Теоретические основы информатики (концептуальные модели и математические основы)

Лекция № 6. Фундаментальная теорема бинарной классификации

А.С. Шундеев

Содержание

Фундаментальная теорема бинарной классификации

Наша ближайшая цель – для важного частного случая задачи бинарной классификации устанавливается эквивалентность свойств PAC-обучаемости, агностической PAC-обучаемости и равномерной сходимости эмпирического риска.

Фундаментальная теорема бинарной классификации

Наша ближайшая цель – для важного частного случая задачи бинарной классификации устанавливается эквивалентность свойств PAC-обучаемости, агностической PAC-обучаемости и равномерной сходимости эмпирического риска.

Более того, будет дан точный критерий одновременного выполнения или невыполнения этих свойств в терминах так называемой размерности Вапника-Червоненкиса.

Фундаментальная теорема бинарной классификации

Наша ближайшая цель – для важного частного случая задачи бинарной классификации устанавливается эквавалентность свойств PAC-обучаемости, агностической PAC-обучаемости и равномерной сходимости эмпирического риска.

Более того, будет дан точный критерий одновременного выполнения или невыполнения этих свойств в терминах так называемой размерности Вапника-Червоненкиса.

Соответствующее утверждение носит название фундаментальной теоремы бинарной классификации.

Фундаментальная теорема бинарной классификации

Наша ближайшая цель – для важного частного случая задачи бинарной классификации устанавливается эквавалентность свойств PAC-обучаемости, агностической PAC-обучаемости и равномерной сходимости эмпирического риска.

Более того, будет дан точный критерий одновременного выполнения или невыполнения этих свойств в терминах так называемой размерности Вапника-Червоненкиса.

Соответствующее утверждение носит название фундаментальной теоремы бинарной классификации.

При доказательстве этой теоремы проявляется особая роль, которую играет свойство равномерной сходимости эмпирического риска, и которая обосновывает важность для его тщательного изучения в дальнейшем.

Фундаментальная теорема бинарной классификации

Наиболее трудоёмкой частью доказательства теоремы является проверка следующей импликации.

Фундаментальная теорема бинарной классификации

Наиболее трудоёмкой частью доказательства теоремы является проверка следующей импликации.

Если класс гипотез $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ обладает конечной размерностью Вапника-Червоненкиса $\mathrm{vc}(\mathcal{H}) < \infty$, то он будет также обладать и свойством равномерной сходимости эмпирического риска.

Фундаментальная теорема бинарной классификации

Наиболее трудоёмкой частью доказательства теоремы является проверка следующей импликации.

Если класс гипотез $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ обладает конечной размерностью Вапника-Червоненкиса $\mathrm{vc}(\mathcal{H}) < \infty$, то он будет также обладать и свойством равномерной сходимости эмпирического риска.

При этом, для функции сложности обучающей выборки устанавливается оценка вида

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

Содержание

Размерность Вапника-Червоненкиса

Перейдем к определению размерности Вапника-Червоненкиса и доказательству её свойств.

Размерность Вапника-Червоненкиса

Перейдем к определению размерности Вапника-Червоненкиса и доказательству её свойств.

Одно из таких свойств, которое в дальнейшем будет всегда подразумеваться и без отдельного упоминания использоваться, состоит в следующем.

Размерность Вапника-Червоненкиса

Перейдем к определению размерности Вапника-Червоненкиса и доказательству её свойств.

Одно из таких свойств, которое в дальнейшем будет всегда подразумеваться и без отдельного упоминания использоваться, состоит в следующем.

Как мы уже знаем, с любым классом гипотез и, в частности, с классом бинарных классификаторов $\mathcal{H} \subseteq \{0,1\}^X$ всегда связан некоторый класс функций $\mathcal{F}$, которые уже заданы на множестве примеров.

Размерность Вапника-Червоненкиса

Перейдем к определению размерности Вапника-Червоненкиса и доказательству её свойств.

Одно из таких свойств, которое в дальнейшем будет всегда подразумеваться и без отдельного упоминания использоваться, состоит в следующем.

Как мы уже знаем, с любым классом гипотез и, в частности, с классом бинарных классификаторов $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ всегда связан некоторый класс функций $\mathcal{F}$, которые уже заданы на множестве примеров.

В рассматриваемом случае он будет иметь вид $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$, а значит для него самого также определено понятие размерности Вапника-Червоненкиса.

Размерность Вапника-Червоненкиса

Перейдем к определению размерности Вапника-Червоненкиса и доказательству её свойств.

Одно из таких свойств, которое в дальнейшем будет всегда подразумеваться и без отдельного упоминания использоваться, состоит в следующем.

Как мы уже знаем, с любым классом гипотез и, в частности, с классом бинарных классификаторов $\mathcal{H} \subseteq \{0, 1\}^{X}$ всегда связан некоторый класс функций $\mathcal{F}$, которые уже заданы на множестве примеров.

В рассматриваемом случае он будет иметь вид $\mathcal{F} \subseteq \{0, 1\}^{Z}$, $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$, а значит для него самого также определено понятие размерности Вапника-Червоненкиса.

Более того, оказывается, что $\mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{F})$.

Размерность Вапника-Червоненкиса

С **размерностью Вапника-Червоненкиса** тесно связано понятие **функции роста**.

Размерность Вапника-Червоненкиса

С размерностью Вапника-Червоненкиса тесно связано понятие функции роста.

С помощью этого понятия может быть дано само определение размерности Вапника-Червоненкиса, но главное, оно даёт практический способ её вычисления.

Размерность Вапника-Червоненкиса

С размерностью Вапника-Червоненкиса тесно связано понятие функции роста.

С помощью этого понятия может быть дано само определение размерности Вапника-Червоненкиса, но главное, оно даёт практический способ её вычисления.

Это будет продемонстрировано ниже на примерах.

Размерность Вапника-Червоненкиса

Основным результатом данного раздела является доказательство леммы Сауэра-Шелаха, которая является одним из предварительных этапов в доказательстве фундаментальной теоремы бинарной классификации.

Размерность Вапника-Червоненкиса

Основным результатом данного раздела является доказательство леммы Сауэра-Шелаха, которая является одним из предварительных этапов в доказательстве фундаментальной теоремы бинарной классификации.

Эта лемма содержит верхнюю оценку для функции роста, выраженную через размерность Вапника-Червоненкиса. При этом естественно предполагается конечность последней.

Размерность Вапника-Червоненкиса

Основным результатом данного раздела является доказательство леммы Сауэра-Шелаха, которая является одним из предварительных этапов в доказательстве фундаментальной теоремы бинарной классификации.

Эта лемма содержит верхнюю оценку для функции роста, выраженную через размерность Вапника-Червоненкиса. При этом естественно предполагается конечность последней.

Начнём с формальных определений.

Содержание

Размерность Вапника-Червоненкиса

Основные определения

Определение 5.1.

Будем говорить, что конечное множество $S \subset \mathrm{X}$ разбивается классом концептов $\mathcal{C} \subseteq 2^{\mathrm{X}}$, если для любого подмножества $S' \subseteq S$ найдётся концепт $C \in \mathcal{C}$ такой, что $S' = S \cap C$.

Размерность Вапника-Червоненкиса

Основные определения

Определение 5.1.

Будем говорить, что конечное множество $S \subset \mathrm{X}$ разбивается классом концептов $\mathcal{C} \subseteq 2^{\mathrm{X}}$, если для любого подмножества $S' \subseteq S$ найдётся концепт $C \in \mathcal{C}$ такой, что $S' = S \cap C$.

Иногда будет удобно использовать эквивалентное определение.

Определение 5.2.

Пусть $\mathcal{C} \subseteq 2^{X}$ – класс концептов и $S \subset X$ – непустое конечное множество. Предположим, что элементы этого множества упорядочены, $S = \{x_1, \ldots, x_n\}$ ($n \in \mathbb{N}$). Обозначим

$$\mathcal{C}(S) \coloneqq \{(\mathbf{1}_C(x_1), \ldots, \mathbf{1}_C(x_n)) \, : \, C \in \mathcal{C}\}.$$

Определение 5.2.

Пусть $\mathcal{C} \subseteq 2^X$ – класс концептов и $S \subset X$ – непустое конечное множество. Предположим, что элементы этого множества упорядочены, $S = \{x_1, \ldots, x_n\}$ $(n \in \mathbb{N})$. Обозначим

$$\mathcal{C}(S) \coloneqq \left\{(\mathbf{1}_C(x_1), \ldots, \mathbf{1}_C(x_n)) \;:\; C \in \mathcal{C}\right\}.$$

Будем говорить, что множество S разбивается классом концептов $\mathcal{C}$, если

$$\mathcal{C}(S) = \{0, 1\}^n.$$

Определение 5.2.

Пусть $\mathcal{C} \subseteq 2^{X}$ – класс концептов и $S \subset X$ – непустое конечное множество. Предположим, что элементы этого множества упорядочены, $S = \{x_1, \ldots, x_n\}$ ($n \in \mathbb{N}$). Обозначим

$$\mathcal{C}(S) \coloneqq \{(\mathbf{1}_C(x_1), \ldots, \mathbf{1}_C(x_n)) \,:\, C \in \mathcal{C}\}.$$

Будем говорить, что множество S разбивается классом концептов $\mathcal{C}$, если

$$\mathcal{C}(S) = \{0, 1\}^n.$$

Дополнительно считаем, что $\mathcal{C}$ разбивает пустое множество $\varnothing$.

Размерность Вапника-Червоненкиса

Основные определения

Определение 5.3.

Размерностью Вапника-Червоненкиса класса концептов $\mathcal{C} \subseteq 2^X$ называется величина

$$\mathrm{vc}(\mathcal{C}) := \sup \left\{ |S| \ : \ S \text{ разбивается } \mathcal{C} \right\}$$

Определение 5.3.

Размерностью Вапника-Червоненкиса класса концептов $\mathcal{C} \subseteq 2^X$ называется величина

$$\mathrm{vc}(\mathcal{C}) := \sup\left\{|S| \;:\; S \text{ разбивается } \mathcal{C}\right\}$$

Если $\mathrm{vc}(\mathcal{C}) < \infty$, то $\mathcal{C}$ называется **классом Вапника-Червоненкиса**.

Размерность Вапника-Червоненкиса

Основные определения

Определение 5.4.

Введём функцию роста $\Gamma_{\mathcal{C}}$ класса концептов $\mathcal{C} \subseteq 2^{\mathrm{X}}$. Для каждого $n \in \mathbb{N}$ по определению положим

$$\Gamma_{\mathcal{C}}(n) := \sup_{S \subset \mathrm{X}, |S| = n} |\mathcal{C}(S)|.$$

Определение 5.4.

Введём функцию роста $\Gamma_{\mathcal{C}}$ класса концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$. Для каждого $n \in \mathbb{N}$ по определению положим

$$\Gamma_{\mathcal{C}}(n) := \sup_{S \subset \mathsf{X}, |S| = n} |\mathcal{C}(S)|.$$

Число $\Gamma_{\mathcal{C}}(n)$ называется n -м коэффициентом разбиения класса концептов $\mathcal{C}$.

Размерность Вапника-Червоненкиса

Основные определения

Определение 5.4.

Введём функцию роста $\Gamma_{\mathcal{C}}$ класса концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$. Для каждого $n \in \mathbb{N}$ по определению положим

$$\Gamma_{\mathcal{C}}(n) := \sup_{S \subset \mathsf{X}, |S| = n} |\mathcal{C}(S)|.$$

Число $\Gamma_{\mathcal{C}}(n)$ называется n -м коэффициентом разбиения класса концептов $\mathcal{C}$.

Непосредственно из определения функции роста вытекает следующее утверждение.

Определение 5.4.

Введём функцию роста $\Gamma_{\mathcal{C}}$ класса концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$. Для каждого $n \in \mathbb{N}$ по определению положим

$$\Gamma_{\mathcal{C}}(n) := \sup_{S \subset \mathsf{X}, |S| = n} |\mathcal{C}(S)|.$$

Число $\Gamma_{\mathcal{C}}(n)$ называется n -м коэффициентом разбиения класса концептов $\mathcal{C}$.

Непосредственно из определения функции роста вытекает следующее утверждение.

Утверждение 5.1.

Пусть $\mathcal{C} \subseteq 2^{\mathsf{X}}$. Тогда $\mathrm{vc}(\mathcal{C}) = \sup \left\{ n \in \mathbb{N} \ : \ \Gamma_{\mathcal{C}}(n) = 2^n \right\}$.

Лемма 5.1.

Пусть $\mathcal{C} \subseteq 2^{\mathsf{X}}$ и $\Gamma_{\mathcal{C}}(n) < 2^n$ для некоторого $n \in \mathbb{N}$. Тогда $\Gamma_{\mathcal{C}}(m) < 2^m$ для всех $m > n$.

Лемма 5.1.

Пусть $\mathcal{C} \subseteq 2^{\mathsf{X}}$ и $\Gamma_{\mathcal{C}}(n) < 2^n$ для некоторого $n \in \mathbb{N}$. Тогда $\Gamma_{\mathcal{C}}(m) < 2^m$ для всех $m > n$.

◀ Предположим противное.

Лемма 5.1.

Пусть $\mathcal{C} \subseteq 2^{\mathsf{X}}$ и $\Gamma_{\mathcal{C}}(n) < 2^n$ для некоторого $n \in \mathbb{N}$. Тогда $\Gamma_{\mathcal{C}}(m) < 2^m$ для всех $m > n$.

◀ Предположим противное.

Это означает, что существует $m > n$ такое, что $\Gamma_{\mathcal{C}}(m) = 2^m$.

Размерность Вапника-Червоненкиса

Основные определения

Лемма 5.1.

Пусть $\mathcal{C} \subseteq 2^{\mathrm{X}}$ и $\Gamma_{\mathcal{C}}(n) < 2^n$ для некоторого $n \in \mathbb{N}$. Тогда $\Gamma_{\mathcal{C}}(m) < 2^m$ для всех $m > n$.

◀ Предположим противное.

Это означает, что существует $m > n$ такое, что $\Gamma_{\mathcal{C}}(m) = 2^m$.

Но тогда существует множество $\{x_1, \ldots, x_n, x_{n+1}, \ldots, x_m\}$, которое разбивается классом концептов $\mathcal{C}$.

Размерность Вапника-Червоненкиса

Основные определения

Лемма 5.1.

Пусть $\mathcal{C} \subseteq 2^X$ и $\Gamma_{\mathcal{C}}(n) < 2^n$ для некоторого $n \in \mathbb{N}$. Тогда $\Gamma_{\mathcal{C}}(m) < 2^m$ для всех $m > n$.

◀ Предположим противное.

Это означает, что существует $m > n$ такое, что $\Gamma_{\mathcal{C}}(m) = 2^m$.

Но тогда существует множество $\{x_1, \ldots, x_n, x_{n+1}, \ldots, x_m\}$, которое разбивается классом концептов $\mathcal{C}$.

Следовательно, для любых $b_1, \ldots, b_n \in \{0, 1\}$ найдётся некоторый концепт $C \in \mathcal{C}$ такой, что

$$(\mathbf{1}_C(x_1), \ldots, \mathbf{1}_C(x_n), \mathbf{1}_C(x_{n+1}), \ldots, \mathbf{1}_C(x_n)) = (b_1, \ldots, b_n, 0, \ldots, 0).$$

Это в свою очередь означает, что множество $\{x_1, \dots, x_n\}$ разбивается классом концептов $\mathcal{C}$, а значит $\Gamma_{\mathcal{C}}(n) = 2^n$.

