Теоретические основы информатики (концептуальные модели и математические основы)

Лекция № 5. Формальные модели обучения

А.С. Шундеев

Содержание

Содержание

Модель PAC-обучения

Опишем формальную модель для реализуемого случая задачи обучения концептов, которая носит название PAC-обучение.

Модель PAC-обучения

Опишем формальную модель для реализуемого случая задачи обучения концептов, которая носит название PAC-обучение.

В рамках этой модели фиксируется некоторый класс концептов.

Модель PAC-обучения

Опишем формальную модель для реализуемого случая задачи обучения концептов, которая носит название PAC-обучение.

В рамках этой модели фиксируется некоторый класс концептов.

Каждый из его элементов может выступать в роли целевого концепта.

Модель PAC-обучения

Опишем формальную модель для реализуемого случая задачи обучения концептов, которая носит название PAC-обучение.

В рамках этой модели фиксируется некоторый класс концептов.

Каждый из его элементов может выступать в роли целевого концепта.

При этом варианты других целевых концептов не рассматриваются.

Модель PAC-обучения

Опишем формальную модель для реализуемого случая задачи обучения концептов, которая носит название PAC-обучение.

В рамках этой модели фиксируется некоторый класс концептов.

Каждый из его элементов может выступать в роли целевого концепта.

При этом варианты других целевых концептов не рассматриваются.

Предполагается, что рассматриваемые алгоритмы обучения выбирают гипотезы только из этого класса.

Модель PAC-обучения

Опишем формальную модель для реализуемого случая задачи обучения концептов, которая носит название PAC-обучение.

В рамках этой модели фиксируется некоторый класс концептов.

Каждый из его элементов может выступать в роли целевого концепта.

При этом варианты других целевых концептов не рассматриваются.

Предполагается, что рассматриваемые алгоритмы обучения выбирают гипотезы только из этого класса.

Дополнительно фиксируется некоторое семейство вероятностных мер, заданных на множестве объектов.

Модель PAC-обучения

Наличие PAC-обучаемости у класса концептов означает существование алгоритма обучения, обладающего следующим свойствам.

Модель PAC-обучения

Наличие PAC-обучаемости у класса концептов означает существование алгоритма обучения, обладающего следующим свойствам.

Всегда может быть явно указана нижняя граница для размера обучающих выборок, при использовании которых этот алгоритм с заданной вероятностью будет строить приближение к целевому концепту с заданной точностью.

Модель PAC-обучения

Наличие PAC-обучаемости у класса концептов означает существование алгоритма обучения, обладающего следующим свойствам.

Всегда может быть явно указана нижняя граница для размера обучающих выборок, при использовании которых этот алгоритм с заданной вероятностью будет строить приближение к целевому концепту с заданной точностью.

В качестве оценки точности построенного приближения используется значение соответствующего ожидаемого риска.

Модель PAC-обучения

Наличие PAC-обучаемости у класса концептов означает существование алгоритма обучения, обладающего следующим свойствам.

Всегда может быть явно указана нижняя граница для размера обучающих выборок, при использовании которых этот алгоритм с заданной вероятностью будет строить приближение к целевому концепту с заданной точностью.

В качестве оценки точности построенного приближения используется значение соответствующего ожидаемого риска.

При этом предполагается, что подобная нижняя оценка не должна зависеть ни от выбора целевого концепта, ни от выбора вероятностной меры из рассматриваемого семейства.

Содержание

Модель PAC-обучения

Основные определения

Определение 4.1.

Класс концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$ называется PAC-обучаемым относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$, если существует алгоритм обучения $\mathcal{A}$, обладающий следующим свойством.

Определение 4.1.

Класс концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$ называется PAC-обучаемым относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$, если существует алгоритм обучения $\mathcal{A}$, обладающий следующим свойством.

Для любого $\varepsilon \in (0, 1)$ выполняется условие

$$\lim_{n\to\infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, \varepsilon) = 0, \tag{1}$$

где

$$r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, \varepsilon) := \sup_{\mathsf{P}_X \in \mathcal{P}_{\mathsf{X}}} \sup_{C' \in \mathcal{C}} \mathsf{P}^n_X \{ \mathbf{x} \in \mathsf{X}^n \ : \ C = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}), \, R(\mathsf{P}_X; C', C) > \varepsilon \}.$$

Модель PAC-обучения

Основные определения

Определение 4.1.

Класс концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$ называется PAC-обучаемым относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}_{+}^{1}(\mathsf{X})$, если существует алгоритм обучения $\mathcal{A}$, обладающий следующим свойством.

Для любого $\varepsilon \in (0, 1)$ выполняется условие

$$\lim_{n \to \infty} r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, \varepsilon) = 0, \tag{1}$$

где

$$r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, \varepsilon) := \sup_{\mathsf{P}_X \in \mathcal{P}_{\mathsf{X}}} \sup_{C' \in \mathcal{C}} \mathsf{P}_X^n\left\{\mathbf{x} \in \mathsf{X}^n \,:\, C = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}),\ R(\mathsf{P}_X; C', C) > \varepsilon\right\}.$$

При этом алгоритм $\mathcal{A}$ называется PAC-учителем для класса $\mathcal{C}$.

Модель PAC-обучения

Основные определения

Наибольший интерес представляют два семейства вероятностных мер.

Модель PAC-обучения

Основные определения

Наибольший интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}_{\mathsf{X}}$ выбирается $\mathcal{M}^1_+(\mathsf{X}, \mathcal{S}_{\mathsf{X}})$, где $\mathcal{S}_{\mathsf{X}}$ – некоторая фиксированная σ-алгебра.

Модель PAC-обучения

Основные определения

Наибольший интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}_\mathsf{X}$ выбирается $\mathcal{M}^1_+(\mathsf{X}, \mathcal{S}_\mathsf{X})$, где $\mathcal{S}_\mathsf{X}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса концептов $\mathcal{C}$, семейство $\mathcal{P}_\mathsf{X}$ образуют все вероятностные меры P_X такие, что любой концепт из $\mathcal{C}$ является P_X-измеримым.

Модель PAC-обучения

Основные определения

Наибольший интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}_{\mathsf{X}}$ выбирается $\mathcal{M}_{+}^{1}(\mathsf{X}, \mathcal{S}_{\mathsf{X}})$, где $\mathcal{S}_{\mathsf{X}}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса концептов $\mathcal{C}$, семейство $\mathcal{P}_{\mathsf{X}}$ образуют все вероятностные меры P_X такие, что любой концепт из $\mathcal{C}$ является P_X-измеримым.

Если явно не оговорено противное, то всегда будет неявно подразумеваться второй случай.

Модель PAC-обучения

Основные определения

Наибольший интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}_\mathsf{X}$ выбирается $\mathcal{M}_+^1(\mathsf{X}, \mathcal{S}_\mathsf{X})$, где $\mathcal{S}_\mathsf{X}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса концептов $\mathcal{C}$, семейство $\mathcal{P}_\mathsf{X}$ образуют все вероятностные меры P_X такие, что любой концепт из $\mathcal{C}$ является P_X-измеримым.

Если явно не оговорено противное, то всегда будет неявно подразумеваться второй случай.

С практической точки зрения особый интерес представляет скорость сходимости последовательности (1) к своему пределу.

Модель PAC-обучения

Основные определения

Наибольший интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}_\mathsf{X}$ выбирается $\mathcal{M}^1_+(\mathsf{X}, \mathcal{S}_\mathsf{X})$, где $\mathcal{S}_\mathsf{X}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса концептов $\mathcal{C}$, семейство $\mathcal{P}_\mathsf{X}$ образуют все вероятностные меры P_X такие, что любой концепт из $\mathcal{C}$ является P_X-измеримым.

Если явно не оговорено противное, то всегда будет неявно подразумеваться второй случай.

С практической точки зрения особый интерес представляет скорость сходимости последовательности (1) к своему пределу. Для этого вводится понятие функции сложности выборки.

Определение 4.2.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ называется наименьшая функция вида

$$n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_{\mathsf{X}}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Определение 4.2.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P}_\mathsf{X} \subseteq \mathcal{M}^1_+(\mathsf{X})$ называется наименьшая функция вида

$$n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_\mathsf{X}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$r_\mathcal{A}(\mathcal{P}_\mathsf{X}; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_\mathsf{X}}(\varepsilon, \delta).$$

Модель PAC-обучения

Основные определения

Определение 4.2.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ называется наименьшая функция вида

$$n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_{\mathsf{X}}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$r_{\mathcal{A}}(\mathcal{P}_{\mathsf{X}}; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_{\mathsf{X}}}(\varepsilon, \delta).$$

Если понятно, о каком семействе вероятностных мер идёт речь, то будет использоваться сокращённое обозначение $n^{\mathrm{pac}}_{\mathcal{A}}$.

Определение 4.2.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P}_\mathsf{X} \subseteq \mathcal{M}_+^1(\mathsf{X})$ называется наименьшая функция вида

$$n_{\mathcal{A},\mathcal{P}_\mathsf{X}}^{\mathrm{pac}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$r_\mathcal{A}(\mathcal{P}_\mathsf{X}; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n_{\mathcal{A},\mathcal{P}_\mathsf{X}}^{\mathrm{pac}}(\varepsilon, \delta).$$

Если понятно, о каком семействе вероятностных мер идёт речь, то будет использоваться сокращённое обозначение $n_\mathcal{A}^{\mathrm{pac}}$.

У функций сложности выборки входной параметр ε будем называть точностью, а параметр δ – достоверностью.

Модель PAC-обучения

Основные определения

Из приведённого определения следует, что алгоритм обучения $\mathcal{A}$ является PAC-учителем тогда и только тогда, когда функция $n_{\mathcal{A}}^{\mathrm{pac}}$ принимает только конечные значения.

