Теоретические основы информатики (концептуальные модели и математические основы)

Лекция № 7. Радемахеровская сложность

А.С. Шундеев

Содержание

Содержание

Радемахеровская сложность

В данной лекции продолжается изучение равномерной сходимости эмпирического риска, но уже с более общих позиций.

Радемахеровская сложность

В данной лекции продолжается изучение равномерной сходимости эмпирического риска, но уже с более общих позиций.

В частности, изначально не делается никаких дополнительных предположений относительно множества меток Y класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ и используемой функции потерь.

Радемахеровская сложность

В данной лекции продолжается изучение равномерной сходимости эмпирического риска, но уже с более общих позиций.

В частности, изначально не делается никаких дополнительных предположений относительно множества меток Y класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ и используемой функции потерь.

С этой целью вводится понятие радемахеровской сложности класса функции и излагается основанный на этом понятии подход к получению так называемых эмпирических и предсказывающих оценок.

Радемахеровская сложность

После этого вновь будет рассматриваться задача бинарной классификации, но уже с использованием понятия и свойств радемахеровской сложности.

Радемахеровская сложность

После этого вновь будет рассматриваться задача бинарной классификации, но уже с использованием понятия и свойств радемахеровской сложности.

Основным результатом здесь является доказательство теоремы Дадли, одним из следствий которой является улучшение оценки для функции сложности обучающей выборки, позволяющей убрать логарифм в её правой части:

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon, \delta) = \mathcal{O}\left(\frac{\mathrm{vc}(\mathcal{H}) + \ln(1/\delta)}{\varepsilon^2}\right). \tag{1}$$

Содержание

Радемахеровская сложность

Радемахеровское среднее

В основе определения радемахеровской сложности класса функций лежит понятие радемахеровского среднего ограниченного множества в $\mathbb{R}^n$.

Радемахеровская сложность

Радемахеровское среднее

В основе определения радемахеровской сложности класса функций лежит понятие радемахеровского среднего ограниченного множества в $\mathbb{R}^n$.

С определения этого понятия мы и начнём.

Радемахеровская сложность

Радемахеровское среднее

В основе определения радемахеровской сложности класса функций лежит понятие радемахеровского среднего ограниченного множества в $\mathbb{R}^n$.

С определения этого понятия мы и начнём.

Определение 5.7.

Радемахеровским средним ограниченного множества $A \subset \mathbb{R}^n$ $(n \in \mathbb{N})$ называется число

$$\mathcal{R}_n(A) := \mathbf{E}\left[\sup_{(a_1,\ldots,a_n)\in A} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i a_i\right],$$

где $\varepsilon_1, \ldots, \varepsilon_n$ – независимые радемахеровские случайные величины.

Радемахеровская сложность

Радемахеровское среднее

Следующие свойства непосредственно вытекают из приведённого определения.

Радемахеровская сложность

Радемахеровское среднее

Следующие свойства непосредственно вытекают из приведённого определения.

Утверждение 5.7.

Пусть $n \in \mathbb{N}$, $A, B \subset \mathbb{R}^n$ – ограниченные множества и $c \in \mathbb{R}$.

Радемахеровская сложность

Радемахеровское среднее

Следующие свойства непосредственно вытекают из приведённого определения.

Утверждение 5.7.

Пусть $n \in \mathbb{N}$, $A, B \subset \mathbb{R}^n$ – ограниченные множества и $c \in \mathbb{R}$.

Тогда

1. $\mathcal{R}_n(cA) = |c|\mathcal{R}_n(A)$;

Радемахеровская сложность

Радемахеровское среднее

Следующие свойства непосредственно вытекают из приведённого определения.

Утверждение 5.7.

Пусть $n \in \mathbb{N}$, $A, B \subset \mathbb{R}^n$ – ограниченные множества и $c \in \mathbb{R}$.

Тогда

1. $\mathcal{R}_n(cA) = |c|\mathcal{R}_n(A)$;
2. если $A \subseteq B$, то $\mathcal{R}_n(A) \leqslant \mathcal{R}_n(B)$;

Следующие свойства непосредственно вытекают из приведённого определения.

Утверждение 5.7.

Пусть $n \in \mathbb{N}$, $A, B \subset \mathbb{R}^n$ – ограниченные множества и $c \in \mathbb{R}$.

Тогда

1. $\mathcal{R}_n(cA) = |c|\mathcal{R}_n(A)$;
2. если $A \subseteq B$, то $\mathcal{R}_n(A) \leqslant \mathcal{R}_n(B)$;
3. $\mathcal{R}_n(A + B) = \mathcal{R}_n(A) + \mathcal{R}_n(B)$.

Радемахеровская сложность

Радемахеровское среднее

Следующие свойства непосредственно вытекают из приведённого определения.

Утверждение 5.7.

Пусть $n \in \mathbb{N}$, $A, B \subset \mathbb{R}^n$ – ограниченные множества и $c \in \mathbb{R}$.

Тогда

1. $\mathcal{R}_n(cA) = |c|\mathcal{R}_n(A)$;
2. если $A \subseteq B$, то $\mathcal{R}_n(A) \leqslant \mathcal{R}_n(B)$;
3. $\mathcal{R}_n(A + B) = \mathcal{R}_n(A) + \mathcal{R}_n(B)$.

Установим связь между радемахеровскими средними ограниченного множества и его выпуклой оболочки.

Радемахеровская сложность

Радемахеровское среднее

Утверждение 5.8.

Пусть $n \in \mathbb{N}$, $A \subset \mathbb{R}^n$ – ограниченное множество. Тогда

$$\mathcal{R}_n(A) = \mathcal{R}_n(\mathrm{conv} A).$$

Утверждение 5.8.

Пусть $n \in \mathbb{N}$, $A \subset \mathbb{R}^n$ – ограниченное множество. Тогда

$$\mathcal{R}_n(A) = \mathcal{R}_n(\mathrm{conv} A).$$

◄ Напомним, что под выпуклой оболочкой множества A понимается множество

$$\mathrm{conv} A := \left\{ \lambda \mathbf{a} + (1 - \lambda)\mathbf{a}' : \mathbf{a}, \mathbf{a}' \in A, \lambda \in [0, 1] \right\}.$$

Утверждение 5.8.

Пусть $n \in \mathbb{N}$, $A \subset \mathbb{R}^n$ – ограниченное множество. Тогда

$$\mathcal{R}_n(A) = \mathcal{R}_n(\mathrm{conv} A).$$

◀ Напомним, что под выпуклой оболочкой множества A понимается множество

$$\mathrm{conv} A := \left\{\lambda \mathbf{a} + (1-\lambda)\mathbf{a}' : \mathbf{a}, \mathbf{a}' \in A, \lambda \in [0,1]\right\}.$$

Для доказательства утверждения достаточно заметить, что

Радемахеровская сложность

Радемахеровское среднее

$$\sup_{\mathbf{a},\mathbf{a}'\in A,\ \lambda\in[0,1]} \langle \lambda\mathbf{a} + (1-\lambda)\mathbf{a}', \epsilon\rangle = \sup_{\mathbf{a},\mathbf{a}'\in A,\ \lambda\in[0,1]} \lambda\langle \mathbf{a}, \epsilon\rangle + (1-\lambda)\langle \mathbf{a}', \epsilon\rangle$$

$$= \sup_{\mathbf{a},\mathbf{a}'\in A} \max_{\lambda\in[0,1]} \lambda\langle \mathbf{a}, \epsilon\rangle + (1-\lambda)\langle \mathbf{a}', \epsilon\rangle$$

$$= \sup_{\mathbf{a},\mathbf{a}'\in A} \max_{\lambda\in\{0,1\}} \lambda\langle \mathbf{a}, \epsilon\rangle + (1-\lambda)\langle \mathbf{a}', \epsilon\rangle$$

$$= \sup_{\mathbf{a}\in A} \langle \mathbf{a}, \epsilon\rangle,$$

где через ϵ обозначен вектор, состоящий из n независимых радемахеровских случайных величин.

■

Радемахеровская сложность

Радемахеровское среднее

Существует альтернативное определение, в котором под радемахеровским средним понимается число

$$\widehat{\mathcal{R}}_n(A) := \mathbf{E}\left[\sup_{(a_1,\ldots,a_n)\in A} \frac{1}{n}\left|\sum_{i=1}^{n} \varepsilon_i a_i\right|\right].$$

Радемахеровская сложность

Радемахеровское среднее

Существует альтернативное определение, в котором под радемахеровским средним понимается число

$$\widehat{\mathcal{R}}_n(A) := \mathbf{E}\left[\sup_{(a_1,\ldots,a_n)\in A} \frac{1}{n}\left|\sum_{i=1}^{n} \varepsilon_i a_i\right|\right].$$

Следующее утверждение устанавливает связь между этими двумя определениями.

Радемахеровская сложность

Радемахеровское среднее

Существует альтернативное определение, в котором под радемахеровским средним понимается число

$$\widehat{\mathcal{R}}_n(A) := \mathbf{E}\left[\sup_{(a_1,\ldots,a_n)\in A}\frac{1}{n}\left|\sum_{i=1}^{n}\varepsilon_i a_i\right|\right].$$

Следующее утверждение устанавливает связь между этими двумя определениями.

Утверждение 5.9.

Пусть $n \in \mathbb{N}$ и $A \subset \mathbb{R}^n$ – ограниченные множество. Тогда

$$\mathcal{R}_n(A) \leqslant \mathcal{R}_n(A \cup -A) = \widehat{\mathcal{R}}_n(A).$$

Радемахеровская сложность

Радемахеровское среднее

Отмеченные свойства радемахеровского среднего подсказывают идею его приближённого вычисления.

Радемахеровская сложность

Радемахеровское среднее

Отмеченные свойства радемахеровского среднего подсказывают идею его приближённого вычисления.

Вначале можно попытаться представить ограниченное множество в виде набора несвязных компонент и свести задачу к вычислению радемахеровского среднего для каждой из них.

Радемахеровская сложность

Радемахеровское среднее

Отмеченные свойства радемахеровского среднего подсказывают идею его приближённого вычисления.

Вначале можно попытаться представить ограниченное множество в виде набора несвязных компонент и свести задачу к вычислению радемахеровского среднего для каждой из них.

Отдельные компоненты можно попытаться приблизить выпуклыми множествами.

Радемахеровская сложность

Радемахеровское среднее

Отмеченные свойства радемахеровского среднего подсказывают идею его приближённого вычисления.

Вначале можно попытаться представить ограниченное множество в виде набора несвязных компонент и свести задачу к вычислению радемахеровского среднего для каждой из них.

Отдельные компоненты можно попытаться приблизить выпуклыми множествами.

Вычисление радемахеровского среднего выпуклого множества очевидно сводится к вычислению радемахеровского среднего конечного множества.

Радемахеровская сложность

Радемахеровское среднее

Отмеченные свойства радемахеровского среднего подсказывают идею его приближённого вычисления.

Вначале можно попытаться представить ограниченное множество в виде набора несвязных компонент и свести задачу к вычислению радемахеровского среднего для каждой из них.

Отдельные компоненты можно попытаться приблизить выпуклыми множествами.

Вычисление радемахеровского среднего выпуклого множества очевидно сводится к вычислению радемахеровского среднего конечного множества.

В этой связи особую роль и важность приобретает следующая лемма.

Радемахеровская сложность

Радемахеровское среднее

Лемма 5.4 (лемма о конечном классе, Массар).

Пусть $A = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}\} \subset \mathbb{R}^n$ $(n, N \in \mathbb{N})$ и $L \in \mathbb{R}_+$ такие, что $\|\mathbf{a}^{(j)}\|_2 \leqslant L$ $(j = 1, \ldots, N)$.

Лемма 5.4 (лемма о конечном классе, Массар).

Пусть $A = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}\} \subset \mathbb{R}^n$ $(n, N \in \mathbb{N})$ и $L \in \mathbb{R}_+$ такие, что $\|\mathbf{a}^{(j)}\|_2 \leqslant L$ $(j = 1, \ldots, N)$.

Тогда

$$\mathcal{R}_n(A) \leqslant \frac{L\sqrt{2\ln N}}{n}. \tag{2}$$

Лемма 5.4 (лемма о конечном классе, Массар).

Пусть $A = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}\} \subset \mathbb{R}^n$ $(n, N \in \mathbb{N})$ и $L \in \mathbb{R}_+$ такие, что $\|\mathbf{a}^{(j)}\|_2 \leqslant L$ $(j = 1, \ldots, N)$.

Тогда

$$\mathcal{R}_n(A) \leqslant \frac{L\sqrt{2 \ln N}}{n}. \tag{2}$$

◀ Определим случайные величины

$$\xi_j := \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i a_i^{(j)} \quad (j = 1, \ldots, N).$$

Радемахеровская сложность

Радемахеровское среднее

Напоминание

Лемма(Хёффдинг). Пусть ξ – случайная величина и $\xi \in [a, b]$ (п.н.) для некоторых $a, b \in \mathbb{R}$, $a < b$. Тогда для любого $\varepsilon > 0$ выполняется неравенство

$$\mathbf{E}\left[e^{\varepsilon(\xi - \mathbf{E}\xi)}\right] \leqslant e^{\frac{\varepsilon^2 (b-a)^2}{8}}.$$

Определение. Случайная величина ξ называется *субгауссовской с параметром масштабирования* $s > 0$, если для любого $\varepsilon > 0$ выполняется неравенство

$$\mathbf{E}\left[e^{\varepsilon(\xi - \mathbf{E}\xi)}\right] \leqslant e^{\frac{\varepsilon^2 s^2}{2}}.$$

Лемма(максимальное неравенство). Пусть $\xi_1, \ldots, \xi_n$ – субгауссовские случайные величины с параметром масштабирования $s > 0$ и нулевым математическим ожиданием.
Тогда

$$\mathbf{E}\left[\max_{1 \leqslant i \leqslant n} \xi_i\right] \leqslant s\sqrt{2 \ln n}.$$

Радемахеровская сложность

Радемахеровское среднее

Из леммы Хёффдинга следует, что случайная величина $\frac{a_i^{(j)}}{n}\varepsilon_i$ является субгауссовской с параметром масштабирования равным $\frac{|a_i^{(j)}|}{n}$.

Радемахеровская сложность

Радемахеровское среднее

Из леммы Хёффдинга следует, что случайная величина $\frac{a_i^{(j)}}{n}\varepsilon_i$ является субгауссовской с параметром масштабирования равным $\frac{|a_i^{(j)}|}{n}$.