Размерность Вапника-Червоненкиса

Основные определения

Это в свою очередь означает, что множество $\{x_1, \dots, x_n\}$ разбивается классом концептов $\mathcal{C}$, а значит $\Gamma_{\mathcal{C}}(n) = 2^n$.

Мы пришли к противоречию.

$\blacksquare$

Содержание

Определение 5.5.

Размерностью Вапника-Червоненкиса класса функций $\mathcal{H} \subseteq \{0, 1\}^{X}$ называется величина

$$\mathrm{vc}(\mathcal{H}) := \mathrm{vc}(\mathcal{C}),$$

где $\mathcal{C} \subseteq 2^{X}$ и $\mathcal{C} \simeq_{\mathbf{1}} \mathcal{H}$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Определение 5.5.

Размерностью Вапника-Червоненкиса класса функций $\mathcal{H} \subseteq \{0,1\}^{X}$ называется величина

$$\mathrm{vc}(\mathcal{H}) := \mathrm{vc}(\mathcal{C}),$$

где $\mathcal{C} \subseteq 2^{X}$ и $\mathcal{C} \simeq_{1} \mathcal{H}$.

Определим для класса функций $\mathcal{H}$ также функцию роста, полагая

$$\Gamma_{\mathcal{H}}(n) := \Gamma_{\mathcal{C}}(n) \qquad (n \in \mathbb{N}).$$

Лемма 5.2.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$, $\mathcal{F} \subseteq \{0, 1\}^Z$ и $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Лемма 5.2.

Пусть $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0, 1\}^{\mathsf{Z}}$ и $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$.

Тогда

$$\mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{F}).$$

Лемма 5.2.

Пусть $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$.

Тогда
$$\mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{F}).$$

◀ Сначала покажем, что $\mathrm{vc}(\mathcal{H}) \leqslant \mathrm{vc}(\mathcal{F})$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Лемма 5.2.

Пусть $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$, $\mathcal{F} \subseteq \{0, 1\}^{\mathsf{Z}}$ и $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$.

Тогда
$$\mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{F}).$$

◀ Сначала покажем, что $\mathrm{vc}(\mathcal{H}) \leqslant \mathrm{vc}(\mathcal{F})$.

Имеет место равенство

$$h(x) = l_{01}(h(x), 0) \qquad (h \in \mathcal{H}, x \in \mathsf{X}).$$

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Следовательно, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{x_1, \ldots, x_n\}$ разбивается классом функций $\mathcal{H}$, то и n-элементное множество $\{(x_1, 0), \ldots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Следовательно, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{x_1, \ldots, x_n\}$ разбивается классом функций $\mathcal{H}$, то и n-элементное множество $\{(x_1, 0), \ldots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$.

Это означает, что $\mathrm{vc}(\mathcal{H}) \leqslant \mathrm{vc}(\mathcal{F})$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Следовательно, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{x_1, \ldots, x_n\}$ разбивается классом функций $\mathcal{H}$, то и n-элементное множество $\{(x_1, 0), \ldots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$.

Это означает, что $\mathrm{vc}(\mathcal{H}) \leqslant \mathrm{vc}(\mathcal{F})$.

Теперь докажем обратное неравенство $\mathrm{vc}(\mathcal{H}) \geqslant \mathrm{vc}(\mathcal{F})$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Следовательно, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{x_1, \ldots, x_n\}$ разбивается классом функций $\mathcal{H}$, то и n-элементное множество $\{(x_1, 0), \ldots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$.

Это означает, что $\mathrm{vc}(\mathcal{H}) \leqslant \mathrm{vc}(\mathcal{F})$.

Теперь докажем обратное неравенство $\mathrm{vc}(\mathcal{H}) \geqslant \mathrm{vc}(\mathcal{F})$.

Имеет место условие

$$l_{01}(h(x), 0) \neq l_{01}(h(x), 1) \qquad (h \in \mathcal{H}, x \in \mathrm{X}).$$

Следовательно, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{x_1, \dots, x_n\}$ разбивается классом функций $\mathcal{H}$, то и n-элементное множество $\{(x_1, 0), \dots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$.

Это означает, что $\mathrm{vc}(\mathcal{H}) \leqslant \mathrm{vc}(\mathcal{F})$.

Теперь докажем обратное неравенство $\mathrm{vc}(\mathcal{H}) \geqslant \mathrm{vc}(\mathcal{F})$.

Имеет место условие

$$l_{01}(h(x), 0) \neq l_{01}(h(x), 1) \qquad (h \in \mathcal{H}, x \in \mathrm{X}).$$

Следовательно, не существует множества вида $\{(x, 0), (x, 1)\}$ ($x \in \mathrm{X}$), которое разбивается классом функций $\mathcal{F}$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Поэтому, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ разбивается классом функций $\mathcal{F}$, то все x_i ($i = 1, \ldots, n$) в нём будут различными.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Поэтому, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ разбивается классом функций $\mathcal{F}$, то все x_i ($i = 1, \ldots, n$) в нём будут различными.

Имеет место равенство

$$l_{01}(h(x), y) = y \oplus l_{01}(h(x), 0) \qquad (h \in \mathcal{H}, x \in \mathrm{X}, y \in \{0, 1\}),$$

где символ $\oplus$ обозначает операцию сложения по модулю 2.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Поэтому, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ разбивается классом функций $\mathcal{F}$, то все x_i ($i = 1, \ldots, n$) в нём будут различными.

Имеет место равенство

$$l_{01}(h(x), y) = y \oplus l_{01}(h(x), 0) \qquad (h \in \mathcal{H}, x \in \mathrm{X}, y \in \{0, 1\}),$$

где символ $\oplus$ обозначает операцию сложения по модулю 2.

Следовательно, множество $\{(x_1, 0), \ldots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$, но тогда и n-элементное множество $\{x_1, \ldots, x_n\}$ разбивается классом функций $\mathcal{H}$.

Размерность Вапника-Червоненкиса

Обобщение на класс функций

Поэтому, если некоторое n-элементное ($n \in \mathbb{N}$) множество $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ разбивается классом функций $\mathcal{F}$, то все x_i ($i = 1, \ldots, n$) в нём будут различными.

Имеет место равенство

$$l_{01}(h(x), y) = y \oplus l_{01}(h(x), 0) \qquad (h \in \mathcal{H}, x \in \mathrm{X}, y \in \{0, 1\}),$$

где символ $\oplus$ обозначает операцию сложения по модулю 2.

Следовательно, множество $\{(x_1, 0), \ldots, (x_n, 0)\}$ разбивается классом функций $\mathcal{F}$, но тогда и n-элементное множество $\{x_1, \ldots, x_n\}$ разбивается классом функций $\mathcal{H}$.

Это означает, что $\mathrm{vc}(\mathcal{H}) \geqslant \mathrm{vc}(\mathcal{F})$.

$\blacksquare$

Содержание

Размерность Вапника-Червоненкиса

Примеры

Ранее доказанная лемма 5.1 подсказывает практический способ вычисления размерности Вапника-Червоненкиса.

Размерность Вапника-Червоненкиса

Примеры

Ранее доказанная лемма 5.1 подсказывает практический способ вычисления размерности Вапника-Червоненкиса.

Действительно, если предъявить n-элементное множество, которое разбивается рассматриваемым классом концептов $\mathcal{C}$

Размерность Вапника-Червоненкиса

Примеры

Ранее доказанная лемма 5.1 подсказывает практический способ вычисления размерности Вапника-Червоненкиса.

Действительно, если предъявить n-элементное множество, которое разбивается рассматриваемым классом концептов $\mathcal{C}$ и показать, что никакое $n+1$-элементное множество не разбивается этим классом,

Размерность Вапника-Червоненкиса

Примеры

Ранее доказанная лемма 5.1 подсказывает практический способ вычисления размерности Вапника-Червоненкиса.

Действительно, если предъявить n-элементное множество, которое разбивается рассматриваемым классом концептов $\mathcal{C}$ и показать, что никакое $n+1$-элементное множество не разбивается этим классом, то $\mathrm{vc}(\mathcal{C}) = n$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.1.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(-\infty, a) : a \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 1$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.1.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(-\infty, a) : a \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 1$.

Одноэлементное множество $\{0\}$ разбивается классом $\mathcal{C}$. Например, $0 \notin (-\infty, -1)$ и $0 \in (-\infty, 1)$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.1.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(-\infty, a) : a \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 1$.

Одноэлементное множество $\{0\}$ разбивается классом $\mathcal{C}$. Например, $0 \notin (-\infty, -1)$ и $0 \in (-\infty, 1)$.

Предположим, что существуют $b, c \in \mathbb{R}$ $(b < c)$ такие, что 2-элементное множество $\{b, c\}$ разбивается классом $\mathcal{C}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.1.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(-\infty, a) : a \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 1$.

Одноэлементное множество $\{0\}$ разбивается классом $\mathcal{C}$. Например, $0 \notin (-\infty, -1)$ и $0 \in (-\infty, 1)$.

Предположим, что существуют $b, c \in \mathbb{R}$ $(b < c)$ такие, что 2-элементное множество $\{b, c\}$ разбивается классом $\mathcal{C}$.

Следовательно, для некоторого $a \in \mathbb{R}$ должны одновременно выполняться два условия $b \notin (-\infty, a)$ и $c \in (-\infty, a)$, что невозможно.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.2.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(a, b) : a, b \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 2$.

Пример 5.2.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(a, b) : a, b \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 2$.

Выберем различные $a_1, a_2, a_3, a_4, b, c \in \mathbb{R}$ $(a_1 < a_2 < b < a_3 < c < a_4)$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.2.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(a, b) : a, b \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 2$.

Выберем различные $a_1, a_2, a_3, a_4, b, c \in \mathbb{R}$ ($a_1 < a_2 < b < a_3 < c < a_4$).

Тогда

$$(a_1, a_2) \cap \{b, c\} = \varnothing, \quad (a_2, a_3) \cap \{b, c\} = \{b\},$$
$$(a_3, a_4) \cap \{b, c\} = \{c\}, \quad (a_1, a_4) \cap \{b, c\} = \{b, c\},$$

Пример 5.2.

На множестве $\mathbb{R}$ рассмотрим класс концептов $\mathcal{C} = \{(a, b) : a, b \in \mathbb{R}\}$ и покажем, что $\mathrm{vc}(\mathcal{C}) = 2$.

Выберем различные $a_1, a_2, a_3, a_4, b, c \in \mathbb{R}$ $(a_1 < a_2 < b < a_3 < c < a_4)$.

Тогда

$$(a_1, a_2) \cap \{b, c\} = \varnothing, \quad (a_2, a_3) \cap \{b, c\} = \{b\},$$
$$(a_3, a_4) \cap \{b, c\} = \{c\}, \quad (a_1, a_4) \cap \{b, c\} = \{b, c\},$$

а это значит, что 2-элементное множество $\{b, c\}$ разбивается классом $\mathcal{C}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.2 (продолжение).

Возьмём произвольное 3-элементное множество $\{b, c, d\}$ $(b < c < d)$. Тогда

$$(a_1, a_2) \cap \{b, c, d\} \neq \{b, d\},$$

для любых $a_1, a_2 \in \mathbb{R}$ $(a_1 < a_2)$.

Пример 5.3.

На множестве $\mathbb{R}^2$ рассмотрим класс концептов $\mathcal{C}_{\mathrm{rec}}$, состоящий из всех прямоугольников, стороны которых параллельны координатным осям. Покажем, что $\mathrm{vc}(\mathcal{C}_{\mathrm{rec}}) = 4$.

Пример 5.3.

На множестве $\mathbb{R}^2$ рассмотрим класс концептов $\mathcal{C}_{\mathrm{rec}}$, состоящий из всех прямоугольников, стороны которых параллельны координатным осям. Покажем, что $\mathrm{vc}(\mathcal{C}_{\mathrm{rec}}) = 4$.

Класс $\mathcal{C}_{\mathrm{rec}}$ разбивает 4-элементное множество

$$\{p_1, p_2, p_3, p_4\} = \{(-1, 0), (0, 1), (1, 0), (0, -1)\}.$$

Размерность Вапника-Червоненкиса

Примеры

Пример 5.3.

На множестве $\mathbb{R}^2$ рассмотрим класс концептов $\mathcal{C}_{\mathrm{rec}}$, состоящий из всех прямоугольников, стороны которых параллельны координатным осям. Покажем, что $\mathrm{vc}(\mathcal{C}_{\mathrm{rec}}) = 4$.

Класс $\mathcal{C}_{\mathrm{rec}}$ разбивает 4-элементное множество

$$\{p_1, p_2, p_3, p_4\} = \{(-1, 0), (0, 1), (1, 0), (0, -1)\}.$$

В качестве примера, на следующем слайде рис. а) показано выделение с помощью пересечений с прямоугольниками следующих подмножеств этого множества:

$$\varnothing, \{p_2\}, \{p_1, p_3\}, \{p_2, p_3\}, \{p_1, p_2, p_3\}, \{p_1, p_2, p_3, p_4\}.$$

Размерность Вапника-Червоненкиса

Примеры

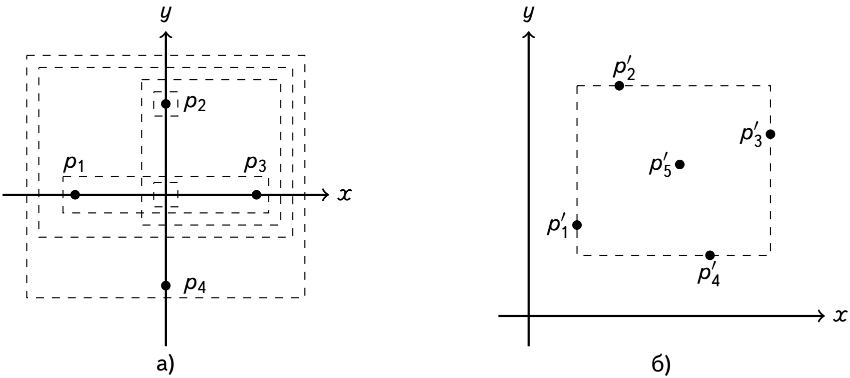


Рис.: а) пример 4-элементного множества, разбиваемого классом $\mathcal{C}_{\mathrm{rec}}$; б) пример, показывающий, что не существует 5-элементного множества, разбиваемого этим классом.

Пример 5.3 (продолжение).

Выберем произвольное 5-элементное множество и перенумеруем его элементы так, как это показано на предыдущем слайде рис. б).

Размерность Вапника-Червоненкиса

Примеры

Пример 5.3 (продолжение).

Выберем произвольное 5-элементное множество и перенумеруем его элементы так, как это показано на предыдущем слайде рис. б).

Через p_1' обозначим элемент с наименьшей x-координатой, а через p_3' – с наибольшей x-координатой.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.3 (продолжение).

Выберем произвольное 5-элементное множество и перенумеруем его элементы так, как это показано на предыдущем слайде рис. б).

Через p_1' обозначим элемент с наименьшей x-координатой, а через p_3' – с наибольшей x-координатой.

Через p_4' обозначим элемент с наименьшей y-координатой, а через p_2' – с наибольшей y-координатой.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.3 (продолжение).

Выберем произвольное 5-элементное множество и перенумеруем его элементы так, как это показано на предыдущем слайде рис. б).

Через p'_1 обозначим элемент с наименьшей x-координатой, а через p'_3 – с наибольшей x-координатой.

Через p'_4 обозначим элемент с наименьшей y-координатой, а через p'_2 – с наибольшей y-координатой.

Оставшийся элемент обозначим через p'_5.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.3 (продолжение).

Выберем произвольное 5-элементное множество и перенумеруем его элементы так, как это показано на предыдущем слайде рис. б).

Через p_1' обозначим элемент с наименьшей x-координатой, а через p_3' – с наибольшей x-координатой.