Модель PAC-обучения

Основные определения

Из приведённого определения следует, что алгоритм обучения $\mathcal{A}$ является PAC-учителем тогда и только тогда, когда функция $n_{\mathcal{A}}^{\mathrm{pac}}$ принимает только конечные значения.

Замечание

Оригинальное определение PAC-обучаемости несколько отличается от приведённого выше.

Из приведённого определения следует, что алгоритм обучения $\mathcal{A}$ является PAC-учителем тогда и только тогда, когда функция $n_{\mathcal{A}}^{\mathrm{pac}}$ принимает только конечные значения.

Замечание

Оригинальное определение PAC-обучаемости несколько отличается от приведённого выше.

В оригинальном определении дополнительно фигурировало понятие сложности описания объекта, а также тербовалось, чтобы функция сложности выборки была ограничена полиномом от величин $1/\varepsilon$ и $1/\delta$.

Модель PAC-обучения

Основные определения

Ранее отмечалось, что особую роль играют методы минимизации эмпирического риска.

Модель PAC-обучения

Основные определения

Ранее отмечалось, что особую роль играют методы минимизации эмпирического риска.

Если задача обучения имеет решение, то такие алгоритмы в некотором смысле являются «взаимозаменяемыми».

Ранее отмечалось, что особую роль играют методы минимизации эмпирического риска.

Если задача обучения имеет решение, то такие алгоритмы в некотором смысле являются «взаимозаменяемыми».

Это является мотивацией для следующего определения.

Модель PAC-обучения

Основные определения

Определение 4.3.

Сложностью выборки класса концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ называется наименьшая функция вида

$$n^{\mathrm{pac}}_{\mathcal{C},\mathcal{P}_{\mathsf{X}}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Модель PAC-обучения

Основные определения

Определение 4.3.

Сложностью выборки класса концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ называется наименьшая функция вида

$$n^{\mathrm{pac}}_{\mathcal{C},\mathcal{P}_{\mathsf{X}}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ и любого метода минимизации эмпирического риска $\mathcal{A}$ выполняется условие

$$n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_{\mathsf{X}}}(\varepsilon,\delta) \leqslant n^{\mathrm{pac}}_{\mathcal{C},\mathcal{P}_{\mathsf{X}}}(\varepsilon,\delta).$$

Модель PAC-обучения

Основные определения

Определение 4.3.

Сложностью выборки класса концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$ называется наименьшая функция вида

$$n^{\mathrm{pac}}_{\mathcal{C},\mathcal{P}_{\mathsf{X}}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ и любого метода минимизации эмпирического риска $\mathcal{A}$ выполняется условие

$$n^{\mathrm{pac}}_{\mathcal{A},\mathcal{P}_{\mathsf{X}}}(\varepsilon,\delta) \leqslant n^{\mathrm{pac}}_{\mathcal{C},\mathcal{P}_{\mathsf{X}}}(\varepsilon,\delta).$$

Если понятно, о каком семействе вероятностных мер идёт речь, то будет использоваться сокращённое обозначение $n^{\mathrm{pac}}_{\mathcal{C}}$.

Далее, будет разобран пример бесконечного PAC-обучаемого класса концептов.

Далее, будет разобран пример бесконечного PAC-обучаемого класса концептов.

После этого будет установлено, что любой конечный класс концептов является PAC-обучаемым.

Модель PAC-обучения

Основные определения

Далее, будет разобран пример бесконечного PAC-обучаемого класса концептов.

После этого будет установлено, что любой конечный класс концептов является PAC-обучаемым.

Возникает закономерный вопрос о существовании не PAC-обучаемых классов концептов.

Модель PAC-обучения

Основные определения

Теорема 4.2.

Предположим, что множество X бесконечно. Тогда класс концептов 2^{X} не является PAC-обучаемым относительно семейства вероятностных мер $\mathcal{M}_{+}^{1}(\mathsf{X}, 2^{\mathsf{X}})$.

Модель PAC-обучения

Основные определения

Теорема 4.2.

Предположим, что множество X бесконечно. Тогда класс концептов 2^X не является PAC-обучаемым относительно семейства вероятностных мер $\mathcal{M}^1_+(X, 2^X)$.

◀ Предположим обратное, что означает существование некоторого PAC-обучающего алгоритма $\mathcal{A}$.

Модель PAC-обучения

Основные определения

Теорема 4.2.

Предположим, что множество X бесконечно. Тогда класс концептов 2^{X} не является PAC-обучаемым относительно семейства вероятностных мер $\mathcal{M}_+^1(\mathsf{X}, 2^{\mathsf{X}})$.

◀ Предположим обратное, что означает существование некоторого PAC-обучающего алгоритма $\mathcal{A}$.

Зафиксируем положительное $\delta < \frac{1}{7}$ и $n > n_{\mathcal{A}}^{\mathrm{pac}}\left(\frac{1}{8}, \delta\right)$.

Теорема 4.2.

Предположим, что множество X бесконечно. Тогда класс концептов 2^{X} не является PAC-обучаемым относительно семейства вероятностных мер $\mathcal{M}_{+}^{1}(\mathrm{X}, 2^{\mathrm{X}})$.

◀ Предположим обратное, что означает существование некоторого PAC-обучающего алгоритма $\mathcal{A}$.

Зафиксируем положительное $\delta < \frac{1}{7}$ и $n > n_{\mathcal{A}}^{\mathrm{pac}}\left(\frac{1}{8}, \delta\right)$.

Тогда для любых $\mathrm{P}_X \in \mathcal{M}_{+}^{1}(\mathrm{X}, 2^{\mathrm{X}})$ и $C \in 2^{\mathrm{X}}$ выполняется неравенство

$$\mathrm{P}_X^n\left\{\mathbf{x} \in \mathrm{X}^n \;:\; C'' = \mathcal{A}(\mathbf{1}_C \circ \mathbf{x}), R\big(\mathrm{P}_X; C, C''\big) > \frac{1}{8}\right\} \leqslant \delta < \frac{1}{7}.$$

Модель PAC-обучения

Основные определения

В то же время, согласно теореме 4.1 (no free lunch) существует вероятностная мера $\widehat{\mathsf{P}}_X \in \mathcal{M}^1_+(\mathsf{X}, 2^{\mathsf{X}})$ и концепт $C' \in 2^{\mathsf{X}}$ такие, что

$$\widehat{\mathsf{P}}^n_X\left\{\mathbf{x} \in \mathsf{X}^n \;:\; C'' = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}), R\left(\widehat{\mathsf{P}}_X; C', C''\right) > \frac{1}{8}\right\} \geqslant \frac{1}{7}.$$

Модель PAC-обучения

Основные определения

В то же время, согласно теореме 4.1 (no free lunch) существует вероятностная мера $\widehat{\mathsf{P}}_X \in \mathcal{M}^1_+(\mathsf{X}, 2^{\mathsf{X}})$ и концепт $C' \in 2^{\mathsf{X}}$ такие, что

$$\widehat{\mathsf{P}}^n_X\left\{\mathbf{x} \in \mathsf{X}^n \ : \ C'' = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}), R\big(\widehat{\mathsf{P}}_X; C', C''\big) > \frac{1}{8}\right\} \geqslant \frac{1}{7}.$$

Мы пришли к противоречию.

■

Модель PAC-обучения

Основные определения

Замечание

Ранее уже отмечалась эквивалентность задачи обучения концептов и задачи бинарной классификации.

Модель PAC-обучения

Основные определения

Замечание

Ранее уже отмечалась эквивалентность задачи обучения концептов и задачи бинарной классификации.

По большому счёту, вопросом удобства является рассмострение класса концептов или соответствующего ему класса характеристических функций. Поэтому понятие PAC-обучения может быть обобщено.

Модель PAC-обучения

Основные определения

Замечание

Ранее уже отмечалась эквивалентность задачи обучения концептов и задачи бинарной классификации.

По большому счёту, вопросом удобства является рассмострение класса концептов или соответствующего ему класса характеристических функций. Поэтому понятие PAC-обучения может быть обобщено.

По определению класс гипотез $\mathcal{H} \subseteq \{0,1\}^{\mathsf{X}}$ является PAC-обучаемым в том и только том случае, когда PAC-обучаем соответствующий ему класс концептов $\mathcal{C} \simeq_{\mathbf{1}} \mathcal{H}$.

Модель PAC-обучения

Основные определения

Замечание

Ранее уже отмечалась эквивалентность задачи обучения концептов и задачи бинарной классификации.

По большому счёту, вопросом удобства является рассмострение класса концептов или соответствующего ему класса характеристических функций. Поэтому понятие PAC-обучения может быть обобщено.

По определению класс гипотез $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ является PAC-обучаемым в том и только том случае, когда PAC-обучаем соответствующий ему класс концептов $\mathcal{C} \simeq_{\mathbf{1}} \mathcal{H}$.

При этом положим $n_{\mathcal{H}}^{\mathrm{pac}} := n_{\mathcal{C}}^{\mathrm{pac}}$.

Содержание

Модель PAC-обучения

Пример 1

В качестве множества объектов X возьмём $\mathbb{R}^2$.

Модель PAC-обучения
Пример 1

В качестве множества объектов X возьмём $\mathbb{R}^2$.

Класс концептов $\mathcal{C}_{\mathrm{rec}}$ будет состоять из всевозможных прямоугольников, стороны которых параллельны координатным осям.

Модель PAC-обучения

Пример 1

В качестве множества объектов X возьмём $\mathbb{R}^2$.

Класс концептов $\mathcal{C}_{\mathrm{rec}}$ будет состоять из всевозможных прямоугольников, стороны которых параллельны координатным осям.

Каждый такой прямоугольник имеет вид

$$R(a_1, b_1, a_2, b_2) := \{(u, v) : a_1 \leqslant u \leqslant b_1,\ a_2 \leqslant v \leqslant b_2\},$$

для некоторых $a_1, b_1, a_2, b_2 \in \mathbb{R}$.