Случайная величина ξ_j является суммой независимых субгауссовских случайных величин, поэтому, она сама является субгауссовской с параметром масштабирования s таким, что

$$s^2 = \frac{1}{n^2}\sum_{i=1}^{n}|a_i^{(j)}|^2 = \frac{\|\mathbf{a}^{(j)}\|_2^2}{n^2} \leqslant \left(\frac{L}{n}\right)^2.$$

Радемахеровская сложность

Радемахеровское среднее

Из леммы Хёффдинга следует, что случайная величина $\frac{a_i^{(j)}}{n}\varepsilon_i$ является субгауссовской с параметром масштабирования равным $\frac{|a_i^{(j)}|}{n}$.

Случайная величина ξ_j является суммой независимых субгауссовских случайных величин, поэтому, она сама является субгауссовской с параметром масштабирования s таким, что

$$s^2 = \frac{1}{n^2}\sum_{i=1}^{n}|a_i^{(j)}|^2 = \frac{\|\mathbf{a}^{(j)}\|_2^2}{n^2} \leqslant \left(\frac{L}{n}\right)^2.$$

Используя представление $\mathcal{R}_n(A) = \mathbf{E}\left[\max\{\xi_1,\ldots,\xi_N\}\right]$ и применяя максимальное неравенство (лемма 2.2), получим неравенство (2).

■

Радемахеровская сложность

Радемахеровское среднее

Аналогичное утверждение может быть получено и для альтернативного определения радемахеровского среднего.

Радемахеровская сложность

Радемахеровское среднее

Аналогичное утверждение может быть получено и для альтернативного определения радемахеровского среднего.

Следствие 5.1.

Пусть $A = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}\} \subset \mathbb{R}^n$ $(n, N \in \mathbb{N})$ и $L \in \mathbb{R}_+$ такие, что $\|\mathbf{a}^{(j)}\|_2 \leqslant L$ $(j = 1, \ldots, N)$.

Аналогичное утверждение может быть получено и для альтернативного определения радемахеровского среднего.

Следствие 5.1.

Пусть $A = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}\} \subset \mathbb{R}^n$ $(n, N \in \mathbb{N})$ и $L \in \mathbb{R}_+$ такие, что $\|\mathbf{a}^{(j)}\|_2 \leqslant L$ $(j = 1, \ldots, N)$.

Тогда

$$\widehat{\mathcal{R}}_n(A) \leqslant \frac{L\sqrt{2\ln(2N)}}{n}, \tag{3}$$

Аналогичное утверждение может быть получено и для альтернативного определения радемахеровского среднего.

Следствие 5.1.

Пусть $A = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}\} \subset \mathbb{R}^n$ $(n, N \in \mathbb{N})$ и $L \in \mathbb{R}_+$ такие, что $\|\mathbf{a}^{(j)}\|_2 \leqslant L$ $(j = 1, \ldots, N)$.

Тогда

$$\widehat{\mathcal{R}}_n(A) \leqslant \frac{L\sqrt{2\ln(2N)}}{n}, \tag{3}$$

в частности

$$\widehat{\mathcal{R}}_n(A) \leqslant \frac{2L\sqrt{\ln N}}{n} \qquad (N \geqslant 2). \tag{4}$$

Радемахеровская сложность

Радемахеровское среднее

◀ Подставляя $\widehat{\mathcal{R}}_n(A) = \mathcal{R}_n(A \cup -A)$ в неравество (2), получим первое неравенство (3).

◀ Подставляя $\widehat{\mathcal{R}}_n(A) = \mathcal{R}_n(A \cup -A)$ в неравество (2), получим первое неравенство (3).

Для доказательства второго неравенства (4) достаточно заметить, что $\ln(2N) \leqslant 2 \ln N$ при $N \geqslant 2$.

■

Содержание

Радемахеровская сложность

Радемахеровская сложность

Радемахеровская сложность определяется для класса функций, которые должны быть ограничены в совокупности.

Радемахеровская сложность

Радемахеровская сложность

Радемахеровская сложность определяется для класса функций, которые должны быть ограничены в совокупности.

Формально говоря, в определении не требуется наличие класса гипотез и функции потерь, которые порождают этот функциональный класс.

Радемахеровская сложность определяется для класса функций, которые должны быть ограничены в совокупности.

Формально говоря, в определении не требуется наличие класса гипотез и функции потерь, которые порождают этот функциональный класс.

Однако подобная интерпретация является полезной.

Радемахеровская сложность

Радемахеровская сложность

Определение 5.8.

Пусть $\mathcal{F} \subseteq [a,b]^Z$ (a < b) и $n \in \mathbb{N}$.

Радемахеровская сложность

Радемахеровская сложность

Определение 5.8.

Пусть $\mathcal{F} \subseteq [a, b]^Z$ (a < b) и $n \in \mathbb{N}$.

Эмпирической радемахеровской сложностью класса функций $\mathcal{F}$ называется функция

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) := \mathcal{R}_n(\{(f(z_1), \ldots, f(z_n)) : f \in \mathcal{F}\}) \qquad (\mathbf{z} \in Z^n).$$

Определение 5.8.

Пусть $\mathcal{F} \subseteq [a,b]^{\mathrm{Z}}$ (a < b) и $n \in \mathbb{N}$.

Эмпирической радемахеровской сложностью класса функций $\mathcal{F}$ называется функция

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) := \mathcal{R}_n(\{(f(z_1), \ldots, f(z_n)) : f \in \mathcal{F}\}) \qquad (\mathbf{z} \in \mathrm{Z}^n).$$

Радемахеровской сложностью класса функций $\mathcal{F}$ относительно вероятностной меры $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$ называется число

$$\mathcal{R}_n(\mathcal{F}, \mathrm{P}) := \mathop{\mathbf{E}}_{\mathbf{z} \sim \mathrm{P}^n} \left[\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \right].$$

Если из контекста понятно, о какой вероятностной мере P идёт речь, то для радемахеровской сложности будет использоваться сокращённое обозначение $\mathcal{R}_n(\mathcal{F})$.

Утверждение 5.10.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq [a, b]^{\mathsf{Z}}$ $(a < b)$, $\varepsilon > 0$ и $n \in \mathbb{N}$.

Утверждение 5.10.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq [a, b]^{\mathsf{Z}}$ $(a < b)$, $\varepsilon > 0$ и $n \in \mathbb{N}$.

Тогда

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \mathcal{R}_n(\mathcal{F}) - \mathcal{R}_n(\mathcal{F}, \mathbf{z}) \geqslant \varepsilon\right\} \leqslant \exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right). \tag{5}$$

Утверждение 5.10.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [a, b]^{\mathrm{Z}}$ $(a < b)$, $\varepsilon > 0$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n \left\{ \mathbf{z} \,:\, \mathcal{R}_n(\mathcal{F}) - \mathcal{R}_n(\mathcal{F}, \mathbf{z}) \geqslant \varepsilon \right\} \leqslant \exp\left(\frac{-2n\varepsilon^2}{(b-a)^2} \right). \tag{5}$$

◀ Рассмотрим произвольные $\mathbf{z}, \mathbf{z}' \in \mathrm{Z}^n$, которые отличаются только в k-ой $(1 \leqslant k \leqslant n)$ компоненте.

Тогда

$$\begin{aligned}
\mathcal{R}_n(\mathcal{F}, \mathbf{z}) &= \frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathrm{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i \big(f(z_i') + f(z_i) - f(z_i') \big) \right] \\
&\leqslant \frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathrm{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i f(z_i') \right] + \frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathrm{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i \big(f(z_i) - f(z_i') \big) \right] \\
&\leqslant \mathcal{R}_n(\mathcal{F}, \mathbf{z}') + \frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathrm{Q}_n} \left[\sup_{f \in \mathcal{F}} \epsilon_k \big(f(z_k) - f(z_k') \big) \right] \\
&\leqslant \mathcal{R}_n(\mathcal{F}, \mathbf{z}') + \frac{|b - a|}{n}.
\end{aligned}$$

Таким образом, функция $\mathcal{R}_n(\mathcal{F}, \cdot)$ обладает свойством покоординатных приращений с коэффициентами $c_i = \frac{|b-a|}{n}$ $(i = 1, \ldots, n)$.

Радемахеровская сложность

Таким образом, функция $\mathcal{R}_n(\mathcal{F},\cdot)$ обладает свойством покоординатных приращений с коэффициентами $c_i = \frac{|b-a|}{n}$ $(i = 1,\ldots,n)$.

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Радемахеровская сложность

Таким образом, функция $\mathcal{R}_n(\mathcal{F}, \cdot)$ обладает свойством покоординатных приращений с коэффициентами $c_i = \frac{|b-a|}{n}$ $(i = 1, \ldots, n)$.

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Для функции $\mathcal{R}_n(\mathcal{F}, \cdot)$ и функций координатных проекций запишем неравенство МакДиармида.

Таким образом, функция $\mathcal{R}_n(\mathcal{F}, \cdot)$ обладает свойством покоординатных приращений с коэффициентами $c_i = \frac{|b-a|}{n}$ $(i = 1, \ldots, n)$.

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Для функции $\mathcal{R}_n(\mathcal{F}, \cdot)$ и функций координатных проекций запишем неравенство МакДиармида.

Получим искомое неравенство (5).

$\blacksquare$

Радемахеровская сложность

Лемма 5.5.

Пусть $\mathsf{P} \in \mathcal{M}_{+}^{1}(\mathsf{Z})$, $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$ и $n \in \mathbb{N}$.

Лемма 5.5.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$ и $n \in \mathbb{N}$.

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) = \mathcal{R}_n(\mathcal{H}, \mathbf{x}) \qquad (\mathbf{z} \in \mathsf{Z}^n, \mathbf{x} = \mathbf{z}|_{\mathsf{X}}). \tag{6}$$

Лемма 5.5.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$ и $n \in \mathbb{N}$.

Тогда
$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) = \mathcal{R}_n(\mathcal{H}, \mathbf{x}) \qquad (\mathbf{z} \in \mathsf{Z}^n, \mathbf{x} = \mathbf{z}|_{\mathsf{X}}). \tag{6}$$

Как следствие,
$$\mathcal{R}_n(\mathcal{F}, \mathsf{P}) = \mathcal{R}_n(\mathcal{H}, \mathsf{P}_X). \tag{7}$$

Лемма 5.5.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$ и $n \in \mathbb{N}$.

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) = \mathcal{R}_n(\mathcal{H}, \mathbf{x}) \qquad (\mathbf{z} \in \mathsf{Z}^n, \mathbf{x} = \mathbf{z}|_{\mathsf{X}}). \tag{6}$$

Как следствие,

$$\mathcal{R}_n(\mathcal{F}, \mathsf{P}) = \mathcal{R}_n(\mathcal{H}, \mathsf{P}_X). \tag{7}$$

◀ Зафиксируем произвольный набор $\mathbf{z} \in \mathsf{Z}^n$ и заметим, что $1 - 2y_i \in \{\pm 1\}$ $(i = 1, \ldots, n)$.

Радемахеровская сложность

Радемахеровская сложность

Лемма 5.5.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$ и $n \in \mathbb{N}$.

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) = \mathcal{R}_n(\mathcal{H}, \mathbf{x}) \qquad (\mathbf{z} \in \mathsf{Z}^n, \mathbf{x} = \mathbf{z}|_{\mathsf{X}}). \tag{6}$$

Как следствие,

$$\mathcal{R}_n(\mathcal{F}, \mathsf{P}) = \mathcal{R}_n(\mathcal{H}, \mathsf{P}_X). \tag{7}$$

◀ Зафиксируем произвольный набор $\mathbf{z} \in \mathsf{Z}^n$ и заметим, что $1 - 2y_i \in \{\pm 1\}$ $(i = 1, \ldots, n)$.

Следовательно, если $\varepsilon_1, \ldots, \varepsilon_n$ – независимые радемахеровские случайные величины,

Лемма 5.5.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$ и $n \in \mathbb{N}$.

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) = \mathcal{R}_n(\mathcal{H}, \mathbf{x}) \qquad (\mathbf{z} \in \mathsf{Z}^n, \mathbf{x} = \mathbf{z}|_{\mathsf{X}}). \tag{6}$$

Как следствие,

$$\mathcal{R}_n(\mathcal{F}, \mathsf{P}) = \mathcal{R}_n(\mathcal{H}, \mathsf{P}_X). \tag{7}$$

◀ Зафиксируем произвольный набор $\mathbf{z} \in \mathsf{Z}^n$ и заметим, что $1 - 2y_i \in \{\pm 1\}$ $(i = 1, \ldots, n)$.

Следовательно, если $\varepsilon_1, \ldots, \varepsilon_n$ – независимые радемахеровские случайные величины,

то и случайные величины $(1 - 2y_1)\varepsilon_1, \ldots, (1 - 2y_n)\varepsilon_n$ также будут независимыми и радемахеровскими.

Радемахеровская сложность

Радемахеровская сложность

Для любых $y, y' \in \{0, 1\}$ имеет место равенство
$l_{01}(y, y') = y + y' - 2yy'$,

Радемахеровская сложность

Радемахеровская сложность

Для любых $y, y' \in \{0, 1\}$ имеет место равенство $l_{01}(y, y') = y + y' - 2yy'$,

а значит

$$\begin{aligned}
\mathcal{R}_n(\mathcal{F}, \mathbf{z}) &= \mathbf{E}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i\, l_{01}(y_i, h(x_i))\right] \\
&= \mathbf{E}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \left(y_i + (1-2y_i)h(x_i)\right)\right] \\
&= \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} y_i\varepsilon_i\right] + \mathbf{E}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i\,(1-2y_i)h(x_i)\right] \\
&= 0 + \mathbf{E}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i\, h(x_i)\right] \\
&= \mathcal{R}_n(\mathcal{H}, \mathbf{x}).
\end{aligned}$$

Радемахеровская сложность

Радемахеровская сложность

Перейдём к доказательству равенства (7).

Перейдём к доказательству равенства (7).