Через p_4' обозначим элемент с наименьшей y-координатой, а через p_2' – с наибольшей y-координатой.

Оставшийся элемент обозначим через p_5'.

Тогда не существует прямоугольника, пересечение с которым даёт подмножество $\{p_1', p_2', p_3', p_4'\}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4.

Пусть $m \in \mathbb{N}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4.

Пусть $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(m)$ множество всех бинарных классификаторов вида

$$\mathbf{u} \longmapsto \operatorname{sgn}\left(\langle \mathbf{w}, \mathbf{u} \rangle\right) \qquad \left(\mathbf{w}, \mathbf{u} \in \mathbb{R}^m\right),$$

где $\operatorname{sgn}(a) := 1$ для всех $a > 0$ и $\operatorname{sgn}(a) := 0$ для всех $a \leqslant 0$.

Пример 5.4.

Пусть $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(m)$ множество всех бинарных классификаторов вида

$$\mathbf{u} \longmapsto \mathrm{sgn}\left(\langle \mathbf{w}, \mathbf{u} \rangle\right) \qquad \left(\mathbf{w}, \mathbf{u} \in \mathbb{R}^m\right),$$

где $\mathrm{sgn}(a) := 1$ для всех $a > 0$ и $\mathrm{sgn}(a) := 0$ для всех $a \leqslant 0$.

Покажем, что $\mathrm{vc}\left(\mathcal{H}_{\mathrm{lin}}(m)\right) = m$.

Пример 5.4.

Пусть $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(m)$ множество всех бинарных классификаторов вида

$$\mathbf{u} \longmapsto \operatorname{sgn}(\langle \mathbf{w}, \mathbf{u} \rangle) \qquad \big(\mathbf{w}, \mathbf{u} \in \mathbb{R}^m\big),$$

где $\operatorname{sgn}(a) := 1$ для всех $a > 0$ и $\operatorname{sgn}(a) := 0$ для всех $a \leqslant 0$.

Покажем, что $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(m)\big) = m$.

Определим множество $E := \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, состоящее из всех базисных векторов в $\mathbb{R}^m$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4.

Пусть $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(m)$ множество всех бинарных классификаторов вида

$$\mathbf{u} \longmapsto \mathrm{sgn}\left(\langle \mathbf{w}, \mathbf{u} \rangle\right) \qquad \left(\mathbf{w}, \mathbf{u} \in \mathbb{R}^m\right),$$

где $\mathrm{sgn}(a) := 1$ для всех $a > 0$ и $\mathrm{sgn}(a) := 0$ для всех $a \leqslant 0$.

Покажем, что $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(m)\big) = m$.

Определим множество $E := \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, состоящее из всех базисных векторов в $\mathbb{R}^m$.

У каждого вектора $\mathbf{e}_j$ $(j = 1, \ldots, m)$ координата с номером j равна 1, а остальные координаты равны 0.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

Для каждого непустого подмножества индексов $J \subseteq \mathbb{N}_m$ положим

$$g_J(\mathbf{u}) := \operatorname{sgn}\left(\sum_{j \in J} \langle \mathbf{e}_j, \mathbf{u} \rangle\right) \qquad (\mathbf{u} \in \mathbb{R}^m).$$

Пример 5.4 (продолжение).

Для каждого непустого подмножества индексов $J \subseteq \mathbb{N}_m$ положим

$$g_J(\mathbf{u}) := \operatorname{sgn}\left(\sum_{j \in J} \langle \mathbf{e}_j, \mathbf{u} \rangle\right) \qquad (\mathbf{u} \in \mathbb{R}^m).$$

Дополнительно положим $g_{\varnothing}(\mathbf{u}) := \operatorname{sgn}(\langle \mathbf{0}, \mathbf{u} \rangle)$, где $\mathbf{0}$ – нулевой вектор в $\mathbb{R}^m$.

Пример 5.4 (продолжение).

Для каждого непустого подмножества индексов $J \subseteq \mathbb{N}_m$ положим

$$g_J(\mathbf{u}) := \operatorname{sgn}\left(\sum_{j \in J} \langle \mathbf{e}_j, \mathbf{u} \rangle\right) \qquad (\mathbf{u} \in \mathbb{R}^m).$$

Дополнительно положим $g_{\varnothing}(\mathbf{u}) := \operatorname{sgn}(\langle \mathbf{0}, \mathbf{u} \rangle)$, где $\mathbf{0}$ – нулевой вектор в $\mathbb{R}^m$.

Построенные классификаторы $\{g_J : J \subseteq \mathbb{N}_m\} \subset \mathcal{H}_{\mathrm{lin}}(m)$ разбивают множество E.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

Для каждого непустого подмножества индексов $J \subseteq \mathbb{N}_m$ положим

$$g_J(\mathbf{u}) := \operatorname{sgn}\left(\sum_{j \in J} \langle \mathbf{e}_j, \mathbf{u} \rangle\right) \qquad (\mathbf{u} \in \mathbb{R}^m).$$

Дополнительно положим $g_\varnothing(\mathbf{u}) := \operatorname{sgn}(\langle \mathbf{0}, \mathbf{u} \rangle)$, где $\mathbf{0}$ – нулевой вектор в $\mathbb{R}^m$.

Построенные классификаторы $\{g_J : J \subseteq \mathbb{N}_m\} \subset \mathcal{H}_{\mathrm{lin}}(m)$ разбивают множество E.

Поэтому $\mathrm{vc}(\mathcal{H}_{\mathrm{lin}}(m)) \geqslant m$.

Пример 5.4 (продолжение).

Зафиксируем произвольный набор, состоящий из $m+1$ различных векторов $\mathbf{u}_1, \ldots, \mathbf{u}_{m+1} \in \mathbb{R}^m$.

Пример 5.4 (продолжение).

Зафиксируем произвольный набор, состоящий из $m+1$ различных векторов $\mathbf{u}_1, \dots, \mathbf{u}_{m+1} \in \mathbb{R}^m$.

Очевидно, что эти вектора будут линейно зависимыми, а значит, существуют числа $a_1, \dots, a_{m+1} \in \mathbb{R}$, не все одновременно равные нулю, такие, что

$$\sum_{j=1}^{m+1} a_j \mathbf{u}_j = \mathbf{0}.$$

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

Зафиксируем произвольный набор, состоящий из $m+1$ различных векторов $\mathbf{u}_1, \ldots, \mathbf{u}_{m+1} \in \mathbb{R}^m$.

Очевидно, что эти вектора будут линейно зависимыми, а значит, существуют числа $a_1, \ldots, a_{m+1} \in \mathbb{R}$, не все одновременно равные нулю, такие, что

$$\sum_{j=1}^{m+1} a_j \mathbf{u}_j = \mathbf{0}.$$

Определим множество индексов $J_+ := \{j : a_j > 0\}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

Зафиксируем произвольный набор, состоящий из $m+1$ различных векторов $\mathbf{u}_1, \ldots, \mathbf{u}_{m+1} \in \mathbb{R}^m$.

Очевидно, что эти вектора будут линейно зависимыми, а значит, существуют числа $a_1, \ldots, a_{m+1} \in \mathbb{R}$, не все одновременно равные нулю, такие, что

$$\sum_{j=1}^{m+1} a_j \mathbf{u}_j = \mathbf{0}.$$

Определим множество индексов $J_+ := \{j : a_j > 0\}$.

Всегда можно предполагать, что $J_+ \neq \varnothing$.

Пример 5.4 (продолжение).

Зафиксируем произвольный набор, состоящий из $m+1$ различных векторов $\mathbf{u}_1, \ldots, \mathbf{u}_{m+1} \in \mathbb{R}^m$.

Очевидно, что эти вектора будут линейно зависимыми, а значит, существуют числа $a_1, \ldots, a_{m+1} \in \mathbb{R}$, не все одновременно равные нулю, такие, что

$$\sum_{j=1}^{m+1} a_j \mathbf{u}_j = \mathbf{0}.$$

Определим множество индексов $J_+ := \{j : a_j > 0\}$.

Всегда можно предполагать, что $J_+ \neq \varnothing$.

Если $J_+ = \varnothing$, то достаточно перейти к рассмотрению набора чисел $-a_1, \ldots, -a_{m+1}$.

Пример 5.4 (продолжение).

Имеет место равенство

$$\sum_{j \in J_+} a_j \mathbf{u}_j = \sum_{j \notin J_+} |a_j| \mathbf{u}_j.$$

Пример 5.4 (продолжение).

Имеет место равенство

$$\sum_{j \in J_+} a_j \mathbf{u}_j = \sum_{j \notin J_+} |a_j| \mathbf{u}_j.$$

Предположим, что класс $\mathcal{H}_{\text{lin}}(m)$ разбивает множество $\{\mathbf{u}_1, \ldots, \mathbf{u}_{m+1}\}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

Имеет место равенство

$$\sum_{j \in J_+} a_j \mathbf{u}_j = \sum_{j \notin J_+} |a_j| \mathbf{u}_j.$$

Предположим, что класс $\mathcal{H}_{\mathrm{lin}}(m)$ разбивает множество $\{\mathbf{u}_1, \ldots, \mathbf{u}_{m+1}\}$.

Это означает существование вектора $\mathbf{w} \in \mathbb{R}^m$ такого, что $\langle \mathbf{w}, \mathbf{u}_j \rangle > 0$ для всех $j \in J_+$ и $\langle \mathbf{w}, \mathbf{u} \rangle \leqslant 0$ для всех $j \notin J_+$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

Имеет место равенство

$$\sum_{j \in J_+} a_j \mathbf{u}_j = \sum_{j \notin J_+} |a_j| \mathbf{u}_j.$$

Предположим, что класс $\mathcal{H}_{\mathrm{lin}}(m)$ разбивает множество $\{\mathbf{u}_1, \ldots, \mathbf{u}_{m+1}\}$.

Это означает существование вектора $\mathbf{w} \in \mathbb{R}^m$ такого, что $\langle \mathbf{w}, \mathbf{u}_j \rangle > 0$ для всех $j \in J_+$ и $\langle \mathbf{w}, \mathbf{u} \rangle \leqslant 0$ для всех $j \notin J_+$.

Но тогда мы приходим к противоречию

$$0 < \sum_{j \in J_+} a_j \langle \mathbf{w}, \mathbf{u}_j \rangle = \sum_{j \notin J_+} |a_j| \langle \mathbf{w}, \mathbf{u}_j \rangle \leqslant 0.$$

Размерность Вапника-Червоненкиса

Примеры

Пример 5.4 (продолжение).

В виду произвольности выбора различных векторов $\mathbf{u}_1, \ldots, \mathbf{u}_{m+1}$ получим искомое равенство $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(m)\big) = m$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.5.

Рассмотрим произвольный конечный класс бинарных классификаторов $\mathcal{H}$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.5.

Рассмотрим произвольный конечный класс бинарных классификаторов $\mathcal{H}$.

Предположим, что он разбивает произвольное конечное подмножество $S \subseteq \mathrm{X}$.

Пример 5.5.

Рассмотрим произвольный конечный класс бинарных классификаторов $\mathcal{H}$.

Предположим, что он разбивает произвольное конечное подмножество $S \subseteq \mathrm{X}$.

Очевидно, что $|\mathcal{H}|_S| \leqslant |\mathcal{H}|$.

Размерность Вапника-Червоненкиса

Примеры

Пример 5.5.

Рассмотрим произвольный конечный класс бинарных классификаторов $\mathcal{H}$.

Предположим, что он разбивает произвольное конечное подмножество $S \subseteq \mathrm{X}$.

Очевидно, что $|\mathcal{H}|_S| \leqslant |\mathcal{H}|$.

Таким образом, S не может быть разбито, если $|\mathcal{H}| < 2^{|S|}$.

Пример 5.5.

Рассмотрим произвольный конечный класс бинарных классификаторов $\mathcal{H}$.

Предположим, что он разбивает произвольное конечное подмножество $S \subseteq \mathrm{X}$.

Очевидно, что $|\mathcal{H}|_S| \leqslant |\mathcal{H}|$.

Таким образом, S не может быть разбито, если $|\mathcal{H}| < 2^{|S|}$.

Следовательно, $\mathrm{vc}(\mathcal{H}) \leqslant \log_2 |\mathcal{H}|$.

Размерность Вапника-Червоненкиса

Примеры

Замечание

Заметим, что размерность Вапника-Червоненкиса конечного класса может существенно отличаться от своей верхней оценки.

Замечание

Заметим, что размерность Вапника-Червоненкиса конечного класса может существенно отличаться от своей верхней оценки.

В качестве множества объектов X возьмём $\mathbb{N}_m$ $(m \in \mathbb{N})$, а в качестве класса бинарных классификаторов $\mathcal{H}$ возьмём ограничение пороговых функций на это множество.

Размерность Вапника-Червоненкиса

Примеры

Замечание

Заметим, что размерность Вапника-Червоненкиса конечного класса может существенно отличаться от своей верхней оценки.

В качестве множества объектов X возьмём $\mathbb{N}_m$ $(m \in \mathbb{N})$, а в качестве класса бинарных классификаторов $\mathcal{H}$ возьмём ограничение пороговых функций на это множество.

Под пороговыми функциями понимаются характеристические функции концептов из примера 5.1.

Размерность Вапника-Червоненкиса

Примеры

Замечание

Заметим, что размерность Вапника-Червоненкиса конечного класса может существенно отличаться от своей верхней оценки.

В качестве множества объектов X возьмём $\mathbb{N}_m$ $(m \in \mathbb{N})$, а в качестве класса бинарных классификаторов $\mathcal{H}$ возьмём ограничение пороговых функций на это множество.

Под пороговыми функциями понимаются характеристические функции концептов из примера 5.1.

В этом случае $|\mathcal{H}| = m$ и $\mathrm{vc}(\mathcal{H}) = 1$.

Размерность Вапника-Червоненкиса

Примеры

Следующий пример в виду его особой важности оформим в виде теоремы.

Размерность Вапника-Червоненкиса

Примеры

Следующий пример в виду его особой важности оформим в виде теоремы.

Этот пример наряду с теоремой 3.1 является ещё одним теоретическим обоснованием практической применимости алгоритма адаптивного бустинга.

Размерность Вапника-Червоненкиса

Примеры

Следующий пример в виду его особой важности оформим в виде теоремы.

Этот пример наряду с теоремой 3.1 является ещё одним теоретическим обоснованием практической применимости алгоритма адаптивного бустинга.

Как мы это видели раньше, его отличительной особенностью является использование вспомогательного слабого учителя, который оперирует с гипотезами из некоторого исходного класса.

Следующий пример в виду его особой важности оформим в виде теоремы.

Этот пример наряду с теоремой 3.1 является ещё одним теоретическим обоснованием практической применимости алгоритма адаптивного бустинга.

Как мы это видели раньше, его отличительной особенностью является использование вспомогательного слабого учителя, который оперирует с гипотезами из некоторого исходного класса.

Сам же алгоритм адаптивного бустинга в процессе обучения строит гипотезу из расширенного класса.

Размерность Вапника-Червоненкиса

Примеры

Теорема, которую мы докажем, говорит о том, что если исходный класс гипотез обладал конечной размерностью Вапника-Червоненкиса, то конечной размерностью будет обладать и расширенный класс гипотез.

Размерность Вапника-Червоненкиса

Примеры

Теорема, которую мы докажем, говорит о том, что если исходный класс гипотез обладал конечной размерностью Вапника-Червоненкиса, то конечной размерностью будет обладать и расширенный класс гипотез.

При этом для последней будет доказана верхняя оценка.