Модель PAC-обучения

Пример 1

Рассмотрим алгоритм обучения $\mathcal{A}_{\mathrm{rec}} : \mathbf{z} \longmapsto C_{\mathbf{z}} := R(a_{\mathbf{z},1}, b_{\mathbf{z},1}, a_{\mathbf{z},2}, b_{\mathbf{z},2})$ ($\mathbf{z} \in \mathrm{Z}^*$), где

$$\begin{aligned} a_{\mathbf{z},1} &= \min_{((u,v),1)\in\mathbf{z}} u, & b_{\mathbf{z},1} &= \max_{((u,v),1)\in\mathbf{z}} u, \\ a_{\mathbf{z},2} &= \min_{((u,v),1)\in\mathbf{z}} v, & b_{\mathbf{z},2} &= \max_{((u,v),1)\in\mathbf{z}} v. \end{aligned}$$

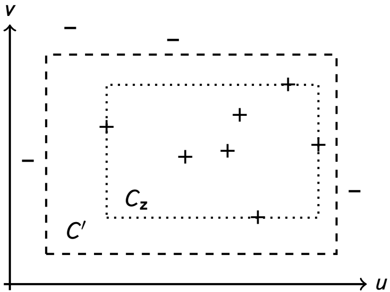
Модель PAC-обучения

Пример 1

Реализуемый случай с целевым концептом C'.

Обучающая выборка имеет вид $\mathbf{z} = \mathbf{1}_{C'} \circ \mathbf{x}$, где $\mathbf{x} \in \mathrm{X}^*$.

Символом $+$ обозначены точки из набора $\mathbf{x}$, принадлежащие C', а символом $-$ обозначены точки, не принадлежащие C'.



Модель PAC-обучения

Пример 1

Непосредственно из определения вытекает следующее свойство этого алгоритма.

Модель PAC-обучения

Пример 1

Непосредственно из определения вытекает следующее свойство этого алгоритма.

Утверждение 4.4.

Пусть $C' \in \mathcal{C}_{\mathrm{rec}}$ и $C_{\mathbf{x}} := \mathcal{A}_{\mathrm{rec}}(\mathbf{1}_{C'} \circ \mathbf{x})$ $(\mathbf{x} \in \mathrm{X}^*)$. Тогда имеет место включение $C_{\mathbf{x}} \subseteq C'$, а значит $C' \vartriangle C_{\mathbf{x}} = C' \setminus C_{\mathbf{x}}$.

Модель PAC-обучения

Пример 1

Непосредственно из определения вытекает следующее свойство этого алгоритма.

Утверждение 4.4.

Пусть $C' \in \mathcal{C}_{\mathrm{rec}}$ и $C_{\mathbf{x}} := \mathcal{A}_{\mathrm{rec}}(\mathbf{1}_{C'} \circ \mathbf{x})$ $(\mathbf{x} \in \mathrm{X}^*)$. Тогда имеет место включение $C_{\mathbf{x}} \subseteq C'$, а значит $C' \bigtriangleup C_{\mathbf{x}} = C' \setminus C_{\mathbf{x}}$.

Докажем PAC-обучаемость класса концептов $\mathcal{C}_{\mathrm{rec}}$.

Модель PAC-обучения

Пример 1

Теорема 4.3.

Класс концептов $\mathcal{C}_{\mathrm{rec}}$ является PAC-обучаемым с помощью алгоритма обучения $\mathcal{A}_{\mathrm{rec}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{rec}}}(\varepsilon, \delta) \leqslant \left\lceil \frac{4}{\varepsilon} \ln\left(\frac{4}{\delta}\right) \right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{2}$$

Модель PAC-обучения

Пример 1

Теорема 4.3.

Класс концептов $\mathcal{C}_{\mathrm{rec}}$ является PAC-обучаемым с помощью алгоритма обучения $\mathcal{A}_{\mathrm{rec}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{rec}}}(\varepsilon,\delta) \leqslant \left\lceil \frac{4}{\varepsilon} \ln\left(\frac{4}{\delta}\right) \right\rceil \qquad \big(\varepsilon,\delta \in (0,1)\big). \tag{2}$$

◀ Зафиксируем вероятностную меру $\mathsf{P}_X \in \mathcal{M}^1_+(\mathsf{X})$, целевой концепт $C' \in \mathcal{C}_{\mathrm{rec}}$, размер обучающих выборок $n \in \mathbb{N}$ и значения параметров $\varepsilon,\delta \in (0,1)$.

Будем использовать обозначения из предыдущего утв. 4.4.

Модель PAC-обучения
Пример 1

Пусть $C' = R(a_1, b_1, a_2, b_2)$ для некоторых $a_1, b_1, a_2, b_2 \in \mathbb{R}$. Введем вспомогательные прямоугольники

$$\begin{array}{rcl} C_1 & = & R_{a_1,b_1,\hat{a}_2,b_2}, \\ C_2 & = & R_{\hat{a}_1,b_1,a_2,b_2}, \\ C_3 & = & R_{a_1,b_1,a_2,\hat{b}_2}, \\ C_4 & = & R_{a_1,\hat{b}_1,a_2,b_2}, \end{array}$$

где числа $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2 \in \mathbb{R}$ выбираются из условий

Модель PAC-обучения
Пример 1

Пусть $C' = R(a_1, b_1, a_2, b_2)$ для некоторых $a_1, b_1, a_2, b_2 \in \mathbb{R}$. Введем вспомогательные прямоугольники

$$\begin{array}{rcl} C_1 & = & R_{a_1, b_1, \hat{a}_2, b_2}, \\ C_2 & = & R_{\hat{a}_1, b_1, a_2, b_2}, \\ C_3 & = & R_{a_1, b_1, a_2, \hat{b}_2}, \\ C_4 & = & R_{a_1, \hat{b}_1, a_2, b_2}, \end{array}$$

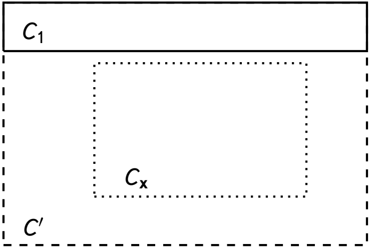
где числа $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2 \in \mathbb{R}$ выбираются из условий

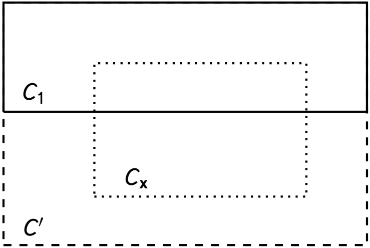
$$\mathrm{P}_X(C_i) = \frac{\varepsilon}{4} \quad (i = 1, 2, 3, 4).$$

Модель PAC-обучения

Пример 1

В качестве примера приведены варианты возможного расположения прямоугольника C_1 относительно прямоугольников C' и $C_{\mathbf{x}}$ ($\mathbf{x} \in \mathrm{X}^n$).





Модель PAC-обучения

Пример 1

Заметим, что из выполнения включения

$$C' \setminus C_{\mathbf{x}} \subseteq \bigcup_{i=1}^{4} C_i, \qquad (3)$$

Модель PAC-обучения

Пример 1

Заметим, что из выполнения включения

$$C' \setminus C_{\mathbf{x}} \subseteq \bigcup_{i=1}^{4} C_i, \tag{3}$$

сразу следует требуемая оценка для точности приближения

$$R(\mathrm{P}_X; C', C_{\mathbf{x}}) = \Big| \text{ Утв. 4.2, 4.4 } \Big| = \mathrm{P}_X(C' \setminus C_{\mathbf{x}}) \leqslant \sum_{i=1}^{4} \mathrm{P}_X(C_i) = \varepsilon.$$

Модель PAC-обучения

Пример 1

Заметим, что из выполнения включения

$$C' \setminus C_{\mathbf{x}} \subseteq \bigcup_{i=1}^{4} C_i, \tag{3}$$

сразу следует требуемая оценка для точности приближения

$$R(\mathrm{P}_X; C', C_{\mathbf{x}}) = \Big| \text{ Утв. 4.2, 4.4 } \Big| = \mathrm{P}_X(C' \setminus C_{\mathbf{x}}) \leqslant \sum_{i=1}^{4} \mathrm{P}_X(C_i) = \varepsilon.$$

Учитывая способ построения приближенного прямоугольника $C_{\mathbf{x}}$, условие (3) можно записать в эквивалентной форме

$$\mathbf{x} \cap C_i \neq \varnothing \quad \text{для всех } i \in \{1, 2, 3, 4\}.$$

Модель PAC-обучения

Пример 1

Но тогда, из предположения $\mathrm{P}_X\big(C' \setminus C_{\mathbf{x}}\big) > \varepsilon$ должно следовать, что

$$\mathbf{x} \cap C_i = \varnothing \qquad \text{для некоторого } i \in \{1, 2, 3, 4\}. \tag{4}$$

Модель PAC-обучения

Пример 1

Но тогда, из предположения $\mathrm{P}_X\big(C' \setminus C_{\mathbf{x}}\big) > \varepsilon$ должно следовать, что

$$\mathbf{x} \cap C_i = \varnothing \qquad \text{для некоторого } i \in \{1, 2, 3, 4\}. \tag{4}$$

Рассмотрим события

$$X_i := \big\{\mathbf{x} \in \mathrm{X}^n : \mathbf{x} \cap C_i = \varnothing\big\} \quad (i = 1, 2, 3, 4)$$

Модель PAC-обучения

Пример 1

Но тогда, из предположения $\mathrm{P}_X(C' \setminus C_{\mathbf{x}}) > \varepsilon$ должно следовать, что

$$\mathbf{x} \cap C_i = \varnothing \qquad \text{для некоторого } i \in \{1, 2, 3, 4\}. \tag{4}$$

Рассмотрим события

$$X_i := \{\mathbf{x} \in \mathrm{X}^n : \mathbf{x} \cap C_i = \varnothing\} \quad (i = 1, 2, 3, 4)$$