Запишем

$$\mathcal{R}_n(\mathcal{F}, \mathrm{P}) = \left| \ (6) \ \right| = \underset{\mathbf{z} \sim \mathrm{P}^n}{\mathbf{E}} \left[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \right] = \left| \ \begin{matrix} \text{Теорема} \\ \text{Фубини} \end{matrix} \ \right|$$

$$= \underset{(x_n, y_n) \sim \mathrm{P}}{\mathbf{E}} \cdots \underset{(x_1, y_1) \sim \mathrm{P}}{\mathbf{E}} \left[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \right] = \left| \ \begin{matrix} \text{Теорема о} \\ \text{дезинтеграции} \end{matrix} \ \right|$$

$$= \underset{(x_n, y_n) \sim \mathrm{P}}{\mathbf{E}} \cdots \underset{(x_2, y_2) \sim \mathrm{P}}{\mathbf{E}} \ \underset{x_1 \sim \mathrm{P}_X}{\mathbf{E}} \ \underset{y_1 \sim \mathrm{P}_{Y|X}(x_1, \cdot)}{\mathbf{E}} \left[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \right]$$

Радемахеровская сложность

Радемахеровская сложность

$$
\begin{aligned}
\mathcal{R}_n(\mathcal{F}, \mathsf{P}) &= \left| \begin{array}{c} \mathcal{R}_n(\mathcal{H}, \mathbf{x}) \\ \text{не зависит от } y_1 \end{array} \right| = \mathop{\mathbf{E}}_{(x_n, y_n)\sim \mathsf{P}} \dots \mathop{\mathbf{E}}_{(x_2, y_2)\sim \mathsf{P}} \mathop{\mathbf{E}}_{x_1 \sim \mathsf{P}_X} \Big[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \Big] \\
&= \left| \begin{array}{c} \text{Теорема} \\ \text{Фубини} \end{array} \right| = \mathop{\mathbf{E}}_{x_1 \sim \mathsf{P}_X} \mathop{\mathbf{E}}_{(x_n, y_n)\sim \mathsf{P}} \dots \mathop{\mathbf{E}}_{(x_2, y_2)\sim \mathsf{P}} \Big[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \Big] \\
&= \dots \\
&= \mathop{\mathbf{E}}_{x_1 \sim \mathsf{P}_X} \dots \mathop{\mathbf{E}}_{x_n \sim \mathsf{P}_X} \Big[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \Big] \\
&= \left| \begin{array}{c} \text{Теорема} \\ \text{Фубини} \end{array} \right| = \mathop{\mathbf{E}}_{\mathbf{x} \sim \mathsf{P}_X^n} \Big[\mathcal{R}_n(\mathcal{H}, \mathbf{x}) \Big] \\
&= \mathcal{R}_n(\mathcal{H}, \mathsf{P}_X).
\end{aligned}
$$

$\blacksquare$

Лемма 5.6.

Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0, 1\}^Z$.

Лемма 5.6.

Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^{Z}$.

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \sqrt{\frac{2\ln\left(\Gamma_{\mathcal{F}}(n)\right)}{n}} \qquad (\mathbf{z} \in Z^n). \tag{8}$$

Лемма 5.6.

Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^Z$.

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \sqrt{\frac{2\ln\left(\Gamma_{\mathcal{F}}(n)\right)}{n}} \qquad (\mathbf{z} \in Z^n). \tag{8}$$

Кроме того, если дополнительно $\mathrm{vc}(\mathcal{F}) < \infty$, то

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \sqrt{\frac{2\mathrm{vc}(\mathcal{F})\ln(n+1)}{n}} \qquad (\mathbf{z} \in Z^n). \tag{9}$$

◀ Неравенство (8) вытекает из леммы Массара о конечном классе.

◀ Неравенство (8) вытекает из леммы Массара о конечном классе.

В качестве A выступает $\mathcal{F}(\mathbf{z})$, в качестве L выступает $\sqrt{n}$, и $N = |\mathcal{F}(\mathbf{z})| \leqslant \Gamma_{\mathcal{F}}(n)$.

◀ Неравенство (8) вытекает из леммы Массара о конечном классе.

В качестве A выступает $\mathcal{F}(\mathbf{z})$, в качестве L выступает $\sqrt{n}$, и $N = |\mathcal{F}(\mathbf{z})| \leqslant \Gamma_{\mathcal{F}}(n)$.

Неравенство (9) следует из предыдущего неравенства (8) и леммы Сауэра-Шелаха.

■

Содержание

Эмпирические и предсказывающие оценки

Отталкиваясь от понятия радемахеровской сложности, перейдём к получению эмпирических и предсказывающих оценок, о которых говорилось в начале этой лекции.

Содержание

Эмпирические и предсказывающие оценки

Неравенство симметризации

Теорема 5.6.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$ и $n \in \mathbb{N}$.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Теорема 5.6.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $n \in \mathbb{N}$.

Определим две функции

$$\Delta^+_{\mathcal{F}}(\mathbf{z}) := \sup_{f \in \mathcal{F}} \big(R(f) - r(f, \mathbf{z})\big), \quad \Delta^-_{\mathcal{F}}(\mathbf{z}) := \sup_{f \in \mathcal{F}} \big(r(f, \mathbf{z}) - R(f)\big) \qquad (\mathbf{z} \in \mathrm{Z}^n).$$

Теорема 5.6.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $n \in \mathbb{N}$.

Определим две функции

$$\Delta^+_{\mathcal{F}}(\mathbf{z}) := \sup_{f\in\mathcal{F}} \big(R(f) - r(f,\mathbf{z})\big), \quad \Delta^-_{\mathcal{F}}(\mathbf{z}) := \sup_{f\in\mathcal{F}} \big(r(f,\mathbf{z}) - R(f)\big) \qquad (\mathbf{z} \in \mathrm{Z}^n).$$

Тогда

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim\mathrm{P}^n} \big[\Delta^+_{\mathcal{F}}(\mathbf{z})\big] \leqslant 2\,\mathcal{R}_n(\mathcal{F}), \tag{10}$$

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim\mathrm{P}^n} \big[\Delta^-_{\mathcal{F}}(\mathbf{z})\big] \leqslant 2\,\mathcal{R}_n(\mathcal{F}). \tag{11}$$

Теорема 5.6.

Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $n \in \mathbb{N}$.

Определим две функции

$$\Delta_{\mathcal{F}}^{+}(\mathbf{z}) := \sup_{f \in \mathcal{F}} \big(R(f) - r(f,\mathbf{z})\big), \quad \Delta_{\mathcal{F}}^{-}(\mathbf{z}) := \sup_{f \in \mathcal{F}} \big(r(f,\mathbf{z}) - R(f)\big) \qquad (\mathbf{z} \in \mathrm{Z}^n).$$

Тогда

$$\mathop{\mathbf{E}}_{\mathbf{z} \sim \mathrm{P}^n} \left[\Delta_{\mathcal{F}}^{+}(\mathbf{z})\right] \leqslant 2\,\mathcal{R}_n(\mathcal{F}), \tag{10}$$

$$\mathop{\mathbf{E}}_{\mathbf{z} \sim \mathrm{P}^n} \left[\Delta_{\mathcal{F}}^{-}(\mathbf{z})\right] \leqslant 2\,\mathcal{R}_n(\mathcal{F}). \tag{11}$$

◀ Докажем неравенство (10). Нераенство (11) проверяется аналогичным образом.

Будем использовать вспомогательное вероятностное пространство $(\mathsf{Z}^{2n}, \otimes^{2n}\mathcal{Z}, \mathsf{P}^{2n})$.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Будем использовать вспомогательное вероятностное пространство $(\mathsf{Z}^{2n}, \otimes^{2n}\mathcal{Z}, \mathsf{P}^{2n})$.

Его элементарные события имеют вид $(z_1, \ldots, z_n, z'_1, \ldots, z'_n)$.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Будем использовать вспомогательное вероятностное пространство $(\mathsf{Z}^{2n}, \otimes^{2n}\mathcal{Z}, \mathsf{P}^{2n})$.

Его элементарные события имеют вид $(z_1, \ldots, z_n, z'_1, \ldots, z'_n)$.

Соответствующие функции координатных проекций $Z_1, \ldots, Z_n, Z'_1, \ldots, Z'_n$ являются независимыми случайными элементами с распределением вероятностей P.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Будем использовать вспомогательное вероятностное пространство $(\mathsf{Z}^{2n}, \otimes^{2n}\mathcal{Z}, \mathsf{P}^{2n})$.

Его элементарные события имеют вид $(z_1, \ldots, z_n, z_1', \ldots, z_n')$.

Соответствующие функции координатных проекций $Z_1, \ldots, Z_n, Z_1', \ldots, Z_n'$ являются независимыми случайными элементами с распределением вероятностей P.

Случайные элементы $\mathbf{Z} := (Z_1, \ldots, Z_n)$ и $\mathbf{Z}' := (Z_1', \ldots, Z_n')$ также являются независимыми и одинаково распределёнными.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Используя свойства условного математического ожидания, запишем

Эмпирические и предсказывающие оценки

Неравенство симметризации

Используя свойства условного математического ожидания, запишем

$$\begin{aligned}
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\right)\,\middle|\,\mathbf{Z}\right]\geqslant\left|\begin{array}{c}\text{Утв. 1.17}\\(3)\end{array}\right|\geqslant\\
&\sup_{f\in\mathcal{F}}\mathbf{E}\left[\left(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\right)\,\middle|\,\mathbf{Z}\right]=\left|\begin{array}{c}\text{Утв. 1.17}\\(1,\ 4)\end{array}\right|=\\
&\sup_{f\in\mathcal{F}}\left(R(f)-r(f,\mathbf{Z})\right)=\\
&\Delta^{+}_{\mathcal{F}}(\mathbf{Z})\quad\text{(п.н.)}.
\end{aligned}\tag{12}$$

Эмпирические и предсказывающие оценки

Неравенство симметризации

Используя свойства условного математического ожидания, запишем

$$\begin{aligned}
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\right)\,\middle|\,\mathbf{Z}\right] \geqslant \left|\begin{matrix}\text{Утв. 1.17}\\ (3)\end{matrix}\right| \geqslant \\
&\sup_{f\in\mathcal{F}}\mathbf{E}\left[\left(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\right)\,\middle|\,\mathbf{Z}\right] = \left|\begin{matrix}\text{Утв. 1.17}\\ (1,\ 4)\end{matrix}\right| = \\
&\sup_{f\in\mathcal{F}}\left(R(f)-r(f,\mathbf{Z})\right) = \\
&\Delta^{+}_{\mathcal{F}}(\mathbf{Z}) \quad (\text{п.н.}).
\end{aligned} \tag{12}$$

Переходя к математическому ожиданию в левой и правой части неравенства (12), получим

Эмпирические и предсказывающие оценки

Неравенство симметризации

Используя свойства условного математического ожидания, запишем

$$\begin{aligned}
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\big(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\big)\,\Big|\,\mathbf{Z}\right] \geqslant \Big|\ \substack{\text{Утв. 1.17}\\ (3)}\ \Big| \geqslant\\
&\sup_{f\in\mathcal{F}}\mathbf{E}\left[\big(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\big)\,\Big|\,\mathbf{Z}\right] = \Big|\ \substack{\text{Утв. 1.17}\\ (1,\ 4)}\ \Big| =\\
&\sup_{f\in\mathcal{F}}\big(R(f)-r(f,\mathbf{Z})\big) =\\
&\Delta^{+}_{\mathcal{F}}(\mathbf{Z}) \quad (\text{п.н.}).
\end{aligned} \tag{12}$$

Переходя к математическому ожиданию в левой и правой части неравенства (12), получим

$$\mathbf{E}\left[\Delta^{+}_{\mathcal{F}}(\mathbf{Z})\right] \leqslant \mathbf{E}\left[\sup_{f\in\mathcal{F}}\big(r(f,\mathbf{Z}')-r(f,\mathbf{Z})\big)\right]. \tag{13}$$

Эмпирические и предсказывающие оценки

Неравенство симметризации

Для любых $\epsilon_1, \ldots, \epsilon_n \in \{\pm 1\}$ и любой функции $f \in \mathcal{F}$ случайные величины

$$\epsilon_1(f(Z_1) - f(Z_1')), \ldots, \epsilon_n(f(Z_n) - f(Z_n'))$$

будут независимыми и одинаково распределёнными.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Для любых $\epsilon_1, \ldots, \epsilon_n \in \{\pm 1\}$ и любой функции $f \in \mathcal{F}$ случайные величины

$$\epsilon_1(f(Z_1) - f(Z_1')), \ldots, \epsilon_n(f(Z_n) - f(Z_n'))$$

будут независимыми и одинаково распределёнными.

Следовательно, одинаково распределены и случайные величины

$$\frac{1}{n}\sum_{i=1}^{n}\left(f(Z_i') - f(Z_i)\right) \sim \frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left(f(Z_i') - f(Z_i)\right).$$

Эмпирические и предсказывающие оценки

Неравенство симметризации

Для любых $\epsilon_1, \ldots, \epsilon_n \in \{\pm 1\}$ и любой функции $f \in \mathcal{F}$ случайные величины

$$\epsilon_1(f(Z_1) - f(Z_1')), \ldots, \epsilon_n(f(Z_n) - f(Z_n'))$$

будут независимыми и одинаково распределёнными.

Следовательно, одинаково распределены и случайные величины

$$\frac{1}{n}\sum_{i=1}^{n}\left(f(Z_i') - f(Z_i)\right) \sim \frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left(f(Z_i') - f(Z_i)\right).$$

Далее, получим

Эмпирические и предсказывающие оценки

Неравенство симметризации

$$
\begin{aligned}
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left(r(f,\mathbf{Z}') - r(f,\mathbf{Z})\right)\right] = \\
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\left(f(Z_i') - f(Z_i)\right)\right] = \\
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left(f(Z_i') - f(Z_i)\right)\right] \leqslant \\
&\mathbf{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(Z_i')\right] + \mathbf{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(Z_i)\right].
\end{aligned}
\tag{14}
$$

Эмпирические и предсказывающие оценки

Неравенство симметризации

Объединим неравенства (13) и (14).

Эмпирические и предсказывающие оценки

Неравенство симметризации

Объединим неравенства (13) и (14).

Применяя теорему о замене переменных в интеграле Лебега, запишем

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\Delta^{+}_{\mathcal{F}}(\mathbf{z})\right] \leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(z_i) + \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(z_i)\right] \tag{15}$$

Эмпирические и предсказывающие оценки

Неравенство симметризации

Объединим неравенства (13) и (14).

Применяя теорему о замене переменных в интеграле Лебега, запишем

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\Delta^{+}_{\mathcal{F}}(\mathbf{z})\right] \leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(z_i) + \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(z_i)\right] \qquad (15)$$

Проинтегрируем левую и правую части неравенства (15) по вероятностной мере Q_n.

Эмпирические и предсказывающие оценки

Неравенство симметризации

Объединим неравенства (13) и (14).

Применяя теорему о замене переменных в интеграле Лебега, запишем

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim\mathrm{P}^n}\left[\Delta^{+}_{\mathcal{F}}(\mathbf{z})\right] \leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim\mathrm{P}^n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(z_i) + \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(z_i)\right] \qquad (15)$$

Проинтегрируем левую и правую части неравенства (15) по вероятностной мере Q_n.

В качестве переменной интегрирования возьмём $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$.

Эмпирические и предсказывающие оценки
Неравенство симметризации

Объединим неравенства (13) и (14).

Применяя теорему о замене переменных в интеграле Лебега, запишем

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\Delta^{+}_{\mathcal{F}}(\mathbf{z})\right] \leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(z_i) + \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(z_i)\right] \tag{15}$$

Проинтегрируем левую и правую части неравенства (15) по вероятностной мере Q_n.