Размерность Вапника-Червоненкиса

Примеры

Теорема, которую мы докажем, говорит о том, что если исходный класс гипотез обладал конечной размерностью Вапника-Червоненкиса, то конечной размерностью будет обладать и расширенный класс гипотез.

При этом для последней будет доказана верхняя оценка.

Прежде, чем непосредственно перейти к рассмотрению этой теоремы, докажем вспомогательное утверждение.

Утверждение 5.2.

Пусть $c \geqslant 2$ и $u \geqslant 2c\ln(c)$. Тогда $u \geqslant c\ln(u)$.

Утверждение 5.2.

Пусть $c \geqslant 2$ и $u \geqslant 2c\ln(c)$. Тогда $u \geqslant c\ln(u)$.

◀ Прежде всего покажем, что $c - 2\ln(c) > 0$ при всех $c \geqslant 2$.

Размерность Вапника-Червоненкиса

Примеры

Утверждение 5.2.

Пусть $c \geqslant 2$ и $u \geqslant 2c\ln(c)$. Тогда $u \geqslant c\ln(u)$.

◀ Прежде всего покажем, что $c - 2\ln(c) > 0$ при всех $c \geqslant 2$.

На самом деле, это неравенство конечно будет выполняться для всех $c > 0$, но нам достаточно ограничиться частным случаем.

Размерность Вапника-Червоненкиса

Примеры

Утверждение 5.2.

Пусть $c \geqslant 2$ и $u \geqslant 2c\ln(c)$. Тогда $u \geqslant c\ln(u)$.

◀ Прежде всего покажем, что $c - 2\ln(c) > 0$ при всех $c \geqslant 2$.

На самом деле, это неравенство конечно будет выполняться для всех $c > 0$, но нам достаточно ограничиться частным случаем.

Введём вспомогательную функцию $\psi(c) := c - 2\ln(c)$ и вычислим её производную $\psi'(c) = 1 - 2/c$.

Утверждение 5.2.

Пусть $c \geqslant 2$ и $u \geqslant 2c\ln(c)$. Тогда $u \geqslant c\ln(u)$.

◀ Прежде всего покажем, что $c - 2\ln(c) > 0$ при всех $c \geqslant 2$.

На самом деле, это неравенство конечно будет выполняться для всех $c > 0$, но нам достаточно ограничиться частным случаем.

Введём вспомогательную функцию $\psi(c) := c - 2\ln(c)$ и вычислим её производную $\psi'(c) = 1 - 2/c$.

Производная $\psi'(c) > 0$ при $c > 2$, а значит функция $\psi(c)$ строго возрастает на этом промежутке.

Утверждение 5.2.

Пусть $c \geqslant 2$ и $u \geqslant 2c\ln(c)$. Тогда $u \geqslant c\ln(u)$.

◀ Прежде всего покажем, что $c - 2\ln(c) > 0$ при всех $c \geqslant 2$.

На самом деле, это неравенство конечно будет выполняться для всех $c > 0$, но нам достаточно ограничиться частным случаем.

Введём вспомогательную функцию $\psi(c) := c - 2\ln(c)$ и вычислим её производную $\psi'(c) = 1 - 2/c$.

Производная $\psi'(c) > 0$ при $c > 2$, а значит функция $\psi(c)$ строго возрастает на этом промежутке.

Осталось только заметить, что $\psi(2) > 0$, так как $\ln(2) < 1$.

Размерность Вапника-Червоненкиса

Примеры

Введём вспомогательную функцию $\varphi(x) := x - c\ln(x)$ и вычислим её производную $\varphi'(x) = 1 - c/x$.

Размерность Вапника-Червоненкиса

Примеры

Введём вспомогательную функцию $\varphi(x) := x - c\ln(x)$ и вычислим её производную $\varphi'(x) = 1 - c/x$.

Производная $\varphi'(x) > 0$ при $x > c$, а значит функция $\varphi(x)$ строго возрастает на этом промежутке.

Размерность Вапника-Червоненкиса

Примеры

Введём вспомогательную функцию $\varphi(x) := x - c\ln(x)$ и вычислим её производную $\varphi'(x) = 1 - c/x$.

Производная $\varphi'(x) > 0$ при $x > c$, а значит функция $\varphi(x)$ строго возрастает на этом промежутке.

Осталось только проверить, что эта функция положительна в интересующей нас точке

Размерность Вапника-Червоненкиса

Примеры

Введём вспомогательную функцию $\varphi(x) := x - c\ln(x)$ и вычислим её производную $\varphi'(x) = 1 - c/x$.

Производная $\varphi'(x) > 0$ при $x > c$, а значит функция $\varphi(x)$ строго возрастает на этом промежутке.

Осталось только проверить, что эта функция положительна в интересующей нас точке

$$\begin{aligned}\varphi(2c\ln(c)) &= 2c\ln(c) - c\ln(2c\ln(c))\\ &= c\ln\left(\frac{c}{2\ln(c)}\right) > 0.\end{aligned}$$

■

Размерность Вапника-Червоненкиса

Примеры

Нам понадобится ещё одно утверждение, которое будет доказано чуть позже.

Нам понадобится ещё одно утверждение, которое будет доказано чуть позже.

Лемма 5.3 (Сауэра-Шелаха).

Пусть $\mathcal{C} \subseteq 2^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{C}) \leqslant d < \infty$.

Нам понадобится ещё одно утверждение, которое будет доказано чуть позже.

Лемма 5.3 (Сауэра-Шелаха).

Пусть $\mathcal{C} \subseteq 2^{Z}$ и $\mathrm{vc}(\mathcal{C}) \leqslant d < \infty$.

Тогда для любого $n \geqslant d$ выполняется неравенство

$$\Gamma_{\mathcal{C}}(n) \leqslant \left(\frac{e\,n}{d}\right)^{d}. \tag{2}$$

Размерность Вапника-Червоненкиса

Примеры

Теорема 5.1.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$ и $m \in \mathbb{N}$.

Теорема 5.1.

Пусть $\mathcal{H} \subseteq \{0, 1\}^{X}$ и $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\text{lin}}(\mathcal{H}, m)$ класс всех бинарных классификаторов вида $g(h_1, \ldots, h_m)$, где $g \in \mathcal{H}_{\text{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$.

Теорема 5.1.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$ и $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ класс всех бинарных классификаторов вида $g(h_1, \ldots, h_m)$, где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$.

Предположим, что $3 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$ и $m \geqslant 3$.

Размерность Вапника-Червоненкиса

Примеры

Теорема 5.1.

Пусть $\mathcal{H} \subseteq \{0, 1\}^X$ и $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ класс всех бинарных классификаторов вида $g(h_1, \ldots, h_m)$, где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$.

Предположим, что $3 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$ и $m \geqslant 3$.

Тогда

$$\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m\Big(\mathrm{vc}(\mathcal{H}) + 1\Big)\Big[3 \ln\big(m(\mathrm{vc}(\mathcal{H}) + 1)\big) + 2\Big]. \qquad (3)$$

Теорема 5.1.

Пусть $\mathcal{H} \subseteq \{0,1\}^X$ и $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ класс всех бинарных классификаторов вида $g(h_1, \ldots, h_m)$, где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$.

Предположим, что $3 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$ и $m \geqslant 3$.

Тогда

$$\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m\Big(\mathrm{vc}(\mathcal{H}) + 1\Big)\Big[3 \ln \big(m(\mathrm{vc}(\mathcal{H}) + 1)\big) + 2\Big]. \qquad (3)$$

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

Теорема 5.1.

Пусть $\mathcal{H} \subseteq \{0,1\}^X$ и $m \in \mathbb{N}$.

Обозначим через $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ класс всех бинарных классификаторов вида $g(h_1, \dots, h_m)$, где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \dots, h_m \in \mathcal{H}$.

Предположим, что $3 \leqslant \mathrm{vc}(\mathcal{H}) < \infty$ и $m \geqslant 3$.

Тогда

$$\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m\Big(\mathrm{vc}(\mathcal{H}) + 1\Big)\Big[3\ln\big(m(\mathrm{vc}(\mathcal{H}) + 1)\big) + 2\Big]. \qquad (3)$$

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

Из включения $\mathcal{H} \subseteq \mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ следует неравенство $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \geqslant d$.

Размерность Вапника-Червоненкиса

Примеры

Если $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m$, то неравенство (3) очевидно выполняется.

Размерность Вапника-Червоненкиса

Примеры

Если $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m$, то неравенство (3) очевидно выполняется.

Поэтому далее, будем предполагать, что $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \geqslant m$.

Если $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m$, то неравенство (3) очевидно выполняется.

Поэтому далее, будем предполагать, что $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \geqslant m$.

Зафиксируем произвольное конечное подмножество $S \subseteq \mathrm{X}$ такое, что $|S| \geqslant \max\{d, m\}$.

Размерность Вапника-Червоненкиса

Примеры

Если $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m$, то неравенство (3) очевидно выполняется.

Поэтому далее, будем предполагать, что $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \geqslant m$.

Зафиксируем произвольное конечное подмножество $S \subseteq \mathrm{X}$ такое, что $|S| \geqslant \max\{d, m\}$.

Предположим, что класс $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ разбивает подмножество S.

Размерность Вапника-Червоненкиса

Примеры

Если $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \leqslant m$, то неравенство (3) очевидно выполняется.

Поэтому далее, будем предполагать, что $\mathrm{vc}\big(\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)\big) \geqslant m$.

Зафиксируем произвольное конечное подмножество $S \subseteq \mathrm{X}$ такое, что $|S| \geqslant \max\{d, m\}$.

Предположим, что класс $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)$ разбивает подмножество S.

Это означает, что

$$\big|\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)|_S\big| = 2^{|S|}. \tag{4}$$

Размерность Вапника-Червоненкиса

Примеры

Каждый элемент класса $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)|_S$ может быть представлен в виде

$$g(h_1|_S, \ldots, h_m|_S) = g|_{V(h_1,\ldots,h_m)}(h_1|_S, \ldots, h_m|_S),$$

Размерность Вапника-Червоненкиса

Примеры

Каждый элемент класса $\mathcal{H}_{\text{lin}}(\mathcal{H}, m)|_S$ может быть представлен в виде

$$g(h_1|_S, \ldots, h_m|_S) = g|_{V(h_1,\ldots,h_m)}(h_1|_S, \ldots, h_m|_S),$$

где $g \in \mathcal{H}_{\text{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$ и
$V(h_1, \ldots, h_m) := \{(h_1(x), \ldots, h_m(x)) : x \in S\}$.

Размерность Вапника-Червоненкиса

Примеры

Каждый элемент класса $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)|_S$ может быть представлен в виде

$$g(h_1|_S, \ldots, h_m|_S) = g|_{V(h_1,\ldots,h_m)}(h_1|_S, \ldots, h_m|_S),$$

где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$ и
$V(h_1, \ldots, h_m) := \{(h_1(x), \ldots, h_m(x)) : x \in S\}$.

Очевидно, что $|V(h_1, \ldots, h_m)| \leqslant |S|$.

Размерность Вапника-Червоненкиса

Примеры

Каждый элемент класса $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)|_S$ может быть представлен в виде

$$g(h_1|_S, \ldots, h_m|_S) = g|_{V(h_1,\ldots,h_m)}(h_1|_S, \ldots, h_m|_S),$$

где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$ и
$V(h_1, \ldots, h_m) := \{(h_1(x), \ldots, h_m(x)) : x \in S\}$.

Очевидно, что $|V(h_1, \ldots, h_m)| \leqslant |S|$.

Воспользуемся леммой Сауэра-Шелаха, которая будет доказана позже.

Размерность Вапника-Червоненкиса

Примеры

Каждый элемент класса $\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)|_S$ может быть представлен в виде

$$g(h_1|_S, \ldots, h_m|_S) = g|_{V(h_1,\ldots,h_m)}(h_1|_S, \ldots, h_m|_S),$$

где $g \in \mathcal{H}_{\mathrm{lin}}(m)$, $h_1, \ldots, h_m \in \mathcal{H}$ и
$V(h_1, \ldots, h_m) := \{(h_1(x), \ldots, h_m(x)) : x \in S\}$.

Очевидно, что $|V(h_1, \ldots, h_m)| \leqslant |S|$.

Воспользуемся леммой Сауэра-Шелаха, которая будет доказана позже. Учитывая предположение $d, m \geqslant 3$, запишем

$$\left|\mathcal{H}_{\mathrm{lin}}(\mathcal{H}, m)|_S\right| \leqslant \left(\Gamma_{\mathcal{H}}(|S|)\right)^m \Gamma_{\mathcal{H}_{\mathrm{lin}}(m)}(|S|) \leqslant \left(\frac{e\,|S|}{d}\right)^{dm} \left(\frac{e\,|S|}{m}\right)^m$$
$$\leqslant |S|^{m(d+1)}.$$

Размерность Вапника-Червоненкиса

Примеры

Из (4) следует, что $2^{|S|} \leqslant |S|^{m(d+1)}$.

Размерность Вапника-Червоненкиса

Примеры

Из (4) следует, что $2^{|S|} \leqslant |S|^{m(d+1)}$.

Следовательно,

$$|S| \leqslant m(d+1)\log_2|S| = \frac{m(d+1)}{\ln 2}\ln|S|. \tag{5}$$

Размерность Вапника-Червоненкиса

Примеры

Из (4) следует, что $2^{|S|} \leqslant |S|^{m(d+1)}$.

Следовательно,

$$|S| \leqslant m(d+1)\log_2|S| = \frac{m(d+1)}{\ln 2}\ln|S|. \tag{5}$$

Заметим, что

$$\frac{m(d+1)}{\ln 2} \geqslant 2, \quad \frac{2}{\ln 2} \leqslant 3, \quad 3\left(\ln\frac{1}{\ln 2}\right) \leqslant 2.$$

Размерность Вапника-Червоненкиса

Примеры

Применяя к (5) утв. 5.2 ($c = \frac{m(d+1)}{\ln 2}$, $u = |S|$), получим

$$|S| \leqslant \frac{2m(d+1)}{\ln 2} \ln\left(\frac{m(d+1)}{\ln 2}\right) \leqslant m(d+1)\left[3\ln\left(m(d+1)\right) + 2\right].$$

Из произвольности выбора подмножества S следует справедливость оценки (3).

$\blacksquare$

Содержание

Утверждение 5.3.

Пусть $n, d \in \mathbb{N}$. Обозначим

$$\binom{n}{\leqslant d} := \sum_{i=0}^{d} \binom{n}{i}, \tag{6}$$

где формально считаем $\binom{n}{i} = 0$ при $i > n$.

Утверждение 5.3.

Пусть $n, d \in \mathbb{N}$. Обозначим

$$\binom{n}{\leqslant d} := \sum_{i=0}^{d} \binom{n}{i}, \tag{6}$$

где формально считаем $\binom{n}{i} = 0$ при $i > n$.

Тогда

$$\binom{n}{\leqslant d} \leqslant (n+1)^d. \tag{7}$$

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Утверждение 5.3.

Пусть $n, d \in \mathbb{N}$. Обозначим

$$\binom{n}{\leqslant d} := \sum_{i=0}^{d} \binom{n}{i}, \tag{6}$$

где формально считаем $\binom{n}{i} = 0$ при $i > n$.

Тогда

$$\binom{n}{\leqslant d} \leqslant (n+1)^d. \tag{7}$$

Кроме того, при $n \geqslant d$ выполняется неравенство

$$\binom{n}{\leqslant d} \leqslant \left(\frac{e\,n}{d}\right)^d.$$

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Для доказательсва первого неравенства достаточно заметить, что $\binom{n}{\leqslant d}$ равно числу подмножеств n-элементного множества, которые состоят из не более чем d элементов.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Для доказательсва первого неравенства достаточно заметить, что $\binom{n}{\leqslant d}$ равно числу подмножеств n-элементного множества, которые состоят из не более чем d элементов.