и оценим их вероятность

$$\mathrm{P}_X^n(X_i) = \left(1 - \frac{\varepsilon}{4}\right)^n \leqslant \left|\ \substack{1 + u \leqslant e^u \\ (u \in \mathbb{R})}\ \right| \leqslant e^{-\frac{n\varepsilon}{4}} \quad (i = 1, 2, 3, 4).$$

Модель PAC-обучения
Пример 1

Используя условие (4), запишем включение

$$\{\mathbf{x} \in \mathrm{X}^n : \mathrm{P}_X(C' \setminus C_{\mathbf{x}}) > \varepsilon\} \subseteq \bigcup_{i=1}^{4} X_i,$$

Модель PAC-обучения

Пример 1

Используя условие (4), запишем включение

$$\left\{ \mathbf{x} \in \mathrm{X}^n : \mathrm{P}_X(C' \setminus C_{\mathbf{x}}) > \varepsilon \right\} \subseteq \bigcup_{i=1}^{4} X_i,$$

из которого получим оценку для вероятности

$$\mathrm{P}_X^n \left\{ \mathbf{x} \in \mathrm{X}^n : \mathrm{P}_X\left(C' \setminus C_{\mathbf{x}}\right) > \varepsilon \right\} \leqslant \sum_{i=1}^{4} \mathrm{P}_X^n\left(X_i\right) \leqslant 4\, e^{-\frac{n\varepsilon}{4}}. \tag{5}$$

Модель PAC-обучения

Пример 1

Используя условие (4), запишем включение

$$\left\{\mathbf{x} \in \mathrm{X}^n : \mathrm{P}_X(C' \setminus C_{\mathbf{x}}) > \varepsilon\right\} \subseteq \bigcup_{i=1}^{4} X_i,$$

из которого получим оценку для вероятности

$$\mathrm{P}_X^n\left\{\mathbf{x} \in \mathrm{X}^n : \mathrm{P}_X\left(C' \setminus C_{\mathbf{x}}\right) > \varepsilon\right\} \leqslant \sum_{i=1}^{4} \mathrm{P}_X^n\left(X_i\right) \leqslant 4\, e^{-\frac{n\varepsilon}{4}}. \tag{5}$$

Ограничивая правую часть неравенства (5) числом δ, получим неравенство

$$4\, e^{-\frac{n\varepsilon}{4}} \leqslant \delta,$$

Модель PAC-обучения

Пример 1

Из его решения

$$n \geqslant \left\lceil \frac{4}{\varepsilon} \ln\left(\frac{4}{\delta}\right) \right\rceil$$

Модель PAC-обучения

Пример 1

Из его решения

$$n \geqslant \left\lceil \frac{4}{\varepsilon} \ln\left(\frac{4}{\delta}\right) \right\rceil$$

вытекает требуемая оценка (2).

■

Модель PAC-обучения

Пример 1

Замечание

Разобранный пример можно обобщить на произвольный m-мерный случай ($m \in \mathbb{N}$).

Модель PAC-обучения

Пример 1

Замечание

Разобранный пример можно обобщить на произвольный m-мерный случай ($m \in \mathbb{N}$).

В качестве концептов будут выступать m-мерные параллелепипеды вида

$$R(a_1, b_1, \ldots, a_m, b_m) := \{(u_1, \ldots, u_m) : a_1 \leqslant u_1 \leqslant b_1, \ldots, a_m \leqslant u_m \leqslant b_m\},$$

где $a_1, b_1, \ldots, a_m, b_m \in \mathbb{R}$.

Модель PAC-обучения

Пример 1

Замечание

Разобранный пример можно обобщить на произвольный m-мерный случай ($m \in \mathbb{N}$).

В качестве концептов будут выступать m-мерные параллелепипеды вида

$$R(a_1, b_1, \ldots, a_m, b_m) := \{(u_1, \ldots, u_m) : a_1 \leqslant u_1 \leqslant b_1, \ldots, a_m \leqslant u_m \leqslant b_m\},$$

где $a_1, b_1, \ldots, a_m, b_m \in \mathbb{R}$.

Для класса таких параллелепипедов существует PAC-учитель $\mathcal{A}_{\mathrm{par}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{par}}}(\varepsilon, \delta) \leqslant \left\lceil \frac{2m}{\varepsilon} \ln\left(\frac{2m}{\delta}\right) \right\rceil \qquad (\varepsilon, \delta \in (0, 1)).$$

Содержание

Модель PAC-обучения

Пример 2

Зафиксируем число $m \in \mathbb{N}$.

Модель PAC-обучения

Пример 2

Зафиксируем число $m \in \mathbb{N}$.

Через V обозначим алфавит литералов, состоящий из символов булевых переменных $\mathsf{v}_1, \dots, \mathsf{v}_m$ и символов их отрицаний $\overline{\mathsf{v}}_1, \dots, \overline{\mathsf{v}}_m$.

Модель PAC-обучения

Пример 2

Зафиксируем число $m \in \mathbb{N}$.

Через V обозначим алфавит литералов, состоящий из символов булевых переменных $\mathrm{v}_1, \ldots, \mathrm{v}_m$ и символов их отрицаний $\overline{\mathrm{v}}_1, \ldots, \overline{\mathrm{v}}_m$.

Будем использовать метаобозначения $\mathrm{v}_i^0 := \overline{\mathrm{v}}_i$ и $\mathrm{v}_i^1 := \mathrm{v}_i$ ($i = 1, \ldots, m$).

Модель PAC-обучения

Пример 2

Зафиксируем число $m \in \mathbb{N}$.

Через V обозначим алфавит литералов, состоящий из символов булевых переменных $\mathrm{v}_1, \ldots, \mathrm{v}_m$ и символов их отрицаний $\overline{\mathrm{v}}_1, \ldots, \overline{\mathrm{v}}_m$.

Будем использовать метаобозначения $\mathrm{v}_i^0 := \overline{\mathrm{v}}_i$ и $\mathrm{v}_i^1 := \mathrm{v}_i$ ($i = 1, \ldots, m$).

В качестве множества объектов X возьмём $\{0, 1\}^m$.

Модель PAC-обучения

Пример 2

Зафиксируем число $m \in \mathbb{N}$.

Через V обозначим алфавит литералов, состоящий из символов булевых переменных $\mathrm{v}_1, \ldots, \mathrm{v}_m$ и символов их отрицаний $\overline{\mathrm{v}}_1, \ldots, \overline{\mathrm{v}}_m$.

Будем использовать метаобозначения $\mathrm{v}_i^0 := \overline{\mathrm{v}}_i$ и $\mathrm{v}_i^1 := \mathrm{v}_i$ ($i = 1, \ldots, m$).

В качестве множества объектов X возьмём $\{0, 1\}^m$.

Множество концептов $\mathcal{C}_m$ состоит из всех таких подмножеств $C \subset X$, у которых характеристическая функция может быть представлена в виде конъюнкций, составленных из символов переменных или их отрицаний,

Модель PAC-обучения

Пример 2

Зафиксируем число $m \in \mathbb{N}$.

Через V обозначим алфавит литералов, состоящий из символов булевых переменных $\mathrm{v}_1, \ldots, \mathrm{v}_m$ и символов их отрицаний $\overline{\mathrm{v}}_1, \ldots, \overline{\mathrm{v}}_m$.

Будем использовать метаобозначения $\mathrm{v}_i^0 := \overline{\mathrm{v}}_i$ и $\mathrm{v}_i^1 := \mathrm{v}_i$ $(i = 1, \ldots, m)$.

В качестве множества объектов X возьмём $\{0, 1\}^m$.

Множество концептов $\mathcal{C}_m$ состоит из всех таких подмножеств $C \subset X$, у которых характеристическая функция может быть представлена в виде конъюнкций, составленных из символов переменных или их отрицаний, а именно

$$\mathbf{1}_C(\mathrm{v}_1, \ldots, \mathrm{v}_m) = \mathrm{v}_{i_1}^{\sigma_1} \wedge \ldots \wedge \mathrm{v}_{i_k}^{\sigma_k} \quad (i_1 < \ldots < i_k;\ 1 \leqslant k \leqslant m).$$

Модель PAC-обучения

Пример 2

По определению положим $V_C := \{v_{i_1}^{\sigma_1}, \dots, v_{i_k}^{\sigma_k}\}$.

Модель PAC-обучения

Пример 2

По определению положим $V_C := \{v_{i_1}^{\sigma_1}, \ldots, v_{i_k}^{\sigma_k}\}$.

Утверждение 4.5.

Мощность множества концептов $|\mathcal{C}_m| = 3^m - 1$.

Модель PAC-обучения

Пример 2

По определению положим $\mathrm{V}_C := \{\mathrm{v}_{i_1}^{\sigma_1}, \ldots, \mathrm{v}_{i_k}^{\sigma_k}\}$.

Утверждение 4.5.

Мощность множества концептов $|\mathcal{C}_m| = 3^m - 1$.

◀ Каждый концепт $C \in \mathcal{C}_m$ однозначно задается своим множеством V_C.

Модель PAC-обучения

Пример 2

По определению положим $\mathrm{V}_C := \{\mathrm{v}_{i_1}^{\sigma_1}, \ldots, \mathrm{v}_{i_k}^{\sigma_k}\}$.

Утверждение 4.5.

Мощность множества концептов $|\mathcal{C}_m| = 3^m - 1$.

◀ Каждый концепт $C \in \mathcal{C}_m$ однозначно задается своим множеством V_C.

Для каждого i $(i = 1, \ldots, m)$ возможны три варианта.

Модель PAC-обучения

Пример 2

По определению положим $\mathrm{V}_C := \{\mathrm{v}_{i_1}^{\sigma_1}, \ldots, \mathrm{v}_{i_k}^{\sigma_k}\}$.