В качестве переменной интегрирования возьмём $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$.

Применяя теорему Фубини для замены порядка интегрирования, получим

Эмпирические и предсказывающие оценки

Неравенство симметризации

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\Delta^+_{\mathcal{F}}(\mathbf{z})\right] \leqslant \\
&\mathop{\mathbf{E}}_{\epsilon\sim \mathrm{Q}_n}\mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(z_i) + \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(z_i)\right] = \\
&\mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\mathop{\mathbf{E}}_{\epsilon\sim \mathrm{Q}_n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(z_i)\right] + \mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\mathop{\mathbf{E}}_{\epsilon\sim \mathrm{Q}_n}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}-\epsilon_i f(z_i)\right] = \\
&2\,\mathcal{R}_n(\mathcal{F}).
\end{aligned}
\tag{16}
$$

■

Содержание

Утверждение 5.11.

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Утверждение 5.11.

Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n\left\{\mathbf{z} : \Delta_{\mathcal{F}}^{+}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1 - \delta, \tag{17}$$

$$\mathrm{P}^n\left\{\mathbf{z} : \Delta_{\mathcal{F}}^{-}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1 - \delta. \tag{18}$$

Утверждение 5.11.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathsf{P}^n\left\{\mathbf{z} \;:\; \Delta_{\mathcal{F}}^{+}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1-\delta, \tag{17}$$

$$\mathsf{P}^n\left\{\mathbf{z} \;:\; \Delta_{\mathcal{F}}^{-}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1-\delta. \tag{18}$$

◀ Установим неравенство (17).

Утверждение 5.11.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}^{+}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1-\delta, \tag{17}$$

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}^{-}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1-\delta. \tag{18}$$

◀ Установим неравенство (17).

Неравенство (18) проверяется аналогичным образом.

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим произвольные $\mathbf{z}, \mathbf{z}' \in Z^n$, которые отличаются только в i-ой $(1 \leqslant i \leqslant n)$ компоненте.

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим произвольные $\mathbf{z}, \mathbf{z}' \in Z^n$, которые отличаются только в i-ой ($1 \leqslant i \leqslant n$) компоненте.

Тогда

$$\begin{aligned}
\Delta^+_{\mathcal{F}}(\mathbf{z}) &= \sup_{f \in \mathcal{F}} \Big[\big(R(f) - r(f, \mathbf{z}') \big) + \big(r(f, \mathbf{z}') - r(f, \mathbf{z}) \big) \Big] \\
&\leqslant \Delta^+_{\mathcal{F}}(\mathbf{z}') + \sup_{f \in \mathcal{F}} \frac{1}{n} \big(f(z_i) - f(z_i') \big) \\
&\leqslant \Delta^+_{\mathcal{F}}(\mathbf{z}') + \frac{1}{n}.
\end{aligned}$$

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим произвольные $\mathbf{z}, \mathbf{z}' \in Z^n$, которые отличаются только в i-ой ($1 \leqslant i \leqslant n$) компоненте.

Тогда

$$\begin{aligned}
\Delta^+_{\mathcal{F}}(\mathbf{z}) &= \sup_{f \in \mathcal{F}} \Big[\big(R(f) - r(f, \mathbf{z}')\big) + \big(r(f, \mathbf{z}') - r(f, \mathbf{z})\big) \Big] \\
&\leqslant \Delta^+_{\mathcal{F}}(\mathbf{z}') + \sup_{f \in \mathcal{F}} \frac{1}{n} \big(f(z_i) - f(z'_i)\big) \\
&\leqslant \Delta^+_{\mathcal{F}}(\mathbf{z}') + \frac{1}{n}.
\end{aligned}$$

Таким образом, функция $\Delta^+_{\mathcal{F}}$ обладает свойством покоординатных приращений с коэффициентами $c_i = \frac{1}{n}$ ($i = 1, \ldots, n$).

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Для функции $\Delta^+_{\mathcal{F}}$ и функций координатных проекций запишем неравенство МакДиармида

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta^+_{\mathcal{F}}(\mathbf{z}) > \mathop{\mathbf{E}}_{\mathbf{z}\sim\mathsf{P}^n}\left[\Delta^+_{\mathcal{F}}(\mathbf{z})\right] + \varepsilon\right\} \leqslant \exp(-2n\varepsilon^2) \quad (\varepsilon > 0). \tag{19}$$

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Для функции $\Delta^+_{\mathcal{F}}$ и функций координатных проекций запишем неравенство МакДиармида

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta^+_{\mathcal{F}}(\mathbf{z}) > \mathop{\mathbf{E}}_{\mathbf{z}\sim\mathsf{P}^n}\left[\Delta^+_{\mathcal{F}}(\mathbf{z})\right] + \varepsilon\right\} \leqslant \exp(-2n\varepsilon^2) \quad (\varepsilon > 0). \qquad (19)$$

Выберем

$$\varepsilon := \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}, \qquad (20)$$

Эмпирические и предсказывающие оценки

Эмпирические оценки

Рассмотрим вероятностное пространство $(\mathsf{Z}^n, \otimes^n \mathcal{Z}, \mathsf{P}^n)$.

Для функции $\Delta^{+}_{\mathcal{F}}$ и функций координатных проекций запишем неравенство МакДиармида

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta^{+}_{\mathcal{F}}(\mathbf{z}) > \underset{\mathbf{z}\sim\mathsf{P}^n}{\mathbf{E}}\left[\Delta^{+}_{\mathcal{F}}(\mathbf{z})\right] + \varepsilon\right\} \leqslant \exp(-2n\varepsilon^2) \quad (\varepsilon > 0). \tag{19}$$

Выберем

$$\varepsilon := \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}, \tag{20}$$

тогда

$$-2n\varepsilon^2 = -2n\frac{1}{2n}\ln\left(\frac{1}{\delta}\right) = \ln(\delta). \tag{21}$$

Объединяя вместе (19), (20) и (21), получим неравенство

$$\mathsf{P}^n \left\{ \mathbf{z} \,:\, \Delta_{\mathcal{F}}^{+}(\mathbf{z}) \leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim\mathsf{P}^n} \left[\Delta_{\mathcal{F}}^{+}(\mathbf{z}) \right] + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)} \right\} \geqslant 1 - \delta.$$

Эмпирические и предсказывающие оценки

Эмпирические оценки

Объединяя вместе (19), (20) и (21), получим неравенство

$$\mathrm{P}^n\left\{\mathbf{z} : \Delta_{\mathcal{F}}^{+}(\mathbf{z}) \leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim \mathrm{P}^n}\left[\Delta_{\mathcal{F}}^{+}(\mathbf{z})\right] + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1-\delta.$$

Применяя к нему неравенство симметризации, получим требуемое неравенство (17).

■

Теорема 5.7.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Теорема 5.7.

Пусть $\mathrm{P} \in \mathcal{M}_{+}^{1}(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n\left\{\mathbf{z} : \forall f \in \mathcal{F} : R(f) \leqslant r(f,\mathbf{z}) + 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1-\delta, \tag{22}$$

Теорема 5.7.

Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n \Big\{ \mathbf{z} : \forall f \in \mathcal{F} : R(f) \leqslant r(f,\mathbf{z}) + 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)} \Big\} \geqslant 1-\delta, \tag{22}$$

$$\mathrm{P}^n \Big\{ \mathbf{z} : \forall f \in \mathcal{F} : R(f) \leqslant r(f,\mathbf{z}) + 2\mathcal{R}_n(\mathcal{F},\mathbf{z}) + 3\sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)} \Big\} \geqslant 1-\delta. \tag{23}$$

◄ Заметим, что неравенство (22) является эквивалентной записью неравенства (17) из предыдущего утв. 5.11.

◀ Заметим, что неравенство (22) является эквивалентной записью неравенства (17) из предыдущего утв. 5.11.

Положим $\delta' := \delta/2$.

◀ Заметим, что неравенство (22) является эквивалентной записью неравенства (17) из предыдущего утв. 5.11.

Положим $\delta' := \delta/2$.

Записывая для δ' неравенство (22), получим

$$\mathrm{P}^n\Big\{\mathbf{z} : \forall f \in \mathcal{F} : R(f) \leqslant \\ r(f, \mathbf{z}) + 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta'}\right)}\Big\} \geqslant 1 - \delta'. \tag{24}$$

Эмпирические и предсказывающие оценки

Эмпирические оценки

Напоминание

Утв. 5.10 Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [a,b]^{\mathrm{Z}}$ $(a < b)$, $\varepsilon > 0$ и $n \in \mathbb{N}$. Тогда

$$\mathrm{P}^n\left\{\mathbf{z} \, : \, \mathcal{R}_n(\mathcal{F}) - \mathcal{R}_n(\mathcal{F}, \mathbf{z}) \geqslant \varepsilon\right\} \leqslant \exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right).$$

Эмпирические и предсказывающие оценки

Эмпирические оценки

Напоминание

Утв. 5.10 Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [a,b]^{\mathsf{Z}}$ $(a < b)$, $\varepsilon > 0$ и $n \in \mathbb{N}$. Тогда

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \mathcal{R}_n(\mathcal{F}) - \mathcal{R}_n(\mathcal{F},\mathbf{z}) \geqslant \varepsilon\right\} \leqslant \exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right).$$

Из утв. 5.10 следует, что

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \mathcal{R}_n(\mathcal{F}) \leqslant \mathcal{R}_n(\mathcal{F},\mathbf{z}) + \sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta'}\right)}\right\} \geqslant 1 - \delta'. \qquad (25)$$

Объединяя вместе (24), (25) и применяя известную оценку для вероятности пересечения двух событий, получим

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \forall f \in \mathcal{F} \,:\, R(f) \leqslant r(f,\mathbf{z}) + 2\mathcal{R}_n(\mathcal{F},\mathbf{z}) + 3\sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta'}\right)}\right\} \geqslant 1 - 2\delta',$$

Объединяя вместе (24), (25) и применяя известную оценку для вероятности пересечения двух событий, получим

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \forall f \in \mathcal{F} \,:\, R(f) \leqslant r(f,\mathbf{z}) + 2\mathcal{R}_n(\mathcal{F},\mathbf{z}) + 3\sqrt{\frac{1}{2n}\ln\left(\frac{1}{\delta'}\right)}\right\} \geqslant 1 - 2\delta',$$

а значит верно неравенство (23).

∎

Содержание

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Теорема 5.8.

Пусть $\mathsf{P} \in \mathcal{M}_{+}^{1}(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Теорема 5.8.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^n)$ справедливо неравенство

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, R(f_{\mathbf{z}}) \leqslant R(\mathcal{F}) + 4\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta. \qquad (26)$$

Теорема 5.8.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto f_{\mathbf{z}}$ ($\mathbf{z} \in \mathrm{Z}^n$) справедливо неравенство

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, R(f_{\mathbf{z}}) \leqslant R(\mathcal{F}) + 4\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta. \tag{26}$$

◀ Для любых $f \in \mathcal{F}$ и $\mathbf{z} \in \mathrm{Z}^n$ имеет место равенство

$$R(f_{\mathbf{z}}) = \Big(R(f_{\mathbf{z}}) - r(f_{\mathbf{z}}, \mathbf{z})\Big) + \Big(r(f_{\mathbf{z}}, \mathbf{z}) - r(f, \mathbf{z})\Big) + \Big(r(f, \mathbf{z}) - R(f)\Big) + R(f).$$

Откуда, учитывая произвольность выбора функции $f \in \mathcal{F}$ и определение метода минимизации эмпирического риска, получим неравенство

$$R(f_{\mathbf{z}}) \leqslant \inf_{f \in \mathcal{F}} R(f) + \Delta^{+}_{\mathcal{F}}(\mathbf{z}) + \Delta^{-}_{\mathcal{F}}(\mathbf{z}). \tag{27}$$

Согласно утв. 5.11 имеют место оценки

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \Delta^{+}_{\mathcal{F}}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \frac{\delta}{2}, \tag{28}$$

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \Delta^{-}_{\mathcal{F}}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \frac{\delta}{2}. \tag{29}$$

Объединяя вместе (27), (28) и (29), и применяя известное неравенство для вероятности пересечения событий, получим искомое неравенство (26).

■

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Теорема 5.9.

Пусть $\mathsf{P} \in \mathcal{M}_+^1(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Теорема 5.9.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta. \tag{30}$$

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Теорема 5.9.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n\left\{\mathbf{z} : \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta. \tag{30}$$

◀ Заметим, что для любого $\varepsilon > 0$ имеет место равенство

$$\mathrm{P}^n\{\mathbf{z} : \Delta_{\mathcal{F}}(\mathbf{z}) > \varepsilon\} = \mathrm{P}^n\{\mathbf{z} : \Delta^+_{\mathcal{F}}(\mathbf{z}) > \varepsilon\} + \mathrm{P}^n\{\mathbf{z} : \Delta^-_{\mathcal{F}}(\mathbf{z}) > \varepsilon\}. \tag{31}$$

Теорема 5.9.

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда

$$\mathrm{P}^n\left\{\mathbf{z} : \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta. \tag{30}$$

◀ Заметим, что для любого $\varepsilon > 0$ имеет место равенство

$$\mathrm{P}^n\{\mathbf{z} : \Delta_{\mathcal{F}}(\mathbf{z}) > \varepsilon\} = \mathrm{P}^n\{\mathbf{z} : \Delta^+_{\mathcal{F}}(\mathbf{z}) > \varepsilon\} + \mathrm{P}^n\{\mathbf{z} : \Delta^-_{\mathcal{F}}(\mathbf{z}) > \varepsilon\}. \tag{31}$$

В равенстве (31) положим

$$\varepsilon := 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)}.$$

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Применим неравенства (28) и (29) из доказательства предыдущей теоремы 5.8 к слагаемым, находящимся в правой части равенства (31).

Применим неравенства (28) и (29) из доказательства предыдущей теоремы 5.8 к слагаемым, находящимся в правой части равенства (31).

Получим эквивалентное (30) неравенство

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) > 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta}\right)}\right\} \leqslant \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

■

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Применим неравенства (28) и (29) из доказательства предыдущей теоремы 5.8 к слагаемым, находящимся в правой части равенства (31).

Получим эквивалентное (30) неравенство

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) > 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta}\right)}\right\} \leqslant \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

■

По лемме 5.5 радемахеровская сложность классов функций $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ и $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, связанных соотношением $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$, совпадает.

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Применим неравенства (28) и (29) из доказательства предыдущей теоремы 5.8 к слагаемым, находящимся в правой части равенства (31).

Получим эквивалентное (30) неравенство

$$\mathsf{P}^n\left\{\mathbf{z} : \Delta_{\mathcal{F}}(\mathbf{z}) > 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta}\right)}\right\} \leqslant \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

■

По лемме 5.5 радемахеровская сложность классов функций $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ и $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, связанных соотношением $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$, совпадает.