Эта величина ограничена сверху числом упорядоченных выборок размера d из $n+1$-элементного множества.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Для доказательса первого неравенства достаточно заметить, что $\binom{n}{\leqslant d}$ равно числу подмножеств n-элементного множества, которые состоят из не более чем d элементов.

Эта величина ограничена сверху числом упорядоченных выборок размера d из $n+1$-элементного множества.

Используя неравенство $1 + x \leqslant e^x$ $(x \in \mathbb{R})$, получим второе неравенство

$$\left(\frac{d}{n}\right)^d \sum_{i=0}^{d} \binom{n}{i} \leqslant \sum_{i=0}^{d} \binom{n}{i} \left(\frac{d}{n}\right)^i \leqslant \sum_{i=0}^{n} \binom{n}{i} \left(\frac{d}{n}\right)^i = \left(1 + \frac{d}{n}\right)^n \leqslant e^d.$$

■

Утверждение 5.4.

Пусть $S \subset Z$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^Z$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Утверждение 5.4.

Пусть $S \subset \mathsf{Z}$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^{\mathsf{Z}}$.

Тогда

$$|\mathcal{C}(S)| \leqslant |\{S' \subseteq S : S' \text{ разбивается } \mathcal{C}\}|. \tag{8}$$

Утверждение 5.4.

Пусть $S \subset Z$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^Z$.

Тогда
$$|\mathcal{C}(S)| \leqslant |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C}\}|. \tag{8}$$

◀ Доказательство проведём индукцией по параметру $|S|$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Утверждение 5.4.

Пусть $S \subset Z$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^Z$.

Тогда
$$|\mathcal{C}(S)| \leqslant |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C}\}|. \tag{8}$$

◀ Доказательство проведём индукцией по параметру $|S|$.

Основание индукции, $|S| = 1$.

Утверждение 5.4.

Пусть $S \subset Z$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^Z$.

Тогда

$$|\mathcal{C}(S)| \leqslant |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C}\}|. \tag{8}$$

◀ Доказательство проведём индукцией по параметру $|S|$.

Основание индукции, $|S| = 1$.

В этом случае левая и правая части (8) либо одновременно равны 1, либо одновременно равны 2.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Утверждение 5.4.

Пусть $S \subset Z$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^{Z}$.

Тогда
$$|\mathcal{C}(S)| \leqslant |\{S' \subseteq S : S' \text{ разбивается } \mathcal{C}\}|. \tag{8}$$

◀ Доказательство проведём индукцией по параметру $|S|$.

Основание индукции, $|S| = 1$.

В этом случае левая и правая части (8) либо одновременно равны 1, либо одновременно равны 2.

Если $|\mathcal{C}(S)| = 1$, то это означает, что $\mathcal{C}$ разбивает только пустое подмножество.

Утверждение 5.4.

Пусть $S \subset Z$, $0 < |S| < \infty$ и $\mathcal{C} \subseteq 2^Z$.

Тогда
$$|\mathcal{C}(S)| \leqslant |\{S' \subseteq S : S' \text{ разбивается } \mathcal{C}\}|. \tag{8}$$

◀ Доказательство проведём индукцией по параметру $|S|$.

Основание индукции, $|S| = 1$.

В этом случае левая и правая части (8) либо одновременно равны 1, либо одновременно равны 2.

Если $|\mathcal{C}(S)| = 1$, то это означает, что $\mathcal{C}$ разбивает только пустое подмножество.

Если $|\mathcal{C}(S)| = 2$, то это означает, что $\mathcal{C}$ разбивает ещё и S.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Индуктивный переход.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Индуктивный переход.

Предположим, что утверждение доказано для всех $|S| < n$ $(n \in \mathbb{N}, n \geqslant 2)$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Индуктивный переход.

Предположим, что утверждение доказано для всех $|S| < n$ ($n \in \mathbb{N}, n \geqslant 2$).

Покажем, что оно остаётся верным и для $|S| = n$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Индуктивный переход.

Предположим, что утверждение доказано для всех $|S| < n$ ($n \in \mathbb{N}, n \geqslant 2$).

Покажем, что оно остаётся верным и для $|S| = n$.

Пусть $S = \{z_1, \ldots, z_n\}$ и $\hat{S} = \{z_2, \ldots, z_n\}$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Индуктивный переход.

Предположим, что утверждение доказано для всех $|S| < n$ $(n \in \mathbb{N}, n \geqslant 2)$.

Покажем, что оно остаётся верным и для $|S| = n$.

Пусть $S = \{z_1, \ldots, z_n\}$ и $\hat{S} = \{z_2, \ldots, z_n\}$.

Определим два множества

$$B_0 := \big\{(b_2, \ldots, b_n) \,:\, (0, b_2, \ldots, b_n) \in \mathcal{C}(S) \vee (1, b_2, \ldots, b_n) \in \mathcal{C}(S)\big\},$$
$$B_1 := \big\{(b_2, \ldots, b_n) \,:\, (0, b_2, \ldots, b_n) \in \mathcal{C}(S) \wedge (1, b_2, \ldots, b_n) \in \mathcal{C}(S)\big\}.$$

Индуктивный переход.

Предположим, что утверждение доказано для всех $|S| < n$ $(n \in \mathbb{N}, n \geqslant 2)$.

Покажем, что оно остаётся верным и для $|S| = n$.

Пусть $S = \{z_1, \dots, z_n\}$ и $\hat{S} = \{z_2, \dots, z_n\}$.

Определим два множества

$$\begin{aligned} B_0 &:= \{(b_2, \dots, b_n) \;:\; (0, b_2, \dots, b_n) \in \mathcal{C}(S) \vee (1, b_2, \dots, b_n) \in \mathcal{C}(S)\}, \\ B_1 &:= \{(b_2, \dots, b_n) \;:\; (0, b_2, \dots, b_n) \in \mathcal{C}(S) \wedge (1, b_2, \dots, b_n) \in \mathcal{C}(S)\}. \end{aligned}$$

Заметим, что

$$|\mathcal{C}(S)| = |B_0| + |B_1|. \tag{9}$$

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Применяя индуктивное предположение, запишем

$$\begin{aligned} |B_0| = |\mathcal{C}(\hat{S})| \leqslant |\{S' \subseteq \hat{S} \,:\, S' \text{ разбивается } \mathcal{C}\}| \\ = |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C} \wedge z_1 \notin S'\}|. \end{aligned} \tag{10}$$

Применяя индуктивное предположение, запишем

$$
\begin{aligned}
|B_0| = |\mathcal{C}(\hat{S})| \leqslant |\{S' \subseteq \hat{S} \,:\, S' \text{ разбивается } \mathcal{C}\}| \\
= |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C} \wedge z_1 \notin S'\}|.
\end{aligned}
\tag{10}
$$

Определим

$$
\begin{aligned}
\hat{\mathcal{C}} := \big\{C \in \mathcal{C} \,:\, \exists C' \in \mathcal{C} \,:\, \mathbf{1}_C(z_1) = 1 - \mathbf{1}_{C'}(z_1), \\
\mathbf{1}_C(z_i) = \mathbf{1}_{C'}(z_i) \quad (i = 2, \ldots, n)\big\}.
\end{aligned}
$$

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Применяя индуктивное предположение, запишем

$$\begin{aligned} |B_0| = |\mathcal{C}(\hat{S})| \leqslant |\{S' \subseteq \hat{S} : S' \text{ разбивается } \mathcal{C}\}| \\ = |\{S' \subseteq S : S' \text{ разбивается } \mathcal{C} \wedge z_1 \notin S'\}|. \end{aligned} \tag{10}$$

Определим

$$\begin{aligned} \hat{\mathcal{C}} := \big\{C \in \mathcal{C} : \exists C' \in \mathcal{C} : \mathbf{1}_C(z_1) = 1 - \mathbf{1}_{C'}(z_1), \\ \mathbf{1}_C(z_i) = \mathbf{1}_{C'}(z_i) \quad (i = 2, \dots, n)\big\}. \end{aligned}$$

Из этого определения видно, что подмножество $S' \subseteq \hat{S}$ разбивается классом множеств $\hat{\mathcal{C}}$ тогда и только тогда, когда этот класс разбивает множество $S' \cup \{z_1\}$.

Применяя индуктивное предположение, запишем

$$
\begin{aligned}
|B_1| = |\hat{\mathcal{C}}(\hat{S})| &\leqslant |\{S' \subseteq \hat{S} \,:\, S' \text{ разбивается } \hat{\mathcal{C}}\}| \\
&= |\{S' \subseteq S \,:\, S' \text{ разбивается } \hat{\mathcal{C}} \wedge z_1 \in S'\}| \\
&\leqslant |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C} \wedge z_1 \in S'\}|.
\end{aligned}
\tag{11}
$$

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Применяя индуктивное предположение, запишем

$$
\begin{aligned}
|B_1| = |\hat{\mathcal{C}}(\hat{S})| &\leqslant |\{S' \subseteq \hat{S} \,:\, S' \text{ разбивается } \hat{\mathcal{C}}\}| \\
&= |\{S' \subseteq S \,:\, S' \text{ разбивается } \hat{\mathcal{C}} \wedge z_1 \in S'\}| \\
&\leqslant |\{S' \subseteq S \,:\, S' \text{ разбивается } \mathcal{C} \wedge z_1 \in S'\}|.
\end{aligned} \tag{11}
$$

Объединяя вместе (9), (10) и (11), получим требуемое неравенство (8).

■

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

Лемма 5.3 (Сауэра-Шелаха).

Пусть $\mathcal{C} \subseteq 2^{Z}$ и $\mathrm{vc}(\mathcal{C}) \leqslant d < \infty$.

Лемма 5.3 (Сауэра-Шелаха).

Пусть $\mathcal{C} \subseteq 2^{Z}$ и $\mathrm{vc}(\mathcal{C}) \leqslant d < \infty$.

Тогда для любого $n \in \mathbb{N}$ выполняется неравенство

$$\Gamma_{\mathcal{C}}(n) \leqslant \binom{n}{\leqslant d}, \tag{12}$$

а значит

$$\Gamma_{\mathcal{C}}(n) \leqslant (n+1)^{d}. \tag{13}$$

Лемма 5.3 (Сауэра-Шелаха).

Пусть $\mathcal{C} \subseteq 2^{\mathsf{Z}}$ и $\mathrm{vc}(\mathcal{C}) \leqslant d < \infty$.

Тогда для любого $n \in \mathbb{N}$ выполняется неравенство

$$\Gamma_{\mathcal{C}}(n) \leqslant \binom{n}{\leqslant d}, \tag{12}$$

а значит

$$\Gamma_{\mathcal{C}}(n) \leqslant (n+1)^d. \tag{13}$$

Кроме того, при $n \geqslant d$ выполняется неравенство

$$\Gamma_{\mathcal{C}}(n) \leqslant \left(\frac{e\,n}{d}\right)^d. \tag{14}$$

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Прежде всего заметим, что достаточно установить неравенство (12).

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Прежде всего заметим, что достаточно установить неравенство (12).

После этого оставшиеся неравенства (13) и (14) будут следовать из утв. 5.3.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Прежде всего заметим, что достаточно установить неравенство (12).

После этого оставшиеся неравенства (13) и (14) будут следовать из утв. 5.3.

Возьмём произвольное n-элементное множество $S \subset Z$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Прежде всего заметим, что достаточно установить неравенство (12).

После этого оставшиеся неравенства (13) и (14) будут следовать из утв. 5.3.

Возьмём произвольное n-элементное множество $S \subset \mathrm{Z}$.

По определению размерности Вапника-Червоненкиса не существует подмножества $S' \subseteq S$, которое разбивается классом $\mathcal{C}$ и содержит больше чем d элементов.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Прежде всего заметим, что достаточно установить неравенство (12).

После этого оставшиеся неравенства (13) и (14) будут следовать из утв. 5.3.

Возьмём произвольное n-элементное множество $S \subset \mathrm{Z}$.

По определению размерности Вапника-Червоненкиса не существует подмножества $S' \subseteq S$, которое разбивается классом $\mathcal{C}$ и содержит больше чем d элементов.

Поэтому число подмножеств множества S, которые разбиваются классом $\mathcal{C}$, не превосходит величины $\binom{n}{\leqslant d}$.

Размерность Вапника-Червоненкиса

Лемма Сауэра-Шелаха

◀ Прежде всего заметим, что достаточно установить неравенство (12).

После этого оставшиеся неравенства (13) и (14) будут следовать из утв. 5.3.

Возьмём произвольное n-элементное множество $S \subset \mathrm{Z}$.

По определению размерности Вапника-Червоненкиса не существует подмножества $S' \subseteq S$, которое разбивается классом $\mathcal{C}$ и содержит больше чем d элементов.

Поэтому число подмножеств множества S, которые разбиваются классом $\mathcal{C}$, не превосходит величины $\binom{n}{\leqslant d}$.

Применяя утв. 5.4 к определению функции роста, получим искомое неравенство (12).

■

Содержание

Фундаментальная теорема

Сначала будет дана формулировка этой теоремы и разобрана схема её доказательства.

Фундаментальная теорема

Сначала будет дана формулировка этой теоремы и разобрана схема её доказательства.

Далее, в виде отдельных утверждений будут рассмотрены ключевые шаги этого доказательства.

Содержание

Теорема 5.2.

Предположим, что

- $\mathcal{H} \subseteq \{0, 1\}^{X}$;

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

Предположим, что

- $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$;
- семейство $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}_{+}^{1}(\mathsf{X})$ образуют все вероятностные меры P_X такие, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой;

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

Предположим, что

- $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$;
- семейство $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ образуют все вероятностные меры P_X такие, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой;
- семейство $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, где $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$, является P-измеримой.

Теорема 5.2.

Предположим, что

- $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$;
- семейство $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ образуют все вероятностные меры P_X такие, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой;
- семейство $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, где $\mathcal{H} \simeq_{l_{01}} \mathcal{F}$, является P-измеримой.

Тогда следующие утверждения эквивалентны:

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

1. класс $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства $\mathcal{P}$ и фунции потерь l_{01};

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

1. класс $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
2. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01} с помощью метода минимизации эмпирического риска;

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

1. класс $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
2. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01} с помощью метода минимизации эмпирического риска;
3. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01};

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

1. класс $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
2. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01} с помощью метода минимизации эмпирического риска;
3. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
4. класс $\mathcal{H}$ является PAC-обучаемым относительно семейства $\mathcal{P}_X$ с помощью метода минимизации эмпирического риска;

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

1. класс $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
2. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01} с помощью метода минимизации эмпирического риска;
3. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
4. класс $\mathcal{H}$ является PAC-обучаемым относительно семейства $\mathcal{P}_X$ с помощью метода минимизации эмпирического риска;
5. класс $\mathcal{H}$ является PAC-обучаемым относительно семейства $\mathcal{P}_X$;

Фундаментальная теорема

Формулировка теоремы и схема доказательства

Теорема 5.2.

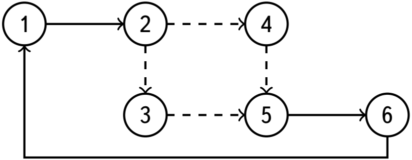
1. класс $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
2. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01} с помощью метода минимизации эмпирического риска;
3. класс $\mathcal{H}$ является агностически PAC-обучаемым относительно семейства $\mathcal{P}$ и фунции потерь l_{01};
4. класс $\mathcal{H}$ является PAC-обучаемым относительно семейства $\mathcal{P}_X$ с помощью метода минимизации эмпирического риска;
5. класс $\mathcal{H}$ является PAC-обучаемым относительно семейства $\mathcal{P}_X$;
6. $\mathrm{vc}(\mathcal{H}) < \infty$.