Утверждение 4.5.

Мощность множества концептов $|\mathcal{C}_m| = 3^m - 1$.

◀ Каждый концепт $C \in \mathcal{C}_m$ однозначно задается своим множеством V_C.

Для каждого i $(i = 1, \ldots, m)$ возможны три варианта.

Множество V_C может содержать только один из литералов v_i или $\overline{\mathrm{v}}_i$, или не содержать ни одного из них.

Модель PAC-обучения

Пример 2

По определению положим $V_C := \{v_{i_1}^{\sigma_1}, \ldots, v_{i_k}^{\sigma_k}\}$.

Утверждение 4.5.

Мощность множества концептов $|\mathcal{C}_m| = 3^m - 1$.

◀ Каждый концепт $C \in \mathcal{C}_m$ однозначно задается своим множеством V_C.

Для каждого i $(i = 1, \ldots, m)$ возможны три варианта.

Множество V_C может содержать только один из литералов v_i или $\overline{v}_i$, или не содержать ни одного из них.

Учитывая, что множество V_C не может быть пустым, получим $|\mathcal{C}_m| = 3^m - 1$.

■

Модель PAC-обучения

Пример 2

Рассмотрим алгоритм обучения $\mathcal{A}_{\mathrm{con}} : \mathbf{z} \longmapsto C_{\mathbf{z}}$ ($\mathbf{z} \in Z^*$), использующий следующее правило.

Модель PAC-обучения

Пример 2

Рассмотрим алгоритм обучения $\mathcal{A}_{\mathrm{con}} : \mathbf{z} \longmapsto C_{\mathbf{z}}$ ($\mathbf{z} \in \mathrm{Z}^*$), использующий следующее правило.

Литерал $\mathrm{v} \in \mathrm{V}$ включается в $\mathrm{V}_{C_{\mathbf{z}}}$ в том и только том случае, когда для любого примера вида $(x, 1) \in \mathbf{z}$ выполняется условие $\mathrm{v}(x) = 1$.

Модель PAC-обучения
Пример 2

Рассмотрим алгоритм обучения $\mathcal{A}_{\mathrm{con}} : \mathbf{z} \longmapsto C_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^*)$, использующий следующее правило.

Литерал $\mathsf{v} \in \mathsf{V}$ включается в $\mathsf{V}_{C_{\mathbf{z}}}$ в том и только том случае, когда для любого примера вида $(x, 1) \in \mathbf{z}$ выполняется условие $\mathsf{v}(x) = 1$.

Например, по набору примеров $\big(((0, 1, 1, 0), 0), ((1, 0, 1, 0), 1), ((1, 0, 1, 1), 1)\big)$ будет построена конъюнкция $\mathsf{v}_1 \wedge \overline{\mathsf{v}}_2 \wedge \mathsf{v}_3$.

Модель PAC-обучения

Пример 2

Рассмотрим алгоритм обучения $\mathcal{A}_{\mathrm{con}} : \mathbf{z} \longmapsto C_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^*)$, использующий следующее правило.

Литерал $\mathsf{v} \in \mathsf{V}$ включается в $\mathsf{V}_{C_{\mathbf{z}}}$ в том и только том случае, когда для любого примера вида $(x, 1) \in \mathbf{z}$ выполняется условие $\mathsf{v}(x) = 1$.

Например, по набору примеров $\big(((0,1,1,0),0),((1,0,1,0),1),((1,0,1,1),1)\big)$ будет построена конъюнкция $\mathsf{v}_1 \wedge \overline{\mathsf{v}}_2 \wedge \mathsf{v}_3$.

Непосредственно из определения вытекает следующее свойство этого алгоритма.

Модель PAC-обучения
Пример 2

Рассмотрим алгоритм обучения $\mathcal{A}_{\mathrm{con}} : \mathbf{z} \longmapsto C_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^*)$, использующий следующее правило.

Литерал $\mathsf{v} \in \mathsf{V}$ включается в $\mathsf{V}_{C_{\mathbf{z}}}$ в том и только том случае, когда для любого примера вида $(x, 1) \in \mathbf{z}$ выполняется условие $\mathsf{v}(x) = 1$.

Например, по набору примеров $\big(((0,1,1,0),0),((1,0,1,0),1),((1,0,1,1),1)\big)$ будет построена конъюнкция $\mathsf{v}_1 \wedge \overline{\mathsf{v}}_2 \wedge \mathsf{v}_3$.

Непосредственно из определения вытекает следующее свойство этого алгоритма.

Утверждение 4.6.

Пусть $C' \in \mathcal{C}_m$ и $C_{\mathbf{x}} := \mathcal{A}_{\mathrm{con}}(\mathbf{1}_{C'} \circ \mathbf{x})$ $(\mathbf{x} \in \mathsf{X}^*)$. Тогда имеет место включение $\mathsf{V}_{C'} \subseteq \mathsf{V}_{C_{\mathbf{x}}}$, а значит $C_{\mathbf{x}} \subseteq C'$.

Модель PAC-обучения

Пример 2

Теорема 4.4.

Класс концептов $\mathcal{C}_m$ является PAC-обучаемым с помощью алгоритма обучения $\mathcal{A}_{\mathrm{con}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{con}}}(\varepsilon, \delta) \leqslant \left\lceil \frac{m}{\varepsilon} \ln\left(\frac{2m}{\delta}\right)\right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{6}$$

Модель PAC-обучения

Пример 2

Теорема 4.4.

Класс концептов $\mathcal{C}_m$ является PAC-обучаемым с помощью алгоритма обучения $\mathcal{A}_{\mathrm{con}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{con}}}(\varepsilon, \delta) \leqslant \left\lceil \frac{m}{\varepsilon} \ln\left(\frac{2m}{\delta}\right) \right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{6}$$

◀ Зафиксируем вероятностную меру $\mathsf{P}_X \in \mathcal{M}^1_+(\mathsf{X})$, целевой концепт $C' \in \mathcal{C}_m$,

Модель PAC-обучения

Пример 2

Теорема 4.4.

Класс концептов $\mathcal{C}_m$ является PAC-обучаемым с помощью алгоритма обучения $\mathcal{A}_{\mathrm{con}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{con}}}(\varepsilon, \delta) \leqslant \left\lceil \frac{m}{\varepsilon} \ln\left(\frac{2m}{\delta}\right) \right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{6}$$

◀ Зафиксируем вероятностную меру $\mathsf{P}_X \in \mathcal{M}^1_+(\mathsf{X})$, целевой концепт $C' \in \mathcal{C}_m$,

размер обучающих выборок $n \in \mathbb{N}$ и значения параметров $\varepsilon, \delta \in (0, 1)$.

Модель PAC-обучения

Пример 2

Теорема 4.4.

Класс концептов $\mathcal{C}_m$ является PAC-обучаемым с помощью алгоритма обучения $\mathcal{A}_{\mathrm{con}}$, для которого имеет место оценка

$$n^{\mathrm{pac}}_{\mathcal{A}_{\mathrm{con}}}(\varepsilon, \delta) \leqslant \left\lceil \frac{m}{\varepsilon} \ln\left(\frac{2m}{\delta}\right) \right\rceil \qquad (\varepsilon, \delta \in (0,1)). \tag{6}$$

◀ Зафиксируем вероятностную меру $\mathsf{P}_X \in \mathcal{M}^1_+(\mathsf{X})$, целевой концепт $C' \in \mathcal{C}_m$,

размер обучающих выборок $n \in \mathbb{N}$ и значения параметров $\varepsilon, \delta \in (0,1)$.

Будем использовать обозначения из предыдущего утв. 4.6.

Модель PAC-обучения
Пример 2

Для каждого литерала $\mathsf{v} \in \mathsf{V}$ определим множество объектов

$$A_{\mathsf{v}} := \{x : x \in C' \text{ и } \mathsf{v}(x) = 0\}.$$

Модель PAC-обучения
Пример 2

Для каждого литерала $\mathsf{v} \in \mathsf{V}$ определим множество объектов

$$A_{\mathsf{v}} := \{x : x \in C' \text{ и } \mathsf{v}(x) = 0\}.$$

Заметим, что если для любого литерала $\mathsf{v} \in \mathsf{V}_{C_{\mathbf{x}}}$ справедливо неравенство $\mathsf{P}_X(A_{\mathsf{v}}) \leqslant \frac{\varepsilon}{m}$, то будет выполняться требуемая оценка точности приближения

$$\begin{aligned} R(\mathsf{P}_X; C', C_{\mathbf{x}}) &= \left| \text{ Утв. 4.2, 4.6 } \right| = \mathsf{P}_X(C' \setminus C_{\mathbf{x}}) \leqslant \sum_{\mathsf{v} \in \mathsf{V}_{C_{\mathbf{x}}}} \mathsf{P}_X(A_{\mathsf{v}}) \\ &\leqslant \sum_{\mathsf{v} \in \mathsf{V}_{C_{\mathbf{x}}}} \frac{\varepsilon}{m} \leqslant m \cdot \frac{\varepsilon}{m} = \varepsilon \qquad (\mathbf{x} \in \mathsf{X}^n). \end{aligned} \tag{7}$$

Модель PAC-обучения

Пример 2

Литерал $\mathsf{b} \in \mathsf{V}$ назовем «плохим», если $\mathsf{P}(A_{\mathsf{b}}) > \frac{\varepsilon}{m}$.

Модель PAC-обучения
Пример 2

Литерал b $\in$ V назовем «плохим», если $\mathrm{P}(A_{\mathrm{b}}) > \frac{\varepsilon}{m}$.

Множество всех плохих литералов обозначим через B.

Модель PAC-обучения

Пример 2

Литерал $\mathrm{b} \in \mathrm{V}$ назовем «плохим», если $\mathrm{P}(A_{\mathrm{b}}) > \frac{\varepsilon}{m}$.

Множество всех плохих литералов обозначим через B.