Поэтому имеет место следующее утверждение, вытекающее из теорем 5.7 и 5.8.

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Следствие 5.2.

Пусть $\mathcal{H} \subseteq \{0,1\}^{X}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Следствие 5.2.

Пусть $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Тогда

$$\mathsf{P}^n\Big\{\mathbf{z} \,:\, \forall h \in \mathcal{H} \,: R(l_{01}; h) \leqslant r(l_{01}; h, \mathbf{z}) + 2\mathcal{R}_n(\mathcal{H}, \mathsf{P}_X) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)}\Big\} \geqslant 1 - \delta, \tag{32}$$

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Следствие 5.2.

Пусть $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Тогда

$$\mathsf{P}^n \Bigg\{ \mathbf{z} \ : \ \forall h \in \mathcal{H} \ : R(l_{01}; h) \leqslant r(l_{01}; h, \mathbf{z}) + 2\mathcal{R}_n(\mathcal{H}, \mathsf{P}_X) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)} \Bigg\} \geqslant 1 - \delta, \tag{32}$$

$$\mathsf{P}^n \Bigg\{ \mathbf{z} \ : \ \forall h \in \mathcal{H} \ : R(l_{01}; h) \leqslant r(l_{01}; h, \mathbf{z}) + 2\mathcal{R}_n(\mathcal{H}, \mathbf{z}|_{\mathsf{X}}) + 3\sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)} \Bigg\} \geqslant 1 - \delta. \tag{33}$$

Следствие 5.2 (продолжение).

Кроме того, для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto h_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^n)$ справедливо неравенство

$$\mathsf{P}^n \Big\{ \mathbf{z} \ : R(l_{01}; h_{\mathbf{z}}) \leqslant \inf_{h \in \mathcal{H}} R(l_{01}; h) + 4\mathcal{R}_n(\mathcal{H}, \mathsf{P}_X) + \sqrt{\frac{2}{n} \ln \left(\frac{2}{\delta} \right)} \Big\} \geqslant 1 - \delta. \tag{34}$$

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Напоминание

Лемма 5.6 Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$. Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \sqrt{\frac{2\mathrm{vc}(\mathcal{F}) \ln(n+1)}{n}} \qquad (\mathbf{z} \in \mathsf{Z}^n).$$

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Напоминание

Лемма 5.6 Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$. Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \sqrt{\frac{2\mathrm{vc}(\mathcal{F}) \ln(n+1)}{n}} \qquad (\mathbf{z} \in \mathsf{Z}^n).$$

Применяя лемму 5.6, получим следующий вариант неравенства Вапника-Червоненкиса.

Следствие 5.3.

Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$.

Эмпирические и предсказывающие оценки

Предсказывающие оценки

Напоминание

Лемма 5.6 Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$. Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \sqrt{\frac{2\mathrm{vc}(\mathcal{F}) \ln(n+1)}{n}} \qquad (\mathbf{z} \in \mathsf{Z}^n).$$

Применяя лемму 5.6, получим следующий вариант неравенства Вапника-Червоненкиса.

Следствие 5.3.

Пусть $n \in \mathbb{N}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$.

Тогда

$$\underset{\mathbf{z} \sim \mathsf{P}^n}{\mathbf{E}} \left[\Delta^{+}_{\mathcal{F}}(\mathbf{z})\right] \leqslant 2\sqrt{\frac{2\mathrm{vc}(\mathcal{F}) \ln(n+1)}{n}}. \tag{35}$$

Содержание

Содержание

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.9.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathsf{M}$ и $\varepsilon > 0$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.9.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Множество A называется ε-покрытием множества B, если для любого $b \in B$ найдется $a \in A$ такое, что $\rho(a, b) \leqslant \varepsilon$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.9.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Множество A называется ε-покрытием множества B, если для любого $b \in B$ найдется $a \in A$ такое, что $\rho(a, b) \leqslant \varepsilon$.

Минимальным ε-покрытием множества B называется ε-покрытие этого множества, обладающее наименьшей мощностью.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.9.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathsf{M}$ и $\varepsilon > 0$.

Множество A называется **ε-покрытием** множества B, если для любого $b \in B$ найдется $a \in A$ такое, что $\rho(a, b) \leqslant \varepsilon$.

Минимальным ε-покрытием множества B называется **ε-покрытие** этого множества, обладающее наименьшей мощностью.

Мощность минимального ε-покрытия множества B называется **числом ε-покрытия** B и обозначается через $N(\varepsilon, B, \rho)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.9.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Множество A называется ε-покрытием множества B, если для любого $b \in B$ найдется $a \in A$ такое, что $\rho(a, b) \leqslant \varepsilon$.

Минимальным ε-покрытием множества B называется ε-покрытие этого множества, обладающее наименьшей мощностью.

Мощность минимального ε-покрытия множества B называется числом ε-покрытия B и обозначается через $N(\varepsilon, B, \rho)$.

Будет использоваться также краткая запись $N(\varepsilon, B)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.10.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathsf{M}$ и $\varepsilon > 0$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.10.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Подмножество $A \subseteq B$ называется ε-упаковкой множества B, если для любый $a, b \in A$ выполняется условие $\rho(a, b) > \varepsilon$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.10.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Подмножество $A \subseteq B$ называется ε-упаковкой множества B, если для любый $a, b \in A$ выполняется условие $\rho(a, b) > \varepsilon$.

Максимальной ε-упаковкой множества B называется ε-упаковка этого множества, обладающая наибольшей мощностью.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.10.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Подмножество $A \subseteq B$ называется ε-упаковкой множества B, если для любый $a, b \in A$ выполняется условие $\rho(a, b) > \varepsilon$.

Максимальной ε-упаковкой множества B называется ε-упаковка этого множества, обладающая наибольшей мощностью.

Мощность максимальной ε-упаковки множества B называется числом ε-упаковки B и обозначается через $M(\varepsilon, B, \rho)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Определение 5.10.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Подмножество $A \subseteq B$ называется ε-упаковкой множества B, если для любый $a, b \in A$ выполняется условие $\rho(a, b) > \varepsilon$.

Максимальной ε-упаковкой множества B называется ε-упаковка этого множества, обладающая наибольшей мощностью.

Мощность максимальной ε-упаковки множества B называется числом ε-упаковки B и обозначается через $M(\varepsilon, B, \rho)$.

Будет использоваться также краткая запись $M(\varepsilon, B)$.

Определение 5.10.

Пусть (M, ρ) – псевдометрическое пространство, $A, B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Подмножество $A \subseteq B$ называется ε-упаковкой множества B, если для любый $a, b \in A$ выполняется условие $\rho(a, b) > \varepsilon$.

Максимальной ε-упаковкой множества B называется ε-упаковка этого множества, обладающая наибольшей мощностью.

Мощность максимальной ε-упаковки множества B называется числом ε-упаковки B и обозначается через $M(\varepsilon, B, \rho)$.

Будет использоваться также краткая запись $M(\varepsilon, B)$.

Следующее утверждение устанавливает связь между введёнными понятиями.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.7.

Пусть (M, ρ) – псевдометрическое пространство, $B \subseteq \mathsf{M}$ и $\varepsilon > 0$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.7.

Пусть (M, ρ) – псевдометрическое пространство, $B \subseteq \mathsf{M}$ и $\varepsilon > 0$.

Тогда
$$N(\varepsilon/2, B) \geqslant M(\varepsilon, B) \geqslant N(\varepsilon, B).$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.7.

Пусть (M, ρ) – псевдометрическое пространство, $B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Тогда
$$N(\varepsilon/2, B) \geqslant M(\varepsilon, B) \geqslant N(\varepsilon, B).$$

◀ Предположим, что $A \subseteq B$ является максимальной ε-упаковкой множества B.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.7.

Пусть (M, ρ) – псевдометрическое пространство, $B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Тогда

$$N(\varepsilon/2, B) \geqslant M(\varepsilon, B) \geqslant N(\varepsilon, B).$$

◀ Предположим, что $A \subseteq B$ является максимальной ε-упаковкой множества B.

Это означает, что для любого $b \in B \setminus A$ множество $A \cup \{b\}$ не обладает этим свойством.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.7.

Пусть (M, ρ) – псевдометрическое пространство, $B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Тогда

$$N(\varepsilon/2, B) \geqslant M(\varepsilon, B) \geqslant N(\varepsilon, B).$$

◀ Предположим, что $A \subseteq B$ является максимальной ε-упаковкой множества B.

Это означает, что для любого $b \in B \setminus A$ множество $A \cup \{b\}$ не обладает этим свойством.

Следовательно, найдётся $a \in A$ такой, что $\rho(a, b) \leqslant \varepsilon$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.7.

Пусть (M, ρ) – псевдометрическое пространство, $B \subseteq \mathrm{M}$ и $\varepsilon > 0$.

Тогда
$$N(\varepsilon/2, B) \geqslant M(\varepsilon, B) \geqslant N(\varepsilon, B).$$

◀ Предположим, что $A \subseteq B$ является максимальной ε-упаковкой множества B.

Это означает, что для любого $b \in B \setminus A$ множество $A \cup \{b\}$ не обладает этим свойством.

Следовательно, найдётся $a \in A$ такой, что $\rho(a, b) \leqslant \varepsilon$.

Таким образом, A одновременно является ε-покрытием множества B, а значит $M(\varepsilon, B) = |A| \geqslant N(\varepsilon, B)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Предположим, что $C \subseteq \mathrm{M}$ является наименьшим $\varepsilon/2$-покрытием множества B, а $A \subseteq B$ является ε-упаковкой этого множества.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Предположим, что $C \subseteq$ M является наименьшим $\varepsilon/2$-покрытием множества B, а $A \subseteq B$ является ε-упаковкой этого множества.

Это означает, что для любого $c \in C$ шар $\{b \in B : \rho(c, b) \leqslant \varepsilon/2\}$ содержит не более одного элемента из A.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Предположим, что $C \subseteq$ M является наименьшим $\varepsilon/2$-покрытием множества B, а $A \subseteq B$ является ε-упаковкой этого множества.

Это означает, что для любого $c \in C$ шар $\{b \in B : \rho(c, b) \leqslant \varepsilon/2\}$ содержит не более одного элемента из A.

Действительно, если предположить существование двух различных элементов $a, a' \in A$, одновремено принадлежащих такому шару, то из неравенства треугольника будет следовать
$\rho(a, a') \leqslant \rho(c, a') + \rho(c, a') \leqslant \varepsilon$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Предположим, что $C \subseteq \mathrm{M}$ является наименьшим $\varepsilon/2$-покрытием множества B, а $A \subseteq B$ является ε-упаковкой этого множества.

Это означает, что для любого $c \in C$ шар $\{b \in B : \rho(c, b) \leqslant \varepsilon/2\}$ содержит не более одного элемента из A.

Действительно, если предположить существование двух различных элементов $a, a' \in A$, одновремено принадлежащих такому шару, то из неравенства треугольника будет следовать
$\rho(a, a') \leqslant \rho(c, a') + \rho(c, a') \leqslant \varepsilon$.

А это нарушает свойство ε-упаковки, которым по предположению обладает множество A.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Предположим, что $C \subseteq$ M является наименьшим $\varepsilon/2$-покрытием множества B, а $A \subseteq B$ является ε-упаковкой этого множества.

Это означает, что для любого $c \in C$ шар $\{b \in B : \rho(c, b) \leqslant \varepsilon/2\}$ содержит не более одного элемента из A.

Действительно, если предположить существование двух различных элементов $a, a' \in A$, одновремено принадлежащих такому шару, то из неравенства треугольника будет следовать
$\rho(a, a') \leqslant \rho(c, a') + \rho(c, a') \leqslant \varepsilon$.

А это нарушает свойство ε-упаковки, которым по предположению обладает множество A.

Следовательно, $|C| \geqslant |A|$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Предположим, что $C \subseteq$ M является наименьшим $\varepsilon/2$-покрытием множества B, а $A \subseteq B$ является ε-упаковкой этого множества.

Это означает, что для любого $c \in C$ шар $\{b \in B : \rho(c, b) \leqslant \varepsilon/2\}$ содержит не более одного элемента из A.

Действительно, если предположить существование двух различных элементов $a, a' \in A$, одновремено принадлежащих такому шару, то из неравенства треугольника будет следовать
$\rho(a, a') \leqslant \rho(c, a') + \rho(c, a') \leqslant \varepsilon$.

А это нарушает свойство ε-упаковки, которым по предположению обладает множество A.

Следовательно, $|C| \geqslant |A|$.

В виду произвольности выбора ε-упаковки A получим неравенство $N(\varepsilon/2, B) \geqslant M(\varepsilon, B)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

В дальнейшем нас будет интересовать специальный случай псевдометрических пространств,

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

В дальнейшем нас будет интересовать специальный случай псевдометрических пространств,

у которых в качестве множества элементов выступают классы функций $\mathcal{F} \subseteq \mathbb{R}^{\mathsf{Z}}$, В дальнейшем нас будет интересовать специальный случай псевдометрических пространств,

у которых в качестве множества элементов выступают классы функций $\mathcal{F} \subseteq \mathbb{R}^{\mathsf{Z}}$,

а псевдометрики порождаются полунормами вида

$$\|f\|_{p,\mathbf{z}} := \left(\frac{1}{n} \sum_{i=1}^{n} |f(z_i)|^p \right)^{1/p} \qquad (f \in \mathcal{F}),$$

где $p \in [1, \infty)$, $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b \ln(a)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b\ln(a)$.

Тогда

$$a \leqslant \frac{e}{e-1}\, b\ln(b). \tag{36}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b\ln(a)$.

Тогда

$$a \leqslant \frac{e}{e-1}\, b\ln(b). \tag{36}$$

◀ Обозначим $\gamma := \frac{e}{e-1}$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b\ln(a)$.

Тогда

$$a \leqslant \frac{e}{e-1}\, b\ln(b). \tag{36}$$

◀ Обозначим $\gamma := \frac{e}{e-1}$.

Для доказательства этого утверждения достаточно показать, что $a \geqslant b\ln(a)$ при $a \geqslant \gamma\, b\ln(b)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b\ln(a)$.

Тогда

$$a \leqslant \frac{e}{e-1}\, b\ln(b). \tag{36}$$

◀ Обозначим $\gamma := \frac{e}{e-1}$.

Для доказательства этого утверждения достаточно показать, что $a \geqslant b\ln(a)$ при $a \geqslant \gamma\, b\ln(b)$.

С этой целью исследуем поведение функции $\varphi(a) := a - b\ln(a)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b\ln(a)$.

Тогда

$$a \leqslant \frac{e}{e-1}\, b\ln(b). \tag{36}$$

◀ Обозначим $\gamma := \frac{e}{e-1}$.