Фундаментальная теорема

Формулировка теоремы и схема доказательства

◀ (схема доказательства).
На следующем рисунке графически изображена схема доказательства теоремы.

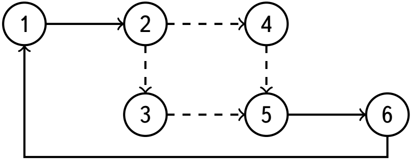
◄ (схема доказательства).
На следующем рисунке графически изображена схема доказательства теоремы.



Фундаментальная теорема

Формулировка теоремы и схема доказательства

◀ (схема доказательства).
На следующем рисунке графически изображена схема доказательства теоремы.

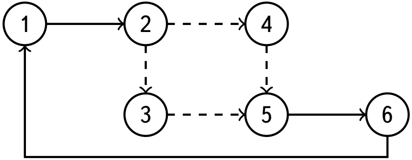


Каждый кружок обозначает утверждение этой теоремы с соответствующим номером.

Фундаментальная теорема

Формулировка теоремы и схема доказательства

◀ (схема доказательства).
На следующем рисунке графически изображена схема доказательства теоремы.

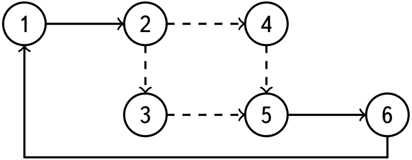


Каждый кружок обозначает утверждение этой теоремы с соответствующим номером.

Стрелками обозначены импликации из одного утверждения теоремы в другое.

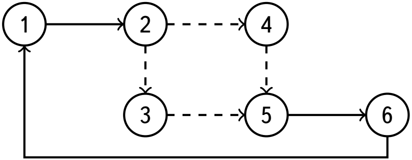
Фундаментальная теорема

Формулировка теоремы и схема доказательства



Фундаментальная теорема

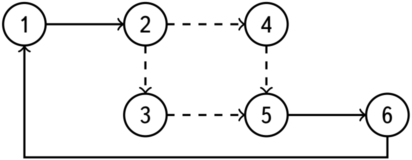
Формулировка теоремы и схема доказательства



Импликации $2 \Rightarrow 3$, $2 \Rightarrow 4$, $3 \Rightarrow 5$ и $4 \Rightarrow 5$ носят очевидный характер.

Фундаментальная теорема

Формулировка теоремы и схема доказательства

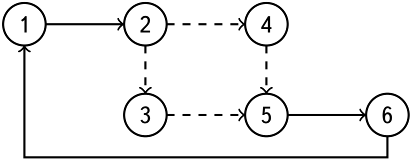


Импликации $2 \Rightarrow 3$, $2 \Rightarrow 4$, $3 \Rightarrow 5$ и $4 \Rightarrow 5$ носят очевидный характер.

Импликация $1 \Rightarrow 2$ уже была доказана в виде теоремы 4.6.

Фундаментальная теорема

Формулировка теоремы и схема доказательства



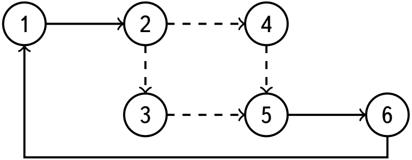
Импликации $2 \Rightarrow 3$, $2 \Rightarrow 4$, $3 \Rightarrow 5$ и $4 \Rightarrow 5$ носят очевидный характер.

Импликация $1 \Rightarrow 2$ уже была доказана в виде теоремы 4.6.

Импликация $5 \Rightarrow 6$ будет доказана как теорема 5.3. Её доказательство базируется на использовании теоремы «no free lunch».

Фундаментальная теорема

Формулировка теоремы и схема доказательства



Импликации $2 \Rightarrow 3$, $2 \Rightarrow 4$, $3 \Rightarrow 5$ и $4 \Rightarrow 5$ носят очевидный характер.

Импликация $1 \Rightarrow 2$ уже была доказана в виде теоремы 4.6.

Импликация $5 \Rightarrow 6$ будет доказана как теорема 5.3. Её доказательство базируется на использовании теоремы «no free lunch».

Наиболее трудоёмкой частью доказательства является проверка импликации $6 \Rightarrow 1$. Она будет установлена в виде теоремы 5.5.

■

Содержание

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Теорема 5.3.

Предположим, что класс гипотез $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ является PAC-обучаемым относительно семейства $\mathcal{P}_{\mathsf{X}}$, составленного из всех вероятностных мер P_X таких, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой.

Теорема 5.3.

Предположим, что класс гипотез $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ является PAC-обучаемым относительно семейства $\mathcal{P}_{\mathsf{X}}$, составленного из всех вероятностных мер P_X таких, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой.

Тогда $\mathrm{vc}(\mathcal{H}) < \infty$.

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Теорема 5.3.

Предположим, что класс гипотез $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ является PAC-обучаемым относительно семейства $\mathcal{P}_{\mathsf{X}}$, составленного из всех вероятностных мер P_X таких, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой.

Тогда $\mathrm{vc}(\mathcal{H}) < \infty$.

◀ Доказательство проведём от противного.

Фундаментальная теорема

Импликация 5 $\Rightarrow$ 6

Теорема 5.3.

Предположим, что класс гипотез $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ является PAC-обучаемым относительно семейства $\mathcal{P}_{\mathsf{X}}$, составленного из всех вероятностных мер P_{X} таких, что любая гипотеза из $\mathcal{H}$ является P_{X}-измеримой.

Тогда $\mathrm{vc}(\mathcal{H}) < \infty$.

◀ Доказательство проведём от противного.

Предположим, что $\mathrm{vc}(\mathcal{H}) = \infty$ и одновременно для класса $\mathcal{H}$ существует PAC-учитель $\mathcal{A}$.

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Теорема 5.3.

Предположим, что класс гипотез $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ является PAC-обучаемым относительно семейства $\mathcal{P}_{\mathsf{X}}$, составленного из всех вероятностных мер P_X таких, что любая гипотеза из $\mathcal{H}$ является P_X-измеримой.

Тогда $\mathrm{vc}(\mathcal{H}) < \infty$.

◀ Доказательство проведём от противного.

Предположим, что $\mathrm{vc}(\mathcal{H}) = \infty$ и одновременно для класса $\mathcal{H}$ существует PAC-учитель $\mathcal{A}$.

Нашей целью будет доказательство неравенства

$$r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, 1/8) \geqslant 1/7 \qquad (n \in \mathbb{N}). \tag{15}$$

Фундаментальная теорема
Импликация 5 ⇒ 6

Из этого неравенства следует, что

$$\lim_{n\to\infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, 1/8) \neq 0,$$

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Из этого неравенства следует, что

$$\lim_{n\to\infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, 1/8) \neq 0,$$

а это противоречит определению PAC-обучаемости.

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Из этого неравенства следует, что

$$\lim_{n\to\infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, 1/8) \neq 0,$$

а это противоречит определению PAC-обучаемости.

Зафиксируем произвольное $n \in \mathbb{N}$.

Из этого неравенства следует, что

$$\lim_{n \to \infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, 1/8) \neq 0,$$

а это противоречит определению PAC-обучаемости.

Зафиксируем произвольное $n \in \mathbb{N}$.

В силу сделанного предположения существует подмножество $\widehat{\mathsf{X}} \subset \mathsf{X}$, $|\widehat{\mathsf{X}}| = 2n$, которое разбивается классом $\mathcal{H}$.

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Из этого неравенства следует, что

$$\lim_{n\to\infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, 1/8) \neq 0,$$

а это противоречит определению PAC-обучаемости.

Зафиксируем произвольное $n \in \mathbb{N}$.

В силу сделанного предположения существует подмножество $\widehat{\mathsf{X}} \subset \mathsf{X}$, $|\widehat{\mathsf{X}}| = 2n$, которое разбивается классом $\mathcal{H}$.

Это означает, что

$$\widehat{\mathcal{H}} := \mathcal{H}|_{\widehat{\mathsf{X}}} = \{0,1\}^{\widehat{\mathsf{X}}}.$$

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

В силу теоремы «no free lunch» существует гипотеза $g \in \mathcal{H}$ и вероятностная мера $\mathsf{P}_{\widehat{X}} \in \mathcal{M}_+^1\big(\widehat{\mathsf{X}}, 2^{\widehat{\mathsf{X}}}\big)$ такие, что

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

В силу теоремы «no free lunch» существует гипотеза $g \in \mathcal{H}$ и вероятностная мера $\mathsf{P}_{\widehat{X}} \in \mathcal{M}_+^1(\widehat{\mathsf{X}}, 2^{\widehat{\mathsf{X}}})$ такие, что

$$\mathsf{P}_{\widehat{X}}^n\left\{\mathbf{x} \in \widehat{\mathsf{X}}^n \,:\, h = \mathcal{A}(g \circ \mathbf{x}), R\big(\mathsf{P}_{\widehat{X}}, l_{01}; g|_{\widehat{\mathsf{X}}}, h|_{\widehat{\mathsf{X}}}\big) > 1/8\right\} \geqslant 1/7. \qquad (16)$$

Фундаментальная теорема

Импликация 5 ⇒ 6

В силу теоремы «no free lunch» существует гипотеза $g \in \mathcal{H}$ и вероятностная мера $\mathrm{P}_{\widehat{X}} \in \mathcal{M}_+^1(\widehat{\mathrm{X}}, 2^{\widehat{\mathrm{X}}})$ такие, что

$$\mathrm{P}_{\widehat{X}}^n\left\{\mathbf{x} \in \widehat{\mathrm{X}}^n \,:\, h = \mathcal{A}(g \circ \mathbf{x}), R\left(\mathrm{P}_{\widehat{X}}, l_{01}; g|_{\widehat{\mathrm{X}}}, h|_{\widehat{\mathrm{X}}}\right) > 1/8\right\} \geqslant 1/7. \qquad (16)$$

Определим вероятностную меру $\mathrm{P}_X \in \mathcal{P}_{\mathrm{X}}$ по правилу

$$\mathrm{P}_X := \mathrm{P}_{\widehat{X}}(B \cap \widehat{\mathrm{X}}) \qquad \left(B \in 2^{\mathrm{X}}\right),$$

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

В силу теоремы «no free lunch» существует гипотеза $g \in \mathcal{H}$ и вероятностная мера $\mathrm{P}_{\widehat{X}} \in \mathcal{M}^1_+(\widehat{\mathrm{X}}, 2^{\widehat{\mathrm{X}}})$ такие, что

$$\mathrm{P}^n_{\widehat{X}}\left\{\mathbf{x} \in \widehat{\mathrm{X}}^n \;:\; h = \mathcal{A}(g \circ \mathbf{x}), R\big(\mathrm{P}_{\widehat{X}}, l_{01}; g|_{\widehat{\mathrm{X}}}, h|_{\widehat{\mathrm{X}}}\big) > 1/8\right\} \geqslant 1/7. \qquad (16)$$

Определим вероятностную меру $\mathrm{P}_X \in \mathcal{P}_{\mathrm{X}}$ по правилу

$$\mathrm{P}_X := \mathrm{P}_{\widehat{X}}\big(B \cap \widehat{\mathrm{X}}\big) \qquad \big(B \in 2^{\mathrm{X}}\big),$$

тогда

$$\begin{aligned} &\mathrm{P}^n_X\left\{\mathbf{x} \in \mathrm{X}^n \;:\; h = \mathcal{A}(g \circ \mathbf{x}), R\big(\mathrm{P}_X, l_{01}; g, h\big) > 1/8\right\} \geqslant \\ &\mathrm{P}^n_{\widehat{X}}\left\{\mathbf{x} \in \widehat{\mathrm{X}}^n \;:\; h = \mathcal{A}(g \circ \mathbf{x}), R\big(\mathrm{P}_{\widehat{X}}, l_{01}; g|_{\widehat{\mathrm{X}}}, h|_{\widehat{\mathrm{X}}}\big) > 1/8\right\}. \end{aligned} \qquad (17)$$

Фундаментальная теорема

Импликация $5 \Rightarrow 6$

Объединяя вместе (16) и (17), а также, учитывая определение функции $r_{\mathcal{A}}$, получим искомое неравенство (15).

$\blacksquare$

Содержание

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Определение 5.6.

Случайная величина, заданная на некотором вероятностном пространстве, называется радемахеровской, если она принимает только два значения -1 и 1. При этом, каждое из этих значений она принимает с вероятностью $\frac{1}{2}$.

Определение 5.6.

Случайная величина, заданная на некотором вероятностном пространстве, называется **радемахеровской**, если она принимает только два значения -1 и 1. При этом, каждое из этих значений она принимает с вероятностью $\frac{1}{2}$.

В дальнейшем, через Q_n будем обозначать распределение случайного вектора, составленного из n независимых радемахеровских случайных величин.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Определение 5.6.

Случайная величина, заданная на некотором вероятностном пространстве, называется **радемахеровской**, если она принимает только два значения -1 и 1. При этом, каждое из этих значений она принимает с вероятностью $\frac{1}{2}$.

В дальнейшем, через Q_n будем обозначать распределение случайного вектора, составленного из n независимых радемахеровских случайных величин.

Утверждение 5.5.

Пусть $\varepsilon_1, \ldots, \varepsilon_n$ $(n \in \mathbb{N})$ – независимые радемахеровские случайные величины и $y_1, \ldots, y_n \in \{\pm 1\}$. Тогда $y_1\varepsilon_1, \ldots, y_n\varepsilon_n$ также независимые радемахеровские случайные величины.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Теорема 5.4 (Вапник-Червоненкис).

Пусть $\mathsf{P} \in \mathcal{M}_{+}^{1}(\mathsf{Z})$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Предположим, что любая функция из $\mathcal{F}$ является P-измеримой.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Теорема 5.4 (Вапник-Червоненкис).

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Предположим, что любая функция из $\mathcal{F}$ является P-измеримой. Тогда

$$\mathsf{P}^n \left\{ \mathbf{z} : \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \sqrt{\frac{8 \ln \left(\Gamma_{\mathcal{F}}(2n) \frac{4}{\delta} \right)}{n}} \right\} \geqslant 1 - \delta. \tag{18}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Теорема 5.4 (Вапник-Червоненкис).

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Предположим, что любая функция из $\mathcal{F}$ является P-измеримой. Тогда

$$\mathsf{P}^n\left\{\mathbf{z} \, : \, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \sqrt{\frac{8\ln\left(\Gamma_{\mathcal{F}}(2n)\frac{4}{\delta}\right)}{n}}\right\} \geqslant 1-\delta. \tag{18}$$

◀ Рассмотрим произвольные $\mathbf{z}, \mathbf{z}' \in \mathsf{Z}^n$, $f \in \mathcal{F}$ и $\varepsilon > 0$ такое, что

$$n \geqslant 4\varepsilon^{-2}\ln 2. \tag{19}$$

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Теорема 5.4 (Вапник-Червоненкис).