Из (7) следует, что если $R\left(\mathrm{P}_X; C', C_{\mathbf{x}}\right) > \varepsilon$, то множество $\mathrm{V}_{C_{\mathbf{x}}}$ должно содержать хотя бы один плохой литерал $\mathrm{b} \in \mathrm{B}$.

Модель PAC-обучения
Пример 2

Литерал $\mathrm{b} \in \mathrm{V}$ назовем «плохим», если $\mathrm{P}(A_{\mathrm{b}}) > \frac{\varepsilon}{m}$.

Множество всех плохих литералов обозначим через B.

Из (7) следует, что если $R\left(\mathrm{P}_X; C', C_{\mathbf{x}}\right) > \varepsilon$, то множество $\mathrm{V}_{C_{\mathbf{x}}}$ должно содержать хотя бы один плохой литерал $\mathrm{b} \in \mathrm{B}$.

Учитывая способ построения $C_{\mathbf{x}}$, получим $\mathbf{x} \cap A_{\mathrm{b}} = \varnothing$.

Модель PAC-обучения

Пример 2

Литерал $\mathsf{b} \in \mathsf{V}$ назовем «плохим», если $\mathsf{P}(A_{\mathsf{b}}) > \frac{\varepsilon}{m}$.

Множество всех плохих литералов обозначим через B.

Из (7) следует, что если $R\big(\mathsf{P}_X; C', C_{\mathbf{x}}\big) > \varepsilon$, то множество $\mathsf{V}_{C_{\mathbf{x}}}$ должно содержать хотя бы один плохой литерал $\mathsf{b} \in \mathsf{B}$.

Учитывая способ построения $C_{\mathbf{x}}$, получим $\mathbf{x} \cap A_{\mathsf{b}} = \varnothing$.

Таким образом,

$$\begin{aligned} \big\{\mathbf{x} \in \mathsf{X}^n \,:\, R\big(\mathsf{P}_X; C', C_{\mathbf{x}}\big) > \varepsilon\big\} &\subseteq \bigcup_{\mathsf{b} \in \mathsf{B}} \big\{\mathbf{x} \in \mathsf{X}^n \,:\, \mathsf{b} \in \mathsf{V}_{C_{\mathbf{x}}}\big\} \\ &\subseteq \bigcup_{\mathsf{b} \in \mathsf{B}} \big\{\mathbf{x} \in \mathsf{X}^n \,:\, \mathbf{x} \cap A_{\mathsf{b}} = \varnothing\big\}. \end{aligned}$$

Модель PAC-обучения

Пример 2

Заметим, что для любого литерала $\mathrm{v} \in \mathrm{V}$ справедливо

$$\mathrm{P}_X^n\left\{\mathbf{x} \in \mathrm{X}^n \;:\; \mathbf{x} \cap A_{\mathrm{v}} = \varnothing\right\} = \left(1 - \mathrm{P}_X(A_{\mathrm{v}})\right)^n,$$

Модель PAC-обучения

Пример 2

Заметим, что для любого литерала $\mathsf{v} \in \mathsf{V}$ справедливо

$$\mathsf{P}_X^n\left\{\mathbf{x} \in \mathsf{X}^n \;:\; \mathbf{x} \cap A_{\mathsf{v}} = \varnothing\right\} = \left(1 - \mathsf{P}_X(A_{\mathsf{v}})\right)^n,$$

но тогда

$$\begin{aligned} \mathsf{P}_X^n\left\{\mathbf{x} \in \mathsf{X}^n \;:\; R\left(\mathsf{P}_X; C', C_{\mathbf{x}}\right) > \varepsilon\right\} &\leqslant \sum_{\mathsf{b} \in \mathsf{B}} \left(1 - \mathsf{P}_X(A_{\mathsf{b}})\right)^n \\ &< |\mathsf{B}| \cdot \left(1 - \frac{\varepsilon}{m}\right)^n \\ &\leqslant 2m \cdot \left(1 - \frac{\varepsilon}{m}\right)^n \\ &\leqslant \left| \begin{array}{c} 1 + u \leqslant e^u \\ (u \in \mathbb{R}) \end{array} \right| \leqslant 2m \cdot e^{-\frac{n\varepsilon}{m}}. \end{aligned} \tag{8}$$

Модель PAC-обучения
Пример 2

Ограничивая правую часть неравенства (8) числом δ, получим неравенство

$$2m \cdot e^{-\frac{n\varepsilon}{m}} \leqslant \delta.$$

Модель PAC-обучения

Пример 2

Ограничивая правую часть неравенства (8) числом δ, получим неравенство

$$2m \cdot e^{-\frac{n\varepsilon}{m}} \leqslant \delta.$$

Из его решения

$$n \geqslant \left\lceil \frac{m}{\varepsilon} \ln \left(\frac{2m}{\delta} \right) \right\rceil$$

вытекает требуемая оценка (6).

■

Содержание

Модель PAC-обучения

Конечный случай

Прежде всего заметим, что в рамках задачи обучения концептов (задачи бинарной классификации) всегда существует метод минимизации эмпирического риска.

Модель PAC-обучения

Конечный случай

Прежде всего заметим, что в рамках задачи обучения концептов (задачи бинарной классификации) всегда существует метод минимизации эмпирического риска.

Используемая функция потерь l_{01} может принимать только два значения 0 или 1.

Модель PAC-обучения

Конечный случай

Прежде всего заметим, что в рамках задачи обучения концептов (задачи бинарной классификации) всегда существует метод минимизации эмпирического риска.

Используемая функция потерь l_{01} может принимать только два значения 0 или 1.

Поэтому эмпирический риск на обучающих выборках фиксированного размера также принимает конечное число значений.

Модель PAC-обучения

Конечный случай

Прежде всего заметим, что в рамках задачи обучения концептов (задачи бинарной классификации) всегда существует метод минимизации эмпирического риска.

Используемая функция потерь l_{01} может принимать только два значения 0 или 1.

Поэтому эмпирический риск на обучающих выборках фиксированного размера также принимает конечное число значений.

Это в свою очередь означает, что для заданной обучающей выборки всегда можно выбрать концепт (гипотезу) с минимальным эмпирическим риском.

Модель PAC-обучения

Конечный случай

Теорема 4.5.

Конечный класс концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$, $|\mathcal{C}| < \infty$ является PAC-обучаемым с помощью метода минимизации эмпирического риска и

$$n_{\mathcal{C}}^{\mathrm{pac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{1}{\varepsilon} \ln \left(\frac{|\mathcal{C}|}{\delta} \right) \right\rceil \qquad \bigl(\varepsilon, \delta \in (0, 1)\bigr). \tag{9}$$

Модель PAC-обучения

Конечный случай

Теорема 4.5.

Конечный класс концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$, $|\mathcal{C}| < \infty$ является PAC-обучаемым с помощью метода минимизации эмпирического риска и

$$n_{\mathcal{C}}^{\mathrm{pac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{1}{\varepsilon} \ln\left(\frac{|\mathcal{C}|}{\delta}\right) \right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{9}$$

◀ Зафиксируем вероятностную меру $\mathsf{P}_X \in \mathcal{M}_+^1(\mathsf{X})$, целевой концепт $C' \in \mathcal{C}$, размер обучающих выборок $n \in \mathbb{N}$ и значения параметров $\varepsilon, \delta \in (0, 1)$.

Модель PAC-обучения

Конечный случай

Теорема 4.5.

Конечный класс концептов $\mathcal{C} \subseteq 2^{\mathsf{X}}$, $|\mathcal{C}| < \infty$ является PAC-обучаемым с помощью метода минимизации эмпирического риска и

$$n_{\mathcal{C}}^{\mathrm{pac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{1}{\varepsilon} \ln\left(\frac{|\mathcal{C}|}{\delta}\right) \right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{9}$$

◀ Зафиксируем вероятностную меру $\mathsf{P}_X \in \mathcal{M}^1_+(\mathsf{X})$, целевой концепт $C' \in \mathcal{C}$, размер обучающих выборок $n \in \mathbb{N}$ и значения параметров $\varepsilon, \delta \in (0, 1)$.

Обозначим через $\mathcal{A}$ произвольный метод минимизации эмпирического риска для класса концептов $\mathcal{C}$.

Модель PAC-обучения
Конечный случай

Нашей основной целью является оценка вероятности следующего события

$$X_{\mathrm{bad}} := \left\{\mathbf{x} \in \mathsf{X}^n \,:\, C = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}),\, R(\mathsf{P}_X; C', C) > \varepsilon\right\}.$$

Модель PAC-обучения

Конечный случай

Нашей основной целью является оценка вероятности следующего события

$$X_{\text{bad}} := \left\{\mathbf{x} \in \mathsf{X}^n \;:\; C = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}),\ R(\mathsf{P}_X; C', C) > \varepsilon\right\}.$$

Каждому концепту $C \in \mathcal{C}$ поставим в соответствие множество наборов объектов

$$X_C := \left\{\mathbf{x} \in \mathsf{X}^n \;:\; r(l_{01}; \mathbf{1}_C, \mathbf{1}_{C'} \circ \mathbf{x}) = 0\right\}$$

Модель PAC-обучения

Конечный случай

Нашей основной целью является оценка вероятности следующего события

$$X_{\mathrm{bad}} := \left\{\mathbf{x} \in \mathsf{X}^n \,:\, C = \mathcal{A}(\mathbf{1}_{C'} \circ \mathbf{x}),\, R(\mathsf{P}_X; C', C) > \varepsilon\right\}.$$

Каждому концепту $C \in \mathcal{C}$ поставим в соответствие множество наборов объектов

$$X_C := \left\{\mathbf{x} \in \mathsf{X}^n \,:\, r(l_{01}; \mathbf{1}_C, \mathbf{1}_{C'} \circ \mathbf{x}) = 0\right\}$$

и заметим, что алгоритм обучения $\mathcal{A}$ может для обучающей выборки вида $\mathbf{1}_{C'} \circ \mathbf{x}$ ($\mathbf{x} \in \mathsf{X}^n$) вернуть концепт C только, если $\mathbf{x} \in X_C$.