Для доказательства этого утверждения достаточно показать, что $a \geqslant b\ln(a)$ при $a \geqslant \gamma\, b\ln(b)$.

С этой целью исследуем поведение функции $\varphi(a) := a - b\ln(a)$.

Производная этой функции $\varphi'(a) = 1 - b/a > 0$ при $a \in (b, +\infty)$,

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Утверждение 5.12.

Пусть $a > 0$, $b \geqslant 2e$ и $a \leqslant b\ln(a)$.

Тогда

$$a \leqslant \frac{e}{e-1}\, b\ln(b). \tag{36}$$

◀ Обозначим $\gamma := \frac{e}{e-1}$.

Для доказательства этого утверждения достаточно показать, что $a \geqslant b\ln(a)$ при $a \geqslant \gamma\, b\ln(b)$.

С этой целью исследуем поведение функции $\varphi(a) := a - b\ln(a)$.

Производная этой функции $\varphi'(a) = 1 - b/a > 0$ при $a \in (b, +\infty)$,

а это означает, что сама функция $\varphi(a)$ возрастает на этом промежутке.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Заметим, что $\gamma\, b \ln(b) \in (b, +\infty)$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Заметим, что $\gamma b \ln(b) \in (b, +\infty)$.

Замечание

$e/(e-1) = 1.581976706869326424385002005109011558546869301075396136266787059 6\ldots$

Действительно, $\gamma b \ln(b) > b \ln(2e) > b$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Заметим, что $\gamma b \ln(b) \in (b, +\infty)$.

Замечание

$e/(e-1) = 1.5819767068693264243850020051090115585468693010753961362667870596\ldots$

Действительно, $\gamma b \ln(b) > b \ln(2e) > b$.

Поэтому нам достаточно будет показать, что $\varphi(\gamma b \ln(b)) \geqslant 0$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Заметим, что $\gamma\, b \ln(b) \in (b, +\infty)$.

Замечание

$e/(e-1) = 1.5819767068693264243850020051090115585468693010753961362667870596\ldots$

Действительно, $\gamma\, b \ln(b) > b \ln(2e) > b$.

Поэтому нам достаточно будет показать, что $\varphi(\gamma\, b \ln(b)) \geqslant 0$.

Проверим это

$$\begin{aligned}\varphi(\gamma\, b \ln(b)) &= \gamma\, b \ln(b) - b \ln(\gamma\, b \ln(b)) = \\ &= b\big[(\gamma - 1)\ln(b) - \ln(\gamma \ln(b))\big] = \\ &= b\big[e^{-1}c - \ln(c)\big],\end{aligned}$$

Заметим, что $\gamma\, b \ln(b) \in (b, +\infty)$.

Замечание

$e/(e-1) = 1.581976706869326424385002005109011558546869301075396136266787059 6\ldots$

Действительно, $\gamma\, b \ln(b) > b \ln(2e) > b$.

Поэтому нам достаточно будет показать, что $\varphi(\gamma\, b \ln(b)) \geqslant 0$.

Проверим это

$$\begin{aligned}\varphi(\gamma\, b \ln(b)) &= \gamma\, b \ln(b) - b \ln(\gamma\, b \ln(b)) = \\ &= b\big[(\gamma - 1) \ln(b) - \ln(\gamma \ln(b))\big] = \\ &= b\big[e^{-1}c - \ln(c)\big],\end{aligned}$$

где $c := \gamma \ln(b)$ и $c \geqslant \gamma \ln(2e) > 0$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Покажем, что функции $\psi(c) := e^{-1}c - \ln(c) \geqslant 0$ при $c > 0$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Покажем, что функции $\psi(c) := e^{-1}c - \ln(c) \geqslant 0$ при $c > 0$.

Производная этой функции $\psi(c) = e^{-1} - 1/c$ обращается в нуль при $c = e$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Покажем, что функции $\psi(c) := e^{-1}c - \ln(c) \geqslant 0$ при $c > 0$.

Производная этой функции $\psi(c) = e^{-1} - 1/c$ обращается в нуль при $c = e$.

Это точка минимума функции $\psi(c)$, в которой $\psi(e) = 0$.

■

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.8.

Пусть $\mathcal{F} \subseteq \{0,1\}^{Z}$, $\varepsilon \in (0,1]$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.8.

Пусть $\mathcal{F} \subseteq \{0,1\}^{Z}$, $\varepsilon \in (0,1]$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon} \ln\left(\frac{2e}{\varepsilon}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{37}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.8.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\varepsilon \in (0,1]$, $\mathbf{z} \in \mathsf{Z}^n$ ($n \in \mathbb{N}$).

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon}\ln\left(\frac{2e}{\varepsilon}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{37}$$

◄ Для доказательства данного утверждения воспользуемся так называемым вероятностным методом.

Лемма 5.8.

Пусть $\mathcal{F} \subseteq \{0, 1\}^{\mathsf{Z}}$, $\varepsilon \in (0, 1]$, $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon} \ln\left(\frac{2e}{\varepsilon}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{37}$$

◀ Для доказательства данного утверждения воспользуемся так называемым вероятностным методом.

Вначале выполним вспомогательные построения.

Лемма 5.8.

Пусть $\mathcal{F} \subseteq \{0, 1\}^{\mathsf{Z}}$, $\varepsilon \in (0, 1]$, $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon} \ln\left(\frac{2e}{\varepsilon}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{37}$$

◀ Для доказательства данного утверждения воспользуемся так называемым вероятностным методом.

Вначале выполним вспомогательные построения.

Обозначим множество индексов $\Omega_n := \{1, \ldots, n\}$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Лемма 5.8.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\varepsilon \in (0,1]$, $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon} \ln\left(\frac{2e}{\varepsilon}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \qquad (37)$$

◀ Для доказательства данного утверждения воспользуемся так называемым вероятностным методом.

Вначале выполним вспомогательные построения.

Обозначим множество индексов $\Omega_n := \{1, \ldots, n\}$.

и определим вероятностное пространство $(\Omega_n, 2^{\Omega_n}, \mathsf{U})$ с равномерно распределённой на Ω_n вероятностной мерой U.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Зафиксируем $m \in \mathbb{N}$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Зафиксируем $m \in \mathbb{N}$.

На вероятностном пространстве $(\Omega_n^m, \otimes^m 2^{\Omega_n}, \mathsf{U}^m)$ функции координатных проекций $I_1, \ldots, I_m$ являются независимыми случайными величинами с распределением U.

Зафиксируем $m \in \mathbb{N}$.

На вероятностном пространстве $(\Omega_n^m, \otimes^m 2^{\Omega_n}, \mathsf{U}^m)$ функции координатных проекций $I_1, \ldots, I_m$ являются независимыми случайными величинами с распределением U.

Пусть $\{f_1, \ldots, f_M\} \subseteq \mathcal{F}$, где $M = M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}})$, – максимальная ε-упаковка множества $\mathcal{F}$.

Зафиксируем $m \in \mathbb{N}$.

На вероятностном пространстве $(\Omega_n^m, \otimes^m 2^{\Omega_n}, \mathsf{U}^m)$ функции координатных проекций $I_1, \dots, I_m$ являются независимыми случайными величинами с распределением U.

Пусть $\{f_1, \dots, f_M\} \subseteq \mathcal{F}$, где $M = M(\varepsilon, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}})$, – максимальная ε-упаковка множества $\mathcal{F}$.

Для любой пары индексов $k, l \in \Omega_n$, $k \neq l$ выполняется неравенство

$$\mathsf{U}\{i \,:\, f_k(z_i) \neq f_l(z_i)\} = \|f_k - f_l\|_{1,\mathbf{z}} > \varepsilon,$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

но тогда

$$\begin{aligned}
\mathsf{U}^m\{f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \ldots, m)\} &= \prod_{j=1}^{m} \mathsf{U}^m\{f_k(z_{l_j}) = f_l(z_{l_j})\} \\
&\leqslant (1-\varepsilon)^m \leqslant \left|\ \begin{matrix} 1+u \leqslant e^u \\ (u \in \mathbb{R}) \end{matrix}\ \right| \\
&\leqslant e^{-m\varepsilon}.
\end{aligned}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

но тогда

$$\begin{aligned}
\mathrm{U}^m\{f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \dots, m)\} &= \prod_{j=1}^{m} \mathrm{U}^m\{f_k(z_{l_j}) = f_l(z_{l_j})\} \\
&\leqslant (1-\varepsilon)^m \leqslant \left| \begin{matrix} 1+u \leqslant e^u \\ (u \in \mathbb{R}) \end{matrix} \right| \\
&\leqslant e^{-m\varepsilon}.
\end{aligned}$$

Используя неравенство для вероятности объединения событий, получим

$$\mathrm{U}^m\{\exists\, k, l \,:\, k \neq l \text{ и } f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \dots, m)\} < M^2 e^{-m\varepsilon}. \tag{38}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

но тогда

$$\mathrm{U}^m\left\{f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \ldots, m)\right\} = \prod_{j=1}^{m} \mathrm{U}^m\left\{f_k(z_{l_j}) = f_l(z_{l_j})\right\}$$
$$\leqslant (1-\varepsilon)^m \leqslant \left|\ \begin{matrix} 1+u \leqslant e^u \\ (u \in \mathbb{R}) \end{matrix}\ \right|$$
$$\leqslant e^{-m\varepsilon}.$$

Используя неравенство для вероятности объединения событий, получим

$$\mathrm{U}^m\left\{\exists\, k, l \,:\, k \neq l \text{ и } f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \ldots, m)\right\} < M^2 e^{-m\varepsilon}. \tag{38}$$

Выберем $m := \frac{2}{\varepsilon}\ln(M)$.

но тогда

$$\begin{aligned}
\mathrm{U}^m\{f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \ldots, m)\} &= \prod_{j=1}^{m} \mathrm{U}^m\{f_k(z_{l_j}) = f_l(z_{l_j})\} \\
&\leqslant (1-\varepsilon)^m \leqslant \left| \begin{smallmatrix} 1+u \leqslant e^u \\ (u \in \mathbb{R}) \end{smallmatrix} \right| \\
&\leqslant e^{-m\varepsilon}.
\end{aligned}$$

Используя неравенство для вероятности объединения событий, получим

$$\mathrm{U}^m\{\exists\, k, l\ :\ k \neq l \text{ и } f_k(z_{l_j}) = f_l(z_{l_j})\ (j = 1, \ldots, m)\} < M^2 e^{-m\varepsilon}. \qquad (38)$$

Выберем $m := \frac{2}{\varepsilon} \ln(M)$.

и заметим, что в этом случае вероятность события в левой части неравенства (38) меньше единицы.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Это, в свою очередь, означает существование последовательности индексов $(i_1, \ldots, i_m) \in \Omega_n^m$ такой, что на множестве $\{z_{i_1}, \ldots, z_{i_m}\}$ функции $f_1, \ldots, f_M$ попарно отличаются друг от друга.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Это, в свою очередь, означает существование последовательности индексов $(i_1, \ldots, i_m) \in \Omega_n^m$ такой, что на множестве $\{z_{i_1}, \ldots, z_{i_m}\}$ функции $f_1, \ldots, f_M$ попарно отличаются друг от друга.

Но тогда

$$M \leqslant \Gamma_{\mathcal{F}}(m). \tag{39}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Это, в свою очередь, означает существование последовательности индексов $(i_1, \ldots, i_m) \in \Omega_n^m$ такой, что на множестве $\{z_{i_1}, \ldots, z_{i_m}\}$ функции $f_1, \ldots, f_M$ попарно отличаются друг от друга.

Но тогда

$$M \leqslant \Gamma_{\mathcal{F}}(m). \tag{39}$$

Далее, возможны два случая.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Это, в свою очередь, означает существование последовательности индексов $(i_1, \dots, i_m) \in \Omega_n^m$ такой, что на множестве $\{z_{i_1}, \dots, z_{i_m}\}$ функции $f_1, \dots, f_M$ попарно отличаются друг от друга.

Но тогда

$$M \leqslant \Gamma_{\mathcal{F}}(m). \tag{39}$$

Далее, возможны два случая.

Для краткости обозначим $d := \mathrm{vc}(\mathcal{F})$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Это, в свою очередь, означает существование последовательности индексов $(i_1, \ldots, i_m) \in \Omega_n^m$ такой, что на множестве $\{z_{i_1}, \ldots, z_{i_m}\}$ функции $f_1, \ldots, f_M$ попарно отличаются друг от друга.

Но тогда

$$M \leqslant \Gamma_{\mathcal{F}}(m). \tag{39}$$

Далее, возможны два случая.

Для краткости обозначим $d := \mathrm{vc}(\mathcal{F})$.

Если $m \leqslant d$, то $\Gamma_{\mathcal{F}}(m) = 2^d$ и неравенство (39) принимает вид $M \leqslant 2^d$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Это, в свою очередь, означает существование последовательности индексов $(i_1, \ldots, i_m) \in \Omega_n^m$ такой, что на множестве $\{z_{i_1}, \ldots, z_{i_m}\}$ функции $f_1, \ldots, f_M$ попарно отличаются друг от друга.

Но тогда

$$M \leqslant \Gamma_{\mathcal{F}}(m). \tag{39}$$

Далее, возможны два случая.

Для краткости обозначим $d := \mathrm{vc}(\mathcal{F})$.

Если $m \leqslant d$, то $\Gamma_{\mathcal{F}}(m) = 2^d$ и неравенство (39) принимает вид $M \leqslant 2^d$.

Откуда сразу вытекает искомое неравенство (37).

Замечание

$9 ln(2e) = 15.2383246250395077847550890931235891126795012092422972870861200 85\ldots$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Если $m > d$, то с учётом леммы Сауэра-Шелаха неравенство (39) может быть преобразовано к виду

$$M^{\frac{1}{d}} \leqslant \frac{e\,m}{d} = \frac{2e}{\varepsilon} \ln\left(M^{\frac{1}{d}}\right).$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Если $m > d$, то с учётом леммы Сауэра-Шелаха неравенство (39) может быть преобразовано к виду

$$M^{\frac{1}{d}} \leqslant \frac{e\,m}{d} = \frac{2e}{\varepsilon} \ln \left(M^{\frac{1}{d}}\right).$$

Но тогда выполняется неравенство (36) из предыдущего утв. 5.12, в котором $a = M^{\frac{1}{d}}$ и $b = \frac{e\,m}{d}$.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Если $m > d$, то с учётом леммы Сауэра-Шелаха неравенство (39) может быть преобразовано к виду

$$M^{\frac{1}{d}} \leqslant \frac{e\,m}{d} = \frac{2e}{\varepsilon} \ln\left(M^{\frac{1}{d}}\right).$$

Но тогда выполняется неравенство (36) из предыдущего утв. 5.12, в котором $a = M^{\frac{1}{d}}$ и $b = \frac{e\,m}{d}$.