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$, $\delta \in (0,1)$ и $n \in \mathbb{N}$. Предположим, что любая функция из $\mathcal{F}$ является P-измеримой. Тогда

$$\mathsf{P}^n \left\{ \mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \sqrt{\frac{8 \ln \left(\Gamma_{\mathcal{F}}(2n) \frac{4}{\delta} \right)}{n}} \right\} \geqslant 1 - \delta. \tag{18}$$

◀ Рассмотрим произвольные $\mathbf{z}, \mathbf{z}' \in \mathsf{Z}^n$, $f \in \mathcal{F}$ и $\varepsilon > 0$ такое, что

$$n \geqslant 4\varepsilon^{-2} \ln 2. \tag{19}$$

Замечание

В дальнейшем будет выбрано

$$\varepsilon = \sqrt{\frac{8 \ln \left(\Gamma_{\mathcal{F}}(2n) \frac{4}{\delta} \right)}{n}}.$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Замечание

Заметим также, что неравенство

$$\mathrm{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \varepsilon\right\} \geqslant 1 - \delta$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Замечание

Заметим также, что неравенство

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \varepsilon\right\} \geqslant 1 - \delta$$

эквивалентно неравенству

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \exists f \in \mathcal{F} \,:\, |R(f) - r(f, \mathbf{z})| > \varepsilon\right\} \leqslant \delta,$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Замечание

Заметим также, что неравенство

$$\mathsf{P}^n\Big\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \varepsilon\Big\} \geqslant 1 - \delta$$

эквивалентно неравенству

$$\mathsf{P}^n\Big\{\mathbf{z} \,:\, \exists f \in \mathcal{F} \,:\, |R(f) - r(f, \mathbf{z})| > \varepsilon\Big\} \leqslant \delta,$$

которое и будет доказываться.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Если одновременно выполняются условия $|R(f) - r(f, \mathbf{z})| > \varepsilon$ и $|R(f) - r(f, \mathbf{z}')| < \frac{\varepsilon}{2}$, то

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Если одновременно выполняются условия $|R(f) - r(f, \mathbf{z})| > \varepsilon$ и $|R(f) - r(f, \mathbf{z}')| < \frac{\varepsilon}{2}$, то

$$|r(f, \mathbf{z}) - r(f, \mathbf{z}')| > \frac{\varepsilon}{2}.$$

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Если одновременно выполняются условия $|R(f) - r(f, \mathbf{z})| > \varepsilon$ и $|R(f) - r(f, \mathbf{z}')| < \frac{\varepsilon}{2}$, то

$$|r(f, \mathbf{z}) - r(f, \mathbf{z}')| > \frac{\varepsilon}{2}.$$

С помощью характеристических функций эта импликация может быть записана в виде

$$\mathbf{1}_{\{\mathbf{z} : |R(f) - r(f,\mathbf{z})| > \varepsilon\}} \mathbf{1}_{\{\mathbf{z}' : |R(f) - r(f,\mathbf{z}')| < \varepsilon/2\}} \leqslant \mathbf{1}_{\{(\mathbf{z},\mathbf{z}') : |r(f,\mathbf{z}) - r(f,\mathbf{z}')| > \varepsilon/2\}}. \tag{20}$$

Фундаментальная теорема

Импликация 6 ⇒ 1

Если одновременно выполняются условия $|R(f) - r(f, \mathbf{z})| > \varepsilon$ и $|R(f) - r(f, \mathbf{z}')| < \frac{\varepsilon}{2}$, то

$$|r(f, \mathbf{z}) - r(f, \mathbf{z}')| > \frac{\varepsilon}{2}.$$

С помощью характеристических функций эта импликация может быть записана в виде

$$\mathbf{1}_{\{\mathbf{z}\,:\,|R(f)-r(f,\mathbf{z})|>\varepsilon\}}\,\mathbf{1}_{\{\mathbf{z}'\,:\,|R(f)-r(f,\mathbf{z}')|<\varepsilon/2\}} \leqslant \mathbf{1}_{\{(\mathbf{z},\mathbf{z}')\,:\,|r(f,\mathbf{z})-r(f,\mathbf{z}')|>\varepsilon/2\}}. \tag{20}$$

Применяя неравенство Хёффдинга, получим

$$\begin{aligned}\mathop{\mathbf{E}}_{\mathbf{z}'\sim \mathrm{P}^n}\left[\mathbf{1}_{\{\mathbf{z}'\,:\,|R(f)-r(f,\mathbf{z}')|<\varepsilon/2\}}\right] &= \mathrm{P}^n\left\{\mathbf{z}' \;:\; |R(f) - r(f, \mathbf{z}')| < \frac{\varepsilon}{2}\right\} \\ &\geqslant 1 - 2\exp\left(-\frac{n\varepsilon^2}{2}\right) \geqslant \frac{1}{2}.\end{aligned} \tag{21}$$

Фундаментальная теорема

Импликация 6 ⇒ 1

Далее, заметим

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbf{z}'\sim \mathrm{P}^n}\left[\mathbf{1}_{\{(\mathbf{z},\mathbf{z}')\,:\,|r(f,\mathbf{z})-r(r,\mathbf{z}')|>\varepsilon/2\}}\right] = \\
&\mathrm{P}^n\left\{\mathbf{z}'\ :\ |r(f,\mathbf{z})-r(f,\mathbf{z}')|>\frac{\varepsilon}{2}\right\}\leqslant \\
&\mathrm{P}^n\left\{\mathbf{z}'\ :\ \exists f\in\mathcal{F}\ :\ |r(f,\mathbf{z})-r(f,\mathbf{z}')|>\frac{\varepsilon}{2}\right\}.
\end{aligned}
\tag{22}
$$

Фундаментальная теорема

Импликация 6 ⇒ 1

Далее, заметим

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbf{z}'\sim \mathrm{P}^n}\left[\mathbf{1}_{\{(\mathbf{z},\mathbf{z}') \,:\, |r(f,\mathbf{z})-r(r,\mathbf{z}')|>\varepsilon/2\}}\right] = \\
&\mathrm{P}^n\left\{\mathbf{z}' \;:\; |r(f,\mathbf{z})-r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\right\} \leqslant \\
&\mathrm{P}^n\left\{\mathbf{z}' \;:\; \exists f\in\mathcal{F} \;:\; |r(f,\mathbf{z})-r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\right\}.
\end{aligned}
\tag{22}
$$

Интегрируя по $\mathbf{z}'$ левую и правую части неравенства (20), и применяя (21) и (22), получим

$$
\mathbf{1}_{\{\mathbf{z} \,:\, |R(f)-r(f,\mathbf{z})|>\varepsilon\}} \leqslant 2\,\mathrm{P}^n\left\{\mathbf{z}' \;:\; \exists f\in\mathcal{F} \;:\; |r(f,\mathbf{z})-r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\right\}. \tag{23}
$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

В левой части неравенства (23) функция $f \in \mathcal{F}$ может быть выбрана произвольным образом. Поэтому

$$\begin{aligned}&\mathbf{1}_{\{\mathbf{z}\,:\,\exists f\in\mathcal{F}\,:\,|R(f)-r(f,\mathbf{z})|>\varepsilon\}} \leqslant \\ &2\,\mathrm{P}^n\left\{\mathbf{z}' \,:\, \exists f \in \mathcal{F} \,:\, |r(f,\mathbf{z}) - r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\right\}.\end{aligned} \tag{24}$$

Фундаментальная теорема

Импликация 6 ⇒ 1

В левой части неравенства (23) функция $f \in \mathcal{F}$ может быть выбрана произвольным образом. Поэтому

$$\mathbf{1}_{\{\mathbf{z}\,:\,\exists f\in\mathcal{F}\,:\,|R(f)-r(f,\mathbf{z})|>\varepsilon\}} \leqslant 2\,\mathsf{P}^n\Big\{\mathbf{z}' \;:\; \exists f \in \mathcal{F} \;:\; |r(f,\mathbf{z}) - r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\Big\}. \tag{24}$$

Проинтегрируем по $\mathbf{z}$ левую и правую части неравенства (24) и, применяя теорему Фубини, получим

$$\mathsf{P}^n\Big\{\mathbf{z} \;:\; \exists f \in \mathcal{F} \;:\; |R(f) - r(f,\mathbf{z})| > \varepsilon\Big\} \leqslant 2\,\mathsf{P}^{2n}\Big\{(\mathbf{z},\mathbf{z}') \;:\; \exists f \in \mathcal{F} \;:\; |r(f,\mathbf{z}) - r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\Big\}. \tag{25}$$

Фундаментальная теорема

Импликация 6 ⇒ 1

В левой части неравенства (23) функция $f \in \mathcal{F}$ может быть выбрана произвольным образом. Поэтому

$$\begin{gathered}\mathbf{1}_{\{\mathbf{z}\,:\,\exists f\in\mathcal{F}\,:\,|R(f)-r(f,\mathbf{z})|>\varepsilon\}} \leqslant \\ 2\,\mathrm{P}^n\left\{\mathbf{z}' \;:\; \exists f \in \mathcal{F} \;:\; |r(f,\mathbf{z}) - r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\right\}.\end{gathered} \tag{24}$$

Проинтегрируем по $\mathbf{z}$ левую и правую части неравенства (24) и, применяя теорему Фубини, получим

$$\begin{gathered}\mathrm{P}^n\left\{\mathbf{z} \;:\; \exists f \in \mathcal{F} \;:\; |R(f) - r(f,\mathbf{z})| > \varepsilon\right\} \leqslant \\ 2\,\mathrm{P}^{2n}\left\{(\mathbf{z},\mathbf{z}') \;:\; \exists f \in \mathcal{F} \;:\; |r(f,\mathbf{z}) - r(f,\mathbf{z}')| > \frac{\varepsilon}{2}\right\}.\end{gathered} \tag{25}$$

Теперь наша основная цель состоит в оценке правой части неравенства (25).

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Для этого будет использовано стандартное неравенство для вероятности конечного объединения событий и неравенство Хёффдинга, которое будет применено к каждому слагаемому получившейся конечной суммы вероятностей.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Для этого будет использовано стандартное неравенство для вероятности конечного объединения событий и неравенство Хёффдинга, которое будет применено к каждому слагаемому получившейся конечной суммы вероятностей.

Этого можно достичь, перейдя от рассмотрения потенциально бесконечного класса функций $\mathcal{F}$ к рассмотрению конечного класса функций.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Для этого будет использовано стандартное неравенство для вероятности конечного объединения событий и неравенство Хёффдинга, которое будет применено к каждому слагаемому получившейся конечной суммы вероятностей.

Этого можно достичь, перейдя от рассмотрения потенциально бесконечного класса функций $\mathcal{F}$ к рассмотрению конечного класса функций.

Заметим, что для фиксированных $\mathbf{z}, \mathbf{z}' \in Z^n$ класс функций $\mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}$ конечен.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Для этого будет использовано стандартное неравенство для вероятности конечного объединения событий и неравенство Хёффдинга, которое будет применено к каждому слагаемому получившейся конечной суммы вероятностей.

Этого можно достичь, перейдя от рассмотрения потенциально бесконечного класса функций $\mathcal{F}$ к рассмотрению конечного класса функций.

Заметим, что для фиксированных $\mathbf{z}, \mathbf{z}' \in Z^n$ класс функций $\mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}$ конечен.

Он представляет собой ограничение $\mathcal{F}$ на конечном множестве элементов из наборов $\mathbf{z}$ и $\mathbf{z}'$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Для этого будет использовано стандартное неравенство для вероятности конечного объединения событий и неравенство Хёффдинга, которое будет применено к каждому слагаемому получившейся конечной суммы вероятностей.

Этого можно достичь, перейдя от рассмотрения потенциально бесконечного класса функций $\mathcal{F}$ к рассмотрению конечного класса функций.

Заметим, что для фиксированных $\mathbf{z}, \mathbf{z}' \in Z^n$ класс функций $\mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}$ конечен.

Он представляет собой ограничение $\mathcal{F}$ на конечном множестве элементов из наборов $\mathbf{z}$ и $\mathbf{z}'$.

Заметим, что $|\mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}| \leqslant \Gamma_{\mathcal{F}}(2n)$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

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

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

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

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

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Используем вспомогательное вероятностное пространство $(\mathsf{Z}^{2n}, \otimes^{2n}\mathscr{Z}, \mathsf{P}^{2n})$.

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.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Используем 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.

Для любой функции $f \in \mathcal{F}$ и произвольного набора $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n) \in \{\pm 1\}^n$ независимыми и одинаково распределёнными будут случайные величины

$$[f(Z_1) - f(Z'_1)]\epsilon_1, \ldots, [f(Z_n) - f(Z'_n)]\epsilon_n,$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Используем 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.

Для любой функции $f \in \mathcal{F}$ и произвольного набора $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \{\pm 1\}^n$ независимыми и одинаково распределёнными будут случайные величины

$$[f(Z_1) - f(Z'_1)]\epsilon_1, \ldots, [f(Z_n) - f(Z'_n)]\epsilon_n,$$

а значит, одинаково распределены и случайные величины

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

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Далее, с учётом теоремы о замене переменных в интеграле Лебега получим

$$\begin{gathered}
\mathsf{P}^{2n}\left\{(\mathbf{z},\mathbf{z}') \;:\; \exists f\in\mathcal{F} \;:\; |r(f,\mathbf{z})-r(f,\mathbf{z}')|>\frac{\varepsilon}{2}\right\} = \\
\mathsf{P}^{2n}\left\{(\mathbf{z},\mathbf{z}') \;:\; \exists f\in\mathcal{F} \;:\; \left|\frac{1}{n}\sum_{i=1}^{n}[f(z_i)-f(z_i')]\epsilon_i\right|>\frac{\varepsilon}{2}\right\}.
\end{gathered} \tag{26}$$

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Далее, с учётом теоремы о замене переменных в интеграле Лебега получим

$$\begin{aligned} &\mathrm{P}^{2n}\left\{(\mathbf{z}, \mathbf{z}') \,:\, \exists f \in \mathcal{F} \,:\, |r(f, \mathbf{z}) - r(f, \mathbf{z}')| > \frac{\varepsilon}{2}\right\} = \\ &\mathrm{P}^{2n}\left\{(\mathbf{z}, \mathbf{z}') \,:\, \exists f \in \mathcal{F} \,:\, \left|\frac{1}{n}\sum_{i=1}^{n}[f(z_i) - f(z_i')]\epsilon_i\right| > \frac{\varepsilon}{2}\right\}. \end{aligned} \tag{26}$$

Проинтегрируем по ϵ левую и правую части неравенства (25) и, применяя (26), а также изменяя с помощью теоремы Фубини порядок интегрирования, получим

$$\begin{aligned} &\mathrm{P}^{n}\left\{\mathbf{z} \,:\, \exists f \in \mathcal{F} \,:\, |R(f) - r(f, \mathbf{z})| > \varepsilon\right\} \leqslant \\ &2 \mathop{\mathbf{E}}_{(\mathbf{z}, \mathbf{z}') \sim \mathrm{P}^{2n}} \mathrm{Q}_n\left\{\epsilon \,:\, \exists f \in \mathcal{F} \,:\, \left|\frac{1}{n}\sum_{i=1}^{n}[f(z_i) - f(z_i')]\epsilon_i\right| > \frac{\varepsilon}{2}\right\}. \end{aligned} \tag{27}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

При фиксированных $\mathbf{z}, \mathbf{z}' \in Z^n$ для каждой функции $\hat{f} \in \mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}$ определим множество

$$E(\hat{f}) := \left\{ \epsilon \, : \, \left| \frac{1}{n} \sum_{i=1}^{n} [\hat{f}(z_i) - \hat{f}(z_i')]\epsilon_i \right| > \frac{\varepsilon}{2} \right\}.$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

При фиксированных $\mathbf{z}, \mathbf{z}' \in Z^n$ для каждой функции $\hat{f} \in \mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}$ определим множество

$$E(\hat{f}) := \left\{ \epsilon \, : \, \left| \frac{1}{n} \sum_{i=1}^{n} [\hat{f}(z_i) - \hat{f}(z_i')]\epsilon_i \right| > \frac{\varepsilon}{2} \right\}.$$