Модель PAC-обучения

Конечный случай

Выделим подмножество концептов

$$\mathcal{C}_{\mathrm{bad}} := \{C \in \mathcal{C} \,:\, R(\mathsf{P}_X; C', C) > \varepsilon\},$$

которые «плохо» приближают целевой концепт C'.

Модель PAC-обучения

Конечный случай

Выделим подмножество концептов

$$\mathcal{C}_{\mathrm{bad}} := \{C \in \mathcal{C} \ : \ R(\mathrm{P}_X; C', C) > \varepsilon\},$$

которые «плохо» приближают целевой концепт C'.

Имеет место включение

$$X_{\mathrm{bad}} \subseteq \bigcup_{C \in \mathcal{C}_{\mathrm{bad}}} X_C. \qquad (10)$$

Модель PAC-обучения

Конечный случай

Выделим подмножество концептов

$$\mathcal{C}_{\mathrm{bad}} := \{C \in \mathcal{C} \,:\, R(\mathsf{P}_X; C', C) > \varepsilon\},$$

которые «плохо» приближают целевой концепт C'.

Имеет место включение

$$X_{\mathrm{bad}} \subseteq \bigcup_{C \in \mathcal{C}_{\mathrm{bad}}} X_C. \tag{10}$$

Для каждого концепта $C \in \mathcal{C}_{\mathrm{bad}}$ выполняется неравенство

$$\mathsf{P}_X\{x \in \mathsf{X} \,:\, \mathbf{1}_{C'}(x) = \mathbf{1}_C(x)\} = 1 - R(\mathsf{P}_X; C', C) < 1 - \varepsilon,$$

Модель PAC-обучения

Конечный случай

Выделим подмножество концептов

$$\mathcal{C}_{\mathrm{bad}} := \left\{C \in \mathcal{C} \,:\, R(\mathsf{P}_X; C', C) > \varepsilon\right\},$$

которые «плохо» приближают целевой концепт C'.

Имеет место включение

$$X_{\mathrm{bad}} \subseteq \bigcup_{C \in \mathcal{C}_{\mathrm{bad}}} X_C. \tag{10}$$

Для каждого концепта $C \in \mathcal{C}_{\mathrm{bad}}$ выполняется неравенство

$$\mathsf{P}_X\left\{x \in \mathsf{X} \,:\, \mathbf{1}_{C'}(x) = \mathbf{1}_C(x)\right\} = 1 - R(\mathsf{P}_X; C', C) < 1 - \varepsilon,$$

а значит

$$\mathsf{P}_X^n(X_C) = \prod_{i=1}^{n} \mathsf{P}_X\left\{x_i \in \mathsf{X} \,:\, \mathbf{1}_{C'}(x_i) = \mathbf{1}_C(x_i)\right\} < (1 - \varepsilon)^n. \tag{11}$$

Модель PAC-обучения

Конечный случай

Объединяя (10) и (11), получим

$$\mathsf{P}_X^n\left(X_{\text{bad}}\right) \leqslant \sum_{C \in \mathcal{C}_{\text{bad}}} \mathsf{P}_X^n\left(X_C\right) < |\mathcal{C}|(1-\varepsilon)^n \leqslant \Big|_{\substack{1+u \leqslant e^u \\ (u \in \mathbb{R})}} \Big| \leqslant |\mathcal{C}|\, \mathrm{e}^{-n\varepsilon}. \qquad (12)$$

Объединяя (10) и (11), получим

$$\mathsf{P}_X^n(X_{\text{bad}}) \leqslant \sum_{c \in \mathcal{C}_{\text{bad}}} \mathsf{P}_X^n(X_c) < |\mathcal{C}|(1-\varepsilon)^n \leqslant \left| \begin{smallmatrix} 1+u \leqslant e^u \\ (u \in \mathbb{R}) \end{smallmatrix} \right| \leqslant |\mathcal{C}|\, e^{-n\varepsilon}. \qquad (12)$$

Ограничим правую часть неравенства (12) значением δ. Получившееся неравенство

$$|\mathcal{C}|\, e^{-n\varepsilon} \leqslant \delta,$$

Модель PAC-обучения

Конечный случай

Объединяя (10) и (11), получим

$$\mathrm{P}_X^n\left(X_{\mathrm{bad}}\right) \leqslant \sum_{C \in \mathcal{C}_{\mathrm{bad}}} \mathrm{P}_X^n\left(X_C\right) < |\mathcal{C}|(1-\varepsilon)^n \leqslant \left| \begin{smallmatrix} 1+u \leqslant e^u \\ (u \in \mathbb{R}) \end{smallmatrix} \right| \leqslant |\mathcal{C}|\, e^{-n\varepsilon}. \qquad (12)$$

Ограничим правую часть неравенства (12) значением δ. Получившееся неравенство

$$|\mathcal{C}|\, e^{-n\varepsilon} \leqslant \delta,$$

имеет решение

$$n \geqslant \left\lceil \frac{1}{\varepsilon} \ln\left(\frac{|\mathcal{C}|}{\delta}\right) \right\rceil. \qquad (13)$$

Модель PAC-обучения

Конечный случай

Учитывая произвольность выбора вероятностной меры P_X и целевого концепта C', получим оценку (9), а значит и PAC-обучаемость класса концептов $\mathcal{C}$.

$\blacksquare$

Содержание

1. Модель PAC-обучения

2. Обобщение модели PAC-обучения
   - Основные определения
   - Интерпретация модели PAC-обучения

3. Равномерная сходимость эмпирического риска

Обобщение модели PAC-обучения

Перейдём к рассмотрению обобщения модели PAC-обучения, которое не накладывает исходных ограничений на множество меток и выбор используемой функции потерь.

Обобщение модели PAC-обучения

Перейдём к рассмотрению обобщения модели PAC-обучения, которое не накладывает исходных ограничений на множество меток и выбор используемой функции потерь.

Отсутствует также предположение о реализуемости, что означает необходимость использования семейств вероятностных мер, заданных на множестве примеров, а не на множестве объектов.

Содержание

Обобщение модели PAC-обучения

Основные определения

Определение 4.4.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ называется агностически PAC-обучаемым относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l, если для него существует алгоритм обучения $\mathcal{A} : \mathbf{z} \longmapsto h_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^*)$, обладающий следующим свойством.

Обобщение модели PAC-обучения

Основные определения

Определение 4.4.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ называется агностически PAC-обучаемым относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l, если для него существует алгоритм обучения $\mathcal{A} : \mathbf{z} \longmapsto h_{\mathbf{z}}$ ($\mathbf{z} \in \mathsf{Z}^*$), обладающий следующим свойством.
Для любого $\varepsilon \in (0, 1)$ выполняется условие

$$\lim_{n\to\infty} \bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) = 0,$$

Обобщение модели PAC-обучения

Основные определения

Определение 4.4.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ называется агностически PAC-обучаемым относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_+^1(\mathsf{Z})$ и функции потерь l, если для него существует алгоритм обучения $\mathcal{A} : \mathbf{z} \longmapsto h_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^*)$, обладающий следующим свойством.
Для любого $\varepsilon \in (0, 1)$ выполняется условие

$$\lim_{n \to \infty} \bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) = 0,$$

где

$$\bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) := \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n \big\{ \mathbf{z} \in \mathsf{Z}^n \; : \; R(\mathsf{P}, l; h_{\mathbf{z}}) > R(\mathsf{P}, l; \mathcal{H}) + \varepsilon \big\}.$$

При этом алгоритм $\mathcal{A}$ называется агностическим PAC-учителем для класса $\mathcal{H}$.

Обобщение модели PAC-обучения

Основные определения

Как и ранее, особый интерес представляют два семейства вероятностных мер.

Обобщение модели PAC-обучения

Основные определения

Как и ранее, особый интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}$ выбирается $\mathcal{M}^1_+(\mathsf{Z}, \mathcal{S}_\mathsf{Z})$, где $\mathcal{S}_\mathsf{Z}$ – некоторая фиксированная σ-алгебра.

Обобщение модели PAC-обучения

Основные определения

Как и ранее, особый интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}$ выбирается $\mathcal{M}^1_+(\mathsf{Z}, \mathcal{S}_\mathsf{Z})$, где $\mathcal{S}_\mathsf{Z}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса гипотез $\mathcal{H}$, семейство $\mathcal{P}$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$, где $\mathcal{H} \simeq_l \mathcal{F}$, является P-измеримой.

Обобщение модели PAC-обучения

Основные определения

Как и ранее, особый интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}$ выбирается $\mathcal{M}_+^1(\mathsf{Z}, \mathcal{S}_\mathsf{Z})$, где $\mathcal{S}_\mathsf{Z}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса гипотез $\mathcal{H}$, семейство $\mathcal{P}$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$, где $\mathcal{H} \simeq_l \mathcal{F}$, является P-измеримой.

Если явно не оговорено противное, то всегда будет неявно подразумеваться второй случай.

Обобщение модели PAC-обучения

Основные определения

Как и ранее, особый интерес представляют два семейства вероятностных мер.

В первом случае, в качестве $\mathcal{P}$ выбирается $\mathcal{M}^1_+(\mathsf{Z}, \mathcal{S}_\mathsf{Z})$, где $\mathcal{S}_\mathsf{Z}$ – некоторая фиксированная σ-алгебра.

Во втором случае, отталкиваясь от рассматриваемого класса гипотез $\mathcal{H}$, семейство $\mathcal{P}$ образуют все вероятностные меры P такие, что любая функция из $\mathcal{F}$, где $\mathcal{H} \simeq_l \mathcal{F}$, является P-измеримой.