Это значит

$$M \leqslant \left(\frac{e}{e-1}\frac{2e}{\varepsilon}\ln\left(\frac{2e}{\varepsilon}\right)\right)^{d}.$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Если $m > d$, то с учётом леммы Сауэра-Шелаха неравенство (39) может быть преобразовано к виду

$$M^{\frac{1}{d}} \leqslant \frac{e\,m}{d} = \frac{2e}{\varepsilon} \ln\left(M^{\frac{1}{d}}\right).$$

Но тогда выполняется неравенство (36) из предыдущего утв. 5.12, в котором $a = M^{\frac{1}{d}}$ и $b = \frac{e\,m}{d}$.

Это значит

$$M \leqslant \left(\frac{e}{e-1}\frac{2e}{\varepsilon}\ln\left(\frac{2e}{\varepsilon}\right)\right)^d.$$

Учитывая оценку $\frac{2e^2}{e-1} < 9$, получим требуемое неравенство (37).

■

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Обобщим только что полученный результат.

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Обобщим только что полученный результат.

Лемма 5.9.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Обобщим только что полученный результат.

Лемма 5.9.

Пусть $\mathcal{F} \subseteq \{0,1\}^{Z}$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon^p} \ln\left(\frac{2e}{\varepsilon^p}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{40}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Обобщим только что полученный результат.

Лемма 5.9.

Пусть $\mathcal{F} \subseteq \{0,1\}^{Z}$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon^p} \ln\left(\frac{2e}{\varepsilon^p}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{40}$$

◀ Заметим, что в рассматриваемом случае выполняются равенства

$$\|f_1 - f_2\|_{p,\mathbf{z}}^p = \|f_1 - f_2\|_{1,\mathbf{z}} \qquad (f_1, f_2 \in \mathcal{F}),$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Обобщим только что полученный результат.

Лемма 5.9.

Пусть $\mathcal{F} \subseteq \{0,1\}^Z$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon^p} \ln\left(\frac{2e}{\varepsilon^p}\right)\right)^{\mathrm{vc}(\mathcal{F})}. \tag{40}$$

◀ Заметим, что в рассматриваемом случае выполняются равенства

$$\|f_1 - f_2\|_{p,\mathbf{z}}^p = \|f_1 - f_2\|_{1,\mathbf{z}} \qquad (f_1, f_2 \in \mathcal{F}),$$

а значит

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) = M(\varepsilon^p, \mathcal{F}, \|\cdot\|_{1,\mathbf{z}}). \tag{41}$$

Оценка сверху для радемахеровской сложности

ε-покрытия и ε-упаковки

Объединяя (37) из предыдущей леммы 5.8 и (41), получим искомое неравенство (40).

■

Содержание

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.10 (Дадли).

Пусть $\mathcal{F} \subseteq \mathbb{R}^{\mathsf{Z}}$ и $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.10 (Дадли).

Пусть $\mathcal{F} \subseteq \mathbb{R}^{\mathsf{Z}}$ и $\mathbf{z} \in \mathsf{Z}^n$ $(n \in \mathbb{N})$.

Предположим, что

$$\gamma_0 := \sup_{f \in \mathcal{F}} \|f\|_{2,\mathbf{z}} < \infty.$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.10 (Дадли).

Пусть $\mathcal{F} \subseteq \mathbb{R}^{Z}$ и $\mathbf{z} \in Z^n$ ($n \in \mathbb{N}$).

Предположим, что
$$\gamma_0 := \sup_{f \in \mathcal{F}} \|f\|_{2,\mathbf{z}} < \infty.$$

Тогда
$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \frac{12}{\sqrt{n}} \int\limits_0^{\gamma_0} \sqrt{\ln N(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})}\, d\varepsilon. \tag{42}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.10 (Дадли).

Пусть $\mathcal{F} \subseteq \mathbb{R}^Z$ и $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Предположим, что

$$\gamma_0 := \sup_{f \in \mathcal{F}} \|f\|_{2,\mathbf{z}} < \infty.$$

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \frac{12}{\sqrt{n}} \int_0^{\gamma_0} \sqrt{\ln N(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})}\, d\varepsilon. \tag{42}$$

◀ Определим числовую последовательность $\gamma_j := 2^{-j}\gamma_0$ $(j \in \mathbb{N})$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.10 (Дадли).

Пусть $\mathcal{F} \subseteq \mathbb{R}^{Z}$ и $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$.

Предположим, что

$$\gamma_0 := \sup_{f \in \mathcal{F}} \|f\|_{2,\mathbf{z}} < \infty.$$

Тогда

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \frac{12}{\sqrt{n}} \int\limits_0^{\gamma_0} \sqrt{\ln N(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})}\, d\varepsilon. \tag{42}$$

◀ Определим числовую последовательность $\gamma_j := 2^{-j}\gamma_0$ $(j \in \mathbb{N})$

и выделим подмножества $\mathcal{F}_j \subseteq \mathbb{R}^{Z}$, представляющие собой минимальные γ_j-покрытия множества $\mathcal{F}$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

Таким образом, $|\mathcal{F}_j| = N(\gamma_j, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

Таким образом, $|\mathcal{F}_j| = N(\gamma_j, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})$.

Для каждой функции $f \in \mathcal{F}$ существует последовательность функций $f_j \in \mathcal{F}_j$ ($j \in \mathbb{N}$) такие, что $\|f - f_j\|_{2,\mathbf{z}} \leqslant \gamma_j$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

Таким образом, $|\mathcal{F}_j| = N(\gamma_j, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})$.

Для каждой функции $f \in \mathcal{F}$ существует последовательность функций $f_j \in \mathcal{F}_j$ ($j \in \mathbb{N}$) такие, что $\|f - f_j\|_{2,\mathbf{z}} \leqslant \gamma_j$.

По определению будем считать $f_0 := 0$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

Таким образом, $|\mathcal{F}_j| = N(\gamma_j, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}})$.

Для каждой функции $f \in \mathcal{F}$ существует последовательность функций $f_j \in \mathcal{F}_j$ ($j \in \mathbb{N}$) такие, что $\|f - f_j\|_{2,\mathbf{z}} \leqslant \gamma_j$.

По определению будем считать $f_0 := 0$.

Тогда

$$\begin{aligned}
\frac{1}{n}\left(\sum_{i=1}^{n}\left|f_j(z_i) - f_{j-1}(z_i)\right|^2\right)^{1/2} &= \frac{1}{\sqrt{n}}\|f_j - f_{j-1}\|_{2,\mathbf{z}} \\
&\leqslant \frac{1}{\sqrt{n}}\Big(\|f_j - f\|_{2,\mathbf{z}} + \|f - f_{j-1}\|_{2,\mathbf{z}}\Big) \qquad (43) \\
&\leqslant \frac{\gamma_j + \gamma_{j-1}}{\sqrt{n}} = \frac{3\gamma_j}{\sqrt{n}} \qquad (j \in \mathbb{N}).
\end{aligned}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Зафиксируем $m \in \mathbb{N}$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Зафиксируем $m \in \mathbb{N}$

и подставим представление $f = f - f_m + \sum_{j=1}^{m}(f_j - f_{j-1})$ в определение эмпирической радемахеровской сложности

Оценка сверху для радемахеровской сложности

Теорема Дадли

Зафиксируем $m \in \mathbb{N}$

и подставим представление $f = f - f_m + \sum_{j=1}^{m}(f_j - f_{j-1})$ в определение эмпирической радемахеровской сложности

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) = \frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathsf{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i \left(f(z_i) - f_m(z_i) + \sum_{j=1}^{m} \left(f_j(z_i) - f_{j-1}(z_i) \right) \right) \right]$$
$$\leqslant \frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathsf{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i \left(f(z_i) - f_m(z_i) \right) \right] +$$
$$\frac{1}{n} \sum_{j=1}^{m} \mathop{\mathbf{E}}_{\epsilon \sim \mathsf{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i \left(f_j(z_i) - f_{j-1}(z_i) \right) \right]. \tag{44}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Оценим по отдельности каждое слагаемое в правой части (44).

Оценим по отдельности каждое слагаемое в правой части (44).

Применяя неравенство Коши-Буняковского, получим

$$\sum_{i=1}^{n} \epsilon_i \big(f(z_i) - f_m(z_i)\big) \leqslant \|\epsilon\|_2 \sqrt{n} \|f - f_m\|_{2,\mathbf{z}} \leqslant n\gamma_m. \tag{45}$$

Оценим по отдельности каждое слагаемое в правой части (44).

Применяя неравенство Коши-Буняковского, получим

$$\sum_{i=1}^{n} \epsilon_i \big(f(z_i) - f_m(z_i)\big) \leqslant \|\epsilon\|_2 \sqrt{n} \|f - f_m\|_{2,\mathbf{z}} \leqslant n\gamma_m. \tag{45}$$

Применяя лемму 5.4 о конечном классе и ранее полученную оценку (43), получим

Оценка сверху для радемахеровской сложности

Теорема Дадли

$$\frac{1}{n} \mathop{\mathbf{E}}_{\epsilon \sim \mathrm{Q}_n} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \epsilon_i \big(f_j(z_i) - f_{j-1}(z_i) \big) \right] \leqslant \frac{3\gamma_j}{\sqrt{n}} \sqrt{2 \ln \left(|\mathcal{F}_j| |\mathcal{F}_{j-1}| \right)}$$

$$\leqslant \frac{3\gamma_j}{\sqrt{n}} \sqrt{2 \ln \left(|\mathcal{F}_j|^2 \right)}$$

$$= \frac{6\gamma_j}{\sqrt{n}} \sqrt{\ln \left(N(\gamma_j, \mathcal{F}) \right)} \qquad (j \in \mathbb{N}). \tag{46}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

$$\frac{1}{n}\mathop{\mathbf{E}}_{\epsilon\sim Q_n}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\epsilon_i\big(f_j(z_i)-f_{j-1}(z_i)\big)\right] \leqslant \frac{3\gamma_j}{\sqrt{n}}\sqrt{2\ln\left(|\mathcal{F}_j||\mathcal{F}_{j-1}|\right)}$$

$$\leqslant \frac{3\gamma_j}{\sqrt{n}}\sqrt{2\ln\left(|\mathcal{F}_j|^2\right)}$$

$$= \frac{6\gamma_j}{\sqrt{n}}\sqrt{\ln\big(N(\gamma_j,\mathcal{F})\big)} \qquad (j\in\mathbb{N}). \tag{46}$$

Применяя полученные оценки (45) и (46) к неравенству (44) и учитывая равенства $\gamma_j = 2(\gamma_j - \gamma_{j+1})$ $(j\in\mathbb{N})$, получим

Оценка сверху для радемахеровской сложности

Теорема Дадли

$$
\begin{aligned}
\mathcal{R}_n(\mathcal{F}, \mathbf{z}) &\leqslant \gamma_m + \frac{6\gamma_j}{\sqrt{n}} \sum_{j=1}^{m} \gamma_j \sqrt{\ln\left(N(\gamma_j, \mathcal{F})\right)} \\
&= \gamma_m + \frac{12}{\sqrt{n}} \sum_{j=1}^{m} (\gamma_j - \gamma_{j+1}) \sqrt{\ln\left(N(\gamma_j, \mathcal{F})\right)} \\
&\leqslant \gamma_m + \frac{12}{\sqrt{n}} \int\limits_{\gamma_{m+1}}^{\gamma_0} \sqrt{\ln N(\varepsilon, \mathcal{F})}\, d\varepsilon.
\end{aligned}
\tag{47}
$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

$$
\begin{aligned}
\mathcal{R}_n(\mathcal{F}, \mathbf{z}) &\leqslant \gamma_m + \frac{6\gamma_j}{\sqrt{n}} \sum_{j=1}^{m} \gamma_j \sqrt{\ln\left(N(\gamma_j, \mathcal{F})\right)} \\
&= \gamma_m + \frac{12}{\sqrt{n}} \sum_{j=1}^{m} (\gamma_j - \gamma_{j+1}) \sqrt{\ln\left(N(\gamma_j, \mathcal{F})\right)} \\
&\leqslant \gamma_m + \frac{12}{\sqrt{n}} \int\limits_{\gamma_{m+1}}^{\gamma_0} \sqrt{\ln N(\varepsilon, \mathcal{F})}\, d\varepsilon.
\end{aligned}
\tag{47}
$$

Переходя к пределу при $m \longrightarrow \infty$ в правой части неравенства (47), получим утверждение теоремы.

■

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.11.

Пусть $\mathcal{F} \subseteq \{0, 1\}^Z$ и $\mathrm{vc}(\mathcal{F}) < \infty$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.11.

Пусть $\mathcal{F} \subseteq \{0, 1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$.

Тогда существует положительная константа $\kappa \leqslant 72$, не зависящая от $\mathcal{F}$, такая, что

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \kappa \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}} \qquad (\mathbf{z} \in \mathsf{Z}^n, n \in \mathbb{N}). \tag{48}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.11.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$.

Тогда существует положительная константа $\kappa \leqslant 72$, не зависящая от $\mathcal{F}$, такая, что

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \kappa \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}} \qquad (\mathbf{z} \in \mathsf{Z}^n, n \in \mathbb{N}). \tag{48}$$

◀ Зафиксируем $\varepsilon > 0$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Теорема 5.11.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{F}) < \infty$.

Тогда существует положительная константа $\kappa \leqslant 72$, не зависящая от $\mathcal{F}$, такая, что

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant \kappa \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}} \qquad (\mathbf{z} \in \mathsf{Z}^n, n \in \mathbb{N}). \tag{48}$$

◀ Зафиксируем $\varepsilon > 0$

и заметим, что согласно лемме 5.7 число ε-покрытия множества $\mathcal{F}$ не превосходит соответствующее число ε-упаковки.

Оценка сверху для радемахеровской сложности

Теорема Дадли

В лемме 5.9 даётся оценка (40) для числа ε-упаковки.

Замечание

Лемма 5.9. Пусть $\mathcal{F} \subseteq \{0,1\}^{Z}$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$. Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon^p} \ln\left(\frac{2e}{\varepsilon^p}\right)\right)^{\mathrm{vc}(\mathcal{F})}.$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

В лемме 5.9 даётся оценка (40) для числа ε-упаковки.

Замечание

Лемма 5.9. Пусть $\mathcal{F} \subseteq \{0,1\}^Z$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$. Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon^p} \ln\left(\frac{2e}{\varepsilon^p}\right)\right)^{\mathrm{vc}(\mathcal{F})}.$$

Логарифмируя левую и правую части этой оценки и применяя неравенство $\ln x \leqslant x/e$ $(x > 0)$,

Оценка сверху для радемахеровской сложности
Теорема Дадли

В лемме 5.9 даётся оценка (40) для числа ε-упаковки.