Тогда с помощью стандартного неравенства для вероятности конечного объединения событий получим

$$\begin{aligned} &Q_n \left\{ \epsilon \, : \, \exists f \in \mathcal{F} \, : \, \left| \frac{1}{n} \sum_{i=1}^{n} [f(z_i) - f(z_i')]\epsilon_i \right| > \frac{\varepsilon}{2} \right\} \\ =&Q_n \bigcup_{\hat{f} \in \mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}} E(\hat{f}) \leqslant \sum_{\hat{f} \in \mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}} Q_n E(\hat{f}). \end{aligned} \tag{28}$$

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Для каждой функции $\hat{f} \in \mathcal{F}|_{\mathbf{z} \cup \mathbf{z}'}$ с помощью неравенства Хёффдинга запишем оценку

$$Q_n\left\{\epsilon \,:\, \left|\frac{1}{n}\sum_{i=1}^{n}[\hat{f}(z_i) - \hat{f}(z_i')]\epsilon_i\right| > \frac{\varepsilon}{2}\right\} \leqslant 2\,\exp\left[-2\left(\frac{n\varepsilon}{2}\right)^2\frac{1}{4n}\right] = 2\,\exp\left[-\frac{n\varepsilon^2}{8}\right]. \tag{29}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Для каждой функции $\hat{f} \in \mathcal{F}|_{\mathbf{z}\cup\mathbf{z}'}$ с помощью неравенства Хёффдинга запишем оценку

$$\begin{aligned} \mathrm{Q}_n\left\{\epsilon \,:\, \left|\frac{1}{n}\sum_{i=1}^{n}[\hat{f}(z_i) - \hat{f}(z_i')]\epsilon_i\right| > \frac{\varepsilon}{2}\right\} &\leqslant 2\,\exp\left[-2\left(\frac{n\varepsilon}{2}\right)^2\frac{1}{4n}\right] \\ &= 2\,\exp\left[-\frac{n\varepsilon^2}{8}\right]. \end{aligned} \tag{29}$$

Объединяя вместе (27), (28) и (29), получим

$$\begin{aligned} \mathrm{P}^n\left\{\mathbf{z} \,:\, \exists f \in \mathcal{F} \,:\, |R(f) - r(f,\mathbf{z})| > \varepsilon\right\} &\leqslant 4\,\left|\mathcal{F}|_{\mathbf{z}\cup\mathbf{z}'}\right|\exp\left[-\frac{n\varepsilon^2}{8}\right] \\ &\leqslant 4\,\Gamma_{\mathcal{F}}(2n)\exp\left[-\frac{n\varepsilon^2}{8}\right]. \end{aligned} \tag{30}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Приравняем правую часть неравенства (30) к δ и решим относительно ε получившее уравнение

$$4\,\Gamma_{\mathcal{F}}(2n)\exp\left[-\frac{n\varepsilon^2}{8}\right] = \delta \tag{31}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Приравняем правую часть неравенства (30) к δ и решим относительно ε получившее уравнение

$$4\,\Gamma_{\mathcal{F}}(2n)\exp\left[-\frac{n\varepsilon^2}{8}\right]=\delta \tag{31}$$

Его решением будет

$$\varepsilon^2=\frac{8}{n}\ln\left(\Gamma_{\mathcal{F}}(2n)\frac{4}{\delta}\right). \tag{32}$$

Приравняем правую часть неравенства (30) к δ и решим относительно ε получившее уравнение

$$4\,\Gamma_{\mathcal{F}}(2n)\exp\left[-\frac{n\varepsilon^2}{8}\right] = \delta \tag{31}$$

Его решением будет

$$\varepsilon^2 = \frac{8}{n}\ln\left(\Gamma_{\mathcal{F}}(2n)\frac{4}{\delta}\right). \tag{32}$$

Для завершения доказательства остаётся проверить, что решение (32) удовлетворяет условию (19).

Подставим (32) в (19).

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Подставим (32) в (19).

Решением получившегося относительно δ неравенства будет

$$\delta \leqslant 2\sqrt{2}\,\Gamma_{\mathcal{F}}(2n). \tag{33}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Подставим (32) в (19).

Решением получившегося относительно δ неравенства будет

$$\delta \leqslant 2\sqrt{2}\,\Gamma_{\mathcal{F}}(2n). \tag{33}$$

Осталось заметить, что правая часть неравенства (33) строго больше единицы.

■

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Нам потребуется следующее техническое утверждение.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Нам потребуется следующее техническое утверждение.

Утверждение 5.6.

Пусть $a, b > 0$. Тогда $\ln a \leqslant ab - \ln b - 1$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Нам потребуется следующее техническое утверждение.

Утверждение 5.6.

Пусть $a, b > 0$. Тогда $\ln a \leqslant ab - \ln b - 1$.

◀ В известном неравенстве $1 + u \leqslant e^u$ $(u \in \mathbb{R})$ достаточно выбрать $u := ab - 1$.

■

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Теорема 5.5.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и семейство $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$ является P-измеримой.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Теорема 5.5.

Пусть $\mathcal{F} \subseteq \{0, 1\}^{\mathsf{Z}}$ и семейство $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$ является P-измеримой. Предположим, что $d := \mathrm{vc}(\mathcal{F}) < \infty$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Теорема 5.5.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и семейство $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$ является P-измеримой. Предположим, что $d := \mathrm{vc}(\mathcal{F}) < \infty$.

Тогда класс функций $\mathcal{F}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P}$.

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Теорема 5.5.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и семейство $\mathcal{P} \subseteq \mathcal{M}_+^1(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$ является P-измеримой. Предположим, что $d := \mathrm{vc}(\mathcal{F}) < \infty$.

Тогда класс функций $\mathcal{F}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P}$.

При этом, если $d \geqslant 2$, то

$$n_{\mathcal{F}}^{\mathrm{uc}}(\varepsilon,\delta) \leqslant \left\lceil \frac{16}{\varepsilon^2}\left(d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{4}{\delta}\right)\right)\right\rceil \quad (\varepsilon,\delta \in (0,1)), \tag{34}$$

Фундаментальная теорема

Импликация 6 ⇒ 1

Теорема 5.5.

Пусть $\mathcal{F} \subseteq \{0,1\}^{\mathsf{Z}}$ и семейство $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$ является P-измеримой. Предположим, что $d := \mathrm{vc}(\mathcal{F}) < \infty$.

Тогда класс функций $\mathcal{F}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P}$.

При этом, если $d \geqslant 2$, то

$$n^{\mathrm{uc}}_{\mathcal{F}}(\varepsilon,\delta) \leqslant \left\lceil \frac{16}{\varepsilon^2}\left(d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{4}{\delta}\right)\right)\right\rceil \quad \bigl(\varepsilon,\delta \in (0,1)\bigr), \tag{34}$$

а если $d = 1$, то

$$n^{\mathrm{uc}}_{\mathcal{F}}(\varepsilon,\delta) \leqslant \left\lceil \frac{16}{\varepsilon^2}\left(\ln\left(\frac{16}{\varepsilon^2}\right) + \ln\left(\frac{8}{\delta}\right)\right)\right\rceil \quad \bigl(\varepsilon,\delta \in (0,1)\bigr). \tag{35}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

◀ Прежде всего заметим, что из оценок (34) и (35), следует конечность функции сложности выборки $n_{\mathcal{F}}^{\mathrm{uc}}$, а значит, будет следовать и свойство равномерной сходимости эмпирического риска класса функций $\mathcal{F}$.

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

◀ Прежде всего заметим, что из оценок (34) и (35), следует конечность функции сложности выборки $n_{\mathcal{F}}^{\mathrm{uc}}$, а значит, будет следовать и свойство равномерной сходимости эмпирического риска класса функций $\mathcal{F}$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$ и $\varepsilon, \delta \in (0, 1)$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

◀ Прежде всего заметим, что из оценок (34) и (35), следует конечность функции сложности выборки $n_{\mathcal{F}}^{\mathrm{uc}}$, а значит, будет следовать и свойство равномерной сходимости эмпирического риска класса функций $\mathcal{F}$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$ и $\varepsilon, \delta \in (0, 1)$.

В оценке (18) из предыдущей теоремы 5.4 ограничим правую часть внутреннего неравенства числом ε.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

◀ Прежде всего заметим, что из оценок (34) и (35), следует конечность функции сложности выборки $n^{\mathrm{uc}}_{\mathcal{F}}$, а значит, будет следовать и свойство равномерной сходимости эмпирического риска класса функций $\mathcal{F}$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$ и $\varepsilon, \delta \in (0, 1)$.

В оценке (18) из предыдущей теоремы 5.4 ограничим правую часть внутреннего неравенства числом ε.

Получим, что при

$$n \geqslant \frac{8}{\varepsilon^2}\left[\ln\left(\Gamma_{\mathcal{F}}(2n)\right) + \ln\left(\frac{4}{\delta}\right)\right] \tag{36}$$

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

◀ Прежде всего заметим, что из оценок (34) и (35), следует конечность функции сложности выборки $n_{\mathcal{F}}^{\mathrm{uc}}$, а значит, будет следовать и свойство равномерной сходимости эмпирического риска класса функций $\mathcal{F}$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$ и $\varepsilon, \delta \in (0, 1)$.

В оценке (18) из предыдущей теоремы 5.4 ограничим правую часть внутреннего неравенства числом ε.

Получим, что при

$$n \geqslant \frac{8}{\varepsilon^2}\left[\ln\left(\Gamma_{\mathcal{F}}(2n)\right) + \ln\left(\frac{4}{\delta}\right)\right] \tag{36}$$

выполняется неравенство

$$\mathsf{P}^n\left\{\mathbf{z} \,:\, \Delta_{\mathcal{F}}(\mathbf{z}) \leqslant \varepsilon\right\} \geqslant 1 - \delta. \tag{37}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d \geqslant 2$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d \geqslant 2$.

В силу произвольности выбора вероятностной меры P и значений $\varepsilon, \delta \in (0, 1)$ для доказательства оценки (34) достаточно показать,

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d \geqslant 2$.

В силу произвольности выбора вероятностной меры P и значений $\varepsilon, \delta \in (0, 1)$ для доказательства оценки (34) достаточно показать,

что для всех

$$n \geqslant \frac{16}{\varepsilon^2}\left[d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{4}{\delta}\right)\right] \tag{38}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d \geqslant 2$.

В силу произвольности выбора вероятностной меры P и значений $\varepsilon, \delta \in (0, 1)$ для доказательства оценки (34) достаточно показать,

что для всех

$$n \geqslant \frac{16}{\varepsilon^2}\left[d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{4}{\delta}\right)\right] \tag{38}$$

выполняется неравенство (36).

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d \geqslant 2$.

В силу произвольности выбора вероятностной меры P и значений $\varepsilon, \delta \in (0, 1)$ для доказательства оценки (34) достаточно показать,

что для всех

$$n \geqslant \frac{16}{\varepsilon^2}\left[d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{4}{\delta}\right)\right] \tag{38}$$

выполняется неравенство (36).

Из неравенства (38) следует, что $n \geqslant d$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d \geqslant 2$.

В силу произвольности выбора вероятностной меры P и значений $\varepsilon, \delta \in (0, 1)$ для доказательства оценки (34) достаточно показать,

что для всех

$$n \geqslant \frac{16}{\varepsilon^2}\left[d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{4}{\delta}\right)\right] \tag{38}$$

выполняется неравенство (36).

Из неравенства (38) следует, что $n \geqslant d$.

По лемме Сауэра-Шелаха для функции роста имеет место оценка $\Gamma_{\mathcal{F}}(2n) \leqslant (2en/d)^d$.

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Следовательно, неравенство (36) будет выполняться для всех

$$n \geqslant \frac{8}{\varepsilon^2}\left[d\ln(n) + d\ln\left(\frac{2e}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \tag{39}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Следовательно, неравенство (36) будет выполняться для всех

$$n \geqslant \frac{8}{\varepsilon^2}\left[d\ln(n) + d\ln\left(\frac{2e}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \tag{39}$$

Воспользуемся утв. 5.6.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Следовательно, неравенство (36) будет выполняться для всех

$$n \geqslant \frac{8}{\varepsilon^2}\left[d \ln(n) + d \ln\left(\frac{2e}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \tag{39}$$

Воспользуемся утв. 5.6.

Полагая $x := n$ и $a := \varepsilon^2/(16d)$, получим

$$\ln(n) \leqslant \frac{n\varepsilon^2}{16d} - \ln\left(\frac{\varepsilon^2}{16d}\right) - 1.$$

Фундаментальная теорема

Импликация 6 $\Rightarrow$ 1

Следовательно, неравенство (36) будет выполняться для всех

$$n \geqslant \frac{8}{\varepsilon^2}\left[d\ln(n) + d\ln\left(\frac{2e}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \tag{39}$$

Воспользуемся утв. 5.6.

Полагая $x := n$ и $a := \varepsilon^2/(16d)$, получим

$$\ln(n) \leqslant \frac{n\varepsilon^2}{16d} - \ln\left(\frac{\varepsilon^2}{16d}\right) - 1.$$

Но тогда неравенство (39) будет выполняться для всех

$$\begin{aligned} n &\geqslant \frac{8}{\varepsilon^2}\left[\frac{n\varepsilon^2}{16} + d\ln\left(\frac{16d}{\varepsilon^2}\right) - d + d + \ln\left(\frac{2}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right] \\ &= \frac{n}{2} + \frac{8}{\varepsilon^2}\left[d\ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{2}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \end{aligned}$$

Это условие может быть переписано в виде

$$n \geqslant \frac{16}{\varepsilon^2}\left[d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{2}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \tag{40}$$

Фундаментальная теорема
Импликация $6 \Rightarrow 1$

Это условие может быть переписано в виде

$$n \geqslant \frac{16}{\varepsilon^2}\left[d \ln\left(\frac{16d}{\varepsilon^2}\right) + \ln\left(\frac{2}{d}\right) + \ln\left(\frac{4}{\delta}\right)\right]. \tag{40}$$

По предположению $\ln(2/d) \leqslant 0$, а это значит, что из неравенства (38) следует выполнение неравенства (40).

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d = 1$.

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d = 1$.

Повторяя начальные шаги доказательства из предыдущего случая, также получим условие (40).

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d = 1$.

Повторяя начальные шаги доказательства из предыдущего случая, также получим условие (40).

При $d = 1$ оно может быть переписано в эквивалентном виде

$$n \geqslant \frac{16}{\varepsilon^2}\left[\ln\left(\frac{16}{\varepsilon^2}\right) + \ln\left(\frac{8}{\delta}\right)\right]. \tag{41}$$

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d = 1$.

Повторяя начальные шаги доказательства из предыдущего случая, также получим условие (40).

При $d = 1$ оно может быть переписано в эквивалентном виде

$$n \geqslant \frac{16}{\varepsilon^2}\left[\ln\left(\frac{16}{\varepsilon^2}\right) + \ln\left(\frac{8}{\delta}\right)\right]. \tag{41}$$

Таким образом, для всех $n \in \mathbb{N}$, удовлетворяющих условию (41), будет выполняться неравенство (36), а значит, будет выполняться и неравенство (37).

Фундаментальная теорема

Импликация $6 \Rightarrow 1$

Рассмотрим случай $d = 1$.

Повторяя начальные шаги доказательства из предыдущего случая, также получим условие (40).

При $d = 1$ оно может быть переписано в эквивалентном виде

$$n \geqslant \frac{16}{\varepsilon^2}\left[\ln\left(\frac{16}{\varepsilon^2}\right) + \ln\left(\frac{8}{\delta}\right)\right]. \tag{41}$$

Таким образом, для всех $n \in \mathbb{N}$, удовлетворяющих условию (41), будет выполняться неравенство (36), а значит, будет выполняться и неравенство (37).

В виду произвольности выбора вероятностной меры P и значений $\varepsilon, \delta \in (0, 1)$ это означает справедливость оценки (35).

■