Замечание

Лемма 5.9. Пусть $\mathcal{F} \subseteq \{0,1\}^{Z}$, $\varepsilon \in (0,1]$, $p \in [1,\infty)$, $\mathbf{z} \in Z^n$ $(n \in \mathbb{N})$. Тогда

$$M(\varepsilon, \mathcal{F}, \|\cdot\|_{p,\mathbf{z}}) \leqslant \left(\frac{9}{\varepsilon^p} \ln\left(\frac{2e}{\varepsilon^p}\right)\right)^{\mathrm{vc}(\mathcal{F})}.$$

Логарифмируя левую и правую части этой оценки и применяя неравенство $\ln x \leqslant x/e$ $(x > 0)$,

получим

$$\ln N(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mathbf{z}}) \leqslant \mathrm{vc}(\mathcal{F}) \ln\left(\frac{18}{\varepsilon^4}\right). \tag{49}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

С учётом (49) неравенство (42) из теоремы Дадли примет вид

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant 12\sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}} \int\limits_0^1 \sqrt{\ln 18 + 4\ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon. \qquad (50)$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

С учётом (49) неравенство (42) из теоремы Дадли примет вид

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant 12\sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}} \int\limits_0^1 \sqrt{\ln 18 + 4 \ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon. \qquad (50)$$

Здесь неявно было использовано то, что в рассматриваемом случае $\gamma_0 \leqslant 1$.

Оценка сверху для радемахеровской сложности

Теорема Дадли

С учётом (49) неравенство (42) из теоремы Дадли примет вид

$$\mathcal{R}_n(\mathcal{F}, \mathbf{z}) \leqslant 12\sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}} \int\limits_0^1 \sqrt{\ln 18 + 4\ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon. \qquad (50)$$

Здесь неявно было использовано то, что в рассматриваемом случае $\gamma_0 \leqslant 1$.

Оценим сверху интеграл в правой части неравенства (50).

Оценка сверху для радемахеровской сложности

Теорема Дадли

Получим

$$\int\limits_0^1 \sqrt{\ln 18 + 4\ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon \leqslant 2\int\limits_0^1 \sqrt{1 + \ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon \leqslant 2\int\limits_0^1 \left(2 + \ln\left(\frac{1}{\varepsilon}\right)\right) d\varepsilon$$
$$= 2\left[3\varepsilon + \varepsilon\ln\left(\frac{1}{\varepsilon}\right)\right]\Bigg|_0^1 = 6. \tag{51}$$

Оценка сверху для радемахеровской сложности

Теорема Дадли

Получим

$$\int_0^1 \sqrt{\ln 18 + 4\ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon \leqslant 2\int_0^1 \sqrt{1 + \ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon \leqslant 2\int_0^1 \left(2 + \ln\left(\frac{1}{\varepsilon}\right)\right) d\varepsilon$$
$$= 2\left[3\varepsilon + \varepsilon\ln\left(\frac{1}{\varepsilon}\right)\right]\Big|_0^1 = 6. \tag{51}$$

Объединение (50) и (51) даёт искомое неравенство (48).

■

Оценка сверху для радемахеровской сложности

Теорема Дадли

Получим

$$\int\limits_0^1 \sqrt{\ln 18 + 4\ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon \leqslant 2\int\limits_0^1 \sqrt{1 + \ln\left(\frac{1}{\varepsilon}\right)}\, d\varepsilon \leqslant 2\int\limits_0^1 \left(2 + \ln\left(\frac{1}{\varepsilon}\right)\right) d\varepsilon$$
$$= 2\left[3\varepsilon + \varepsilon\ln\left(\frac{1}{\varepsilon}\right)\right]\Bigg|_0^1 = 6. \tag{51}$$

Объединение (50) и (51) даёт искомое неравенство (48).

■

Замечание 5.1.

Более точное вычисление интеграла в левой части неравенства (49) даёт оценку константы $\kappa \leqslant 31$.

Содержание

Применим только что доказанную теорему 5.11 к оценкам из теорем 5.7, 5.8 и 5.9.

Учитывая лемму 5.2, получим общее следствие, вытекающее из этих теорем.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Замечание

Теорема 5.7. Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Тогда

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \forall f \in \mathcal{F} \,:\, R(f) \leqslant r(f,\mathbf{z}) + 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)}\right\} \geqslant 1 - \delta.$$

Теорема 5.8. Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Тогда для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathrm{Z}^n)$ справедливо неравенство

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, R(f_{\mathbf{z}}) \leqslant R(\mathcal{F}) + 4\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{2}{n} \ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta.$$

Теорема 5.9. Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Тогда

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant 2\mathcal{R}_n(\mathcal{F}) + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1 - \delta.$$

Лемма 5.2. Пусть $\mathcal{H} \subseteq \{0,1\}^{\mathrm{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathrm{Z}}$ и $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$. Тогда $\mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{F})$.

Теорема 5.12.

Пусть $\mathcal{H} \subseteq \{0,1\}^{X}$, $\mathrm{vc}(\mathcal{H}) < \infty$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.12.

Пусть $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathrm{vc}(\mathcal{H}) < \infty$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда существует положительная константа $\kappa \leqslant 31$ (с учётом замечания 5.1) такая, что

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.12.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$, $\mathrm{vc}(\mathcal{H}) < \infty$, $\delta \in (0, 1)$ и $n \in \mathbb{N}$.

Тогда существует положительная константа $\kappa \leqslant 31$ (с учётом замечания 5.1) такая, что

$$\begin{aligned} \mathrm{P}^n \Bigg\{ \mathbf{z} \,:\, \forall h \in \mathcal{H} \,:\, R(l_{01}; h) \leqslant r(l_{01}; h, \mathbf{z}) + 2\kappa \sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} \\ + \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta}\right)} \Bigg\} \geqslant 1 - \delta. \end{aligned} \tag{52}$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.12.

Пусть $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathrm{vc}(\mathcal{H}) < \infty$, $\delta \in (0,1)$ и $n \in \mathbb{N}$.

Тогда существует положительная константа $\kappa \leqslant 31$ (с учётом замечания 5.1) такая, что

$$\mathsf{P}^n \Bigg\{ \mathbf{z} \,:\, \forall h \in \mathcal{H} \,:\, R(l_{01}; h) \leqslant r(l_{01}; h, \mathbf{z}) + 2\kappa \sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} + \sqrt{\frac{1}{2n} \ln \left(\frac{1}{\delta} \right)} \Bigg\} \geqslant 1 - \delta. \tag{52}$$

Для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto h_{\mathbf{z}}$ ($\mathbf{z} \in \mathsf{Z}^n$) справедливо неравенство

Теорема 5.12 (продолжение).

Для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto h_{\mathbf{z}}$ ($\mathbf{z} \in Z^n$) справедливо неравенство

$$P^n\Big\{\mathbf{z} : R(l_{01}; h_{\mathbf{z}}) \leqslant \inf_{h\in\mathcal{H}} R(l_{01}; h) + 4\kappa\sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\delta}\right)}\Big\} \geqslant 1-\delta. \tag{53}$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.12 (продолжение).

Для любого метода минимизации эмпирического риска $\mathbf{z} \longmapsto h_{\mathbf{z}}$ ($\mathbf{z} \in Z^n$) справедливо неравенство

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, R(l_{01};h_{\mathbf{z}}) \leqslant \inf_{h\in\mathcal{H}} R(l_{01};h) + 4\kappa\sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1-\delta. \tag{53}$$

Тогда

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{H},l_{01}}(\mathbf{z}) \leqslant 2\kappa\sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} + \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta}\right)}\right\} \geqslant 1-\delta. \tag{54}$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.6.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.6.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$

и для рассматриваемых гипотезы $h \in \mathcal{H}$ и набора примеров $\mathbf{z} \in Z^n$ эмпирический риск удовлетворяет неравенству $r(l_{01}; h, \mathbf{z}) \leqslant 0.2$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.6.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$

и для рассматриваемых гипотезы $h \in \mathcal{H}$ и набора примеров $\mathbf{z} \in Z^n$ эмпирический риск удовлетворяет неравенству $r(l_{01}; h, \mathbf{z}) \leqslant 0.2$.

Выберем $\delta = 0.01$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.6.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$

и для рассматриваемых гипотезы $h \in \mathcal{H}$ и набора примеров $\mathbf{z} \in Z^n$ эмпирический риск удовлетворяет неравенству $r(l_{01}; h, \mathbf{z}) \leqslant 0.2$.

Выберем $\delta = 0.01$.

Тогда из оценки (52) следует, что при уровне доверия 0.99

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.6.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$

и для рассматриваемых гипотезы $h \in \mathcal{H}$ и набора примеров $\mathbf{z} \in Z^n$ эмпирический риск удовлетворяет неравенству $r(l_{01}; h, \mathbf{z}) \leqslant 0.2$.

Выберем $\delta = 0.01$.

Тогда из оценки (52) следует, что при уровне доверия 0.99

ожидаемый риск будет удовлетворять неравенству $R(l_{01}; h) \leqslant 1.60$ $(0.34, 0.24)$, если $n = 10^4$ $(10^6, 10^7)$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.13.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$ и $1 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.13.

Пусть $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ и $1 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$.

Тогда для любых $\varepsilon, \delta \in (0, 1)$ выполняются неравенства

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.13.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$ и $1 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$.

Тогда для любых $\varepsilon, \delta \in (0, 1)$ выполняются неравенства

$$n_{\mathcal{H}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{249^2}{\varepsilon^2} \left(\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right)\right\rceil, \tag{55}$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.13.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$ и $1 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$.

Тогда для любых $\varepsilon, \delta \in (0, 1)$ выполняются неравенства

$$n_{\mathcal{H}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{249^2}{\varepsilon^2} \left(\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right) \right) \right\rceil, \tag{55}$$

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon, \delta) \leqslant \left\lceil \frac{125^2}{\varepsilon^2} \left(\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right) \right) \right\rceil. \tag{56}$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.13.

Пусть $\mathcal{H} \subseteq \{0, 1\}^{X}$ и $1 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$.

Тогда для любых $\varepsilon, \delta \in (0, 1)$ выполняются неравенства

$$n_{\mathcal{H}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{249^2}{\varepsilon^2} \left(\mathrm{vc}(\mathcal{H}) + \ln \left(\frac{1}{\delta} \right) \right) \right\rceil, \tag{55}$$

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon, \delta) \leqslant \left\lceil \frac{125^2}{\varepsilon^2} \left(\mathrm{vc}(\mathcal{H}) + \ln \left(\frac{1}{\delta} \right) \right) \right\rceil. \tag{56}$$

◀ Зафиксируем произвольные $\varepsilon, \delta \in (0, 1)$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Теорема 5.13.

Пусть $\mathcal{H} \subseteq \{0,1\}^X$ и $1 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$.

Тогда для любых $\varepsilon, \delta \in (0,1)$ выполняются неравенства

$$n_{\mathcal{H}}^{\mathrm{apac}}(\varepsilon,\delta) \leqslant \left\lceil \frac{249^2}{\varepsilon^2}\left(\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right)\right\rceil, \tag{55}$$

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon,\delta) \leqslant \left\lceil \frac{125^2}{\varepsilon^2}\left(\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right)\right\rceil. \tag{56}$$

◀ Зафиксируем произвольные $\varepsilon, \delta \in (0,1)$.

Для любого $n \in \mathbb{N}$ выполняется цепочка неравенств

$$
\begin{aligned}
&4\kappa \sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} + \sqrt{\frac{2}{n} \ln\left(\frac{2}{\delta}\right)} \leqslant \\
&2\sqrt{16\kappa^2 \frac{\mathrm{vc}(\mathcal{H})}{n} + \frac{2\ln 2}{n} + \frac{2}{n} \ln\left(\frac{1}{\delta}\right)} \leqslant \\
&2\sqrt{(16\kappa^2+2)\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]} \leqslant \left| \begin{array}{c}\text{Замечание 5.1}\\ \kappa \leqslant 31\end{array} \right| \leqslant \\
&249\sqrt{\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]}.
\end{aligned}
\tag{57}
$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Ограничивая в (57) правую часть числом ε, получим неравенство

$$249\sqrt{\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]} \leqslant \varepsilon,$$

которое относительно n имеет решение

$$n \geqslant \frac{249^2}{\varepsilon^2}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]. \tag{58}$$

Объединяя вместе (53) из теоремы 5.12 и (57) с (58), получим оценку (55).

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Для любого $n \in \mathbb{N}$ выполняется цепочка неравенств

$$
\begin{aligned}
&2\kappa \sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} + \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\delta}\right)} \leqslant \\
&2\sqrt{4\kappa^2 \frac{\mathrm{vc}(\mathcal{H})}{n} + \frac{\ln 2}{2n} + \frac{1}{2n} \ln\left(\frac{1}{\delta}\right)} \leqslant \\
&2\sqrt{(4\kappa^2 + 1)\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]} \leqslant \left| \begin{array}{c}\text{Замечание 5.1} \\ \kappa \leqslant 31\end{array} \right| \leqslant \\
&125\sqrt{\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]}.
\end{aligned}
\tag{59}
$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Ограничивая в (59) правую часть числом ε, получим неравенство

$$125\sqrt{\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]} \leqslant \varepsilon,$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Ограничивая в (59) правую часть числом ε, получим неравенство

$$125\sqrt{\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]} \leqslant \varepsilon,$$

которое относительно n имеет решение

$$n \geqslant \frac{125^2}{\varepsilon^2}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right] \qquad (60)$$

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Ограничивая в (59) правую часть числом ε, получим неравенство

$$125\sqrt{\frac{1}{n}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right]} \leqslant \varepsilon,$$

которое относительно n имеет решение

$$n \geqslant \frac{125^2}{\varepsilon^2}\left[\mathrm{vc}(\mathcal{H}) + \ln\left(\frac{1}{\delta}\right)\right] \tag{60}$$

Объединяя вместе (54) из теоремы 5.12 и (59) с (60), получим оценку (56).

■

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.7.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.7.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$.

Выберем $\varepsilon = 0.1$, $\delta = 0.01$.

Оценка сверху для радемахеровской сложности

Вероятностные оценки для бинарной классификации

Пример 5.7.

Предположим, что класс гипотез $\mathcal{H}$ имеет размерность $\mathrm{vc}(\mathcal{H}) = 5$.

Выберем $\varepsilon = 0.1$, $\delta = 0.01$.

Тогда из оценки (55) следует, что с уровнем доверия 0.99 для рассматриваемого набора примеров $\mathbf{z} \in Z^n$ неравенство $R(l_{01}; h_{\mathbf{z}}) \leqslant \inf_{h \in \mathcal{H}} R(l_{01}; h) + 0.1$ будет выполняться при $n \geqslant 59553016$.