Если явно не оговорено противное, то всегда будет неявно подразумеваться второй случай.

Как и в случае модели PAC-обучения определим понятия сложности выборки алгоритма обучения и класса гипотез.

Обобщение модели PAC-обучения

Основные определения

Определение 4.5.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l называется наименьшая функция вида

$$n^{\mathrm{apac}}_{\mathcal{A},\mathcal{P},l} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Определение 4.5.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathrm{Z})$ и функции потерь l называется наименьшая функция вида

$$n^{\mathrm{apac}}_{\mathcal{A},\mathcal{P},l} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$\bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n^{\mathrm{apac}}_{\mathcal{A},\mathcal{P},l}(\varepsilon, \delta).$$

Обобщение модели PAC-обучения

Основные определения

Определение 4.5.

Сложностью выборки алгоритма обучения $\mathcal{A}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathrm{Z})$ и функции потерь l называется наименьшая функция вида

$$n^{\mathrm{apac}}_{\mathcal{A},\mathcal{P},l} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$\bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n^{\mathrm{apac}}_{\mathcal{A},\mathcal{P},l}(\varepsilon, \delta).$$

Если понятно о каком семействе вероятностных мер и функции потерь идёт речь, то будет использоваться сокращённое обозначение $n^{\mathrm{apac}}_{\mathcal{A}}$.

Обобщение модели PAC-обучения

Основные определения

Из приведённого определения следует, что алгоритм обучения $\mathcal{A}$ является агностическим PAC-учителем тогда и только тогда, когда функция $n_{\mathcal{A}}^{\mathrm{apac}}$ принимает только конечные значения.

Обобщение модели PAC-обучения

Основные определения

Определение 4.6.

Сложностью выборки класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, называется наименьшая функция вида

$$n_{\mathcal{H},\mathcal{P},l}^{\mathrm{apac}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Обобщение модели PAC-обучения

Основные определения

Определение 4.6.

Сложностью выборки класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, называется наименьшая функция вида

$$n_{\mathcal{H},\mathcal{P},l}^{\mathrm{apac}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых значений $\varepsilon, \delta \in (0,1)$ и любого метода минимизации эмпирического риска $\mathcal{A}$ выполняется условие

$$n_{\mathcal{A},\mathcal{P},l}^{\mathrm{apac}}(\varepsilon,\delta) \leqslant n_{\mathcal{H},\mathcal{P},l}^{\mathrm{apac}}(\varepsilon,\delta).$$

Обобщение модели PAC-обучения

Основные определения

Определение 4.6.

Сложностью выборки класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, называется наименьшая функция вида

$$n_{\mathcal{H},\mathcal{P},l}^{\mathrm{apac}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых значений $\varepsilon, \delta \in (0,1)$ и любого метода минимизации эмпирического риска $\mathcal{A}$ выполняется условие

$$n_{\mathcal{A},\mathcal{P},l}^{\mathrm{apac}}(\varepsilon,\delta) \leqslant n_{\mathcal{H},\mathcal{P},l}^{\mathrm{apac}}(\varepsilon,\delta).$$

Если понятно о каком семействе вероятностных мер и функции потерь идёт речь, то будет использоваться сокращённое обозначение $n_{\mathcal{H}}^{\mathrm{apac}}$.

Содержание

Обобщение модели PAC-обучения

Интерпретация модели PAC-обучения

Модель агностического PAC-обучения является обобщением модели PAC-обучения.

Обобщение модели PAC-обучения

Интерпретация модели PAC-обучения

Модель агностического PAC-обучения является обобщением модели PAC-обучения.

Формально уточним, что под этим понимается.

Обобщение модели PAC-обучения

Интерпретация модели PAC-обучения

Модель агностического PAC-обучения является обобщением модели PAC-обучения.

Формально уточним, что под этим понимается.

Предположим, что задан некоторый класс бинарных классификаторов $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ и семейство вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$.

Обобщение модели PAC-обучения

Интерпретация модели PAC-обучения

Модель агностического PAC-обучения является обобщением модели PAC-обучения.

Формально уточним, что под этим понимается.

Предположим, что задан некоторый класс бинарных классификаторов $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ и семейство вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$.

Каждая пара $h \in \mathcal{H}$ и $\mathsf{P}_X \in \mathcal{P}_{\mathsf{X}}$ задают распределение случайного элемента $x \longmapsto (x, h(x))$ $(x \in \mathsf{X})$.

Обобщение модели PAC-обучения

Интерпретация модели PAC-обучения

Модель агностического PAC-обучения является обобщением модели PAC-обучения.

Формально уточним, что под этим понимается.

Предположим, что задан некоторый класс бинарных классификаторов $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ и семейство вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$.

Каждая пара $h \in \mathcal{H}$ и $\mathsf{P}_X \in \mathcal{P}_{\mathsf{X}}$ задают распределение случайного элемента $x \longmapsto (x, h(x))$ $(x \in \mathsf{X})$.

Семейство всех таких распределений обозначим через $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$.

Обобщение модели PAC-обучения

Интерпретация модели PAC-обучения

Модель агностического PAC-обучения является обобщением модели PAC-обучения.

Формально уточним, что под этим понимается.

Предположим, что задан некоторый класс бинарных классификаторов $\mathcal{H} \subseteq \{0, 1\}^{\mathsf{X}}$ и семейство вероятностных мер $\mathcal{P}_{\mathsf{X}} \subseteq \mathcal{M}^1_+(\mathsf{X})$.

Каждая пара $h \in \mathcal{H}$ и $\mathsf{P}_X \in \mathcal{P}_{\mathsf{X}}$ задают распределение случайного элемента $x \longmapsto (x, h(x))$ $(x \in \mathsf{X})$.

Семейство всех таких распределений обозначим через $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$.

Очевидно, что в этом случае PAC-обучаемость класса $\mathcal{H}$ относительно семейства $\mathcal{P}_{\mathsf{X}}$ будет эквивалента его агностической PAC-обучаемости относительно семейства $\mathcal{P}$ и фунции потерь l_{01}.

Содержание

Равномерная сходимость эмпирического риска

Ранее было показано, что конечный класс концептов является PAC-обучаемым.

Равномерная сходимость эмпирического риска

Ранее было показано, что конечный класс концептов является PAC-обучаемым.

Аналогичный результат может быть получен относительно агностической PAC-обучаемости конечного класса гипотез.

Равномерная сходимость эмпирического риска

Ранее было показано, что конечный класс концептов является PAC-обучаемым.

Аналогичный результат может быть получен относительно агностической PAC-обучаемости конечного класса гипотез.

На самом деле агностическая PAC-обучаемость является следствием более сильного свойства равномерной сходимости эмпирического риска, которым обладает конечный класс гипотез, и к изучению которого мы перейдём.

Содержание

Равномерная сходимость эмпирического риска

Основные определения

В качестве основного объекта изучения будет выступать следующая функция

$$\Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) = \Delta_{\mathcal{F}}(\mathsf{P}; \mathbf{z}) := \sup_{f \in \mathcal{F}} \left| R(\mathsf{P}; f) - r(f, \mathbf{z}) \right| \qquad (\mathbf{z} \in \mathsf{Z}^*),$$

Равномерная сходимость эмпирического риска

Основные определения

В качестве основного объекта изучения будет выступать следующая функция

$$\Delta_{\mathcal{H}}(\mathrm{P}, l; \mathbf{z}) = \Delta_{\mathcal{F}}(\mathrm{P}; \mathbf{z}) := \sup_{f \in \mathcal{F}} \left| R(\mathrm{P}; f) - r(f, \mathbf{z}) \right| \qquad (\mathbf{z} \in \mathrm{Z}^*),$$

где $\mathcal{H} \subseteq \mathrm{Y}^{\mathrm{X}}$, $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, l – функция потерь и $\mathcal{F} \subseteq [0, 1]^{\mathrm{Z}}$, $\mathcal{H} \simeq_l \mathcal{F}$.

Равномерная сходимость эмпирического риска

Основные определения

Определение 4.7.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l, если любого $\varepsilon \in (0, 1)$ выполняется условие

Определение 4.7.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l, если любого $\varepsilon \in (0,1)$ выполняется условие

$$\lim_{n\to\infty} q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) = 0,$$

где

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) := \sup_{\mathsf{P}\in\mathcal{P}} \mathsf{P}^n\Big\{\mathbf{z} \in \mathsf{Z}^n \ : \ \sup_{h\in\mathcal{H}} \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) > \varepsilon\Big\}.$$

Равномерная сходимость эмпирического риска

Основные определения

Определение 4.7.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, если любого $\varepsilon \in (0, 1)$ выполняется условие

$$\lim_{n \to \infty} q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) = 0,$$

где

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) := \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n \Big\{ \mathbf{z} \in \mathsf{Z}^n \ : \ \sup_{h \in \mathcal{H}} \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) > \varepsilon \Big\}.$$

В дальнейшем, если из контекста понятно, о каком семействе вероятностных мер и функции потерь идёт речь, то будем просто говорить о равномерной сходимости эмпирического риска рассматриваемого класса гипотез.

Равномерная сходимость эмпирического риска

Основные определения

Определение 4.8.

Сложностью выборки равномерной сходимости эмпирического риска класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l называется наименьшая функция вида

$$n^{\mathrm{uc}}_{\mathcal{H},\mathcal{P},l} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Определение 4.8.

Сложностью выборки равномерной сходимости эмпирического риска класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l называется наименьшая функция вида

$$n_{\mathcal{H},\mathcal{P},l}^{\mathrm{uc}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n_{\mathcal{H},\mathcal{P},l}^{\mathrm{uc}}(\varepsilon, \delta).$$

Равномерная сходимость эмпирического риска

Основные определения

Определение 4.8.

Сложностью выборки равномерной сходимости эмпирического риска класса гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l называется наименьшая функция вида

$$n_{\mathcal{H},\mathcal{P},l}^{\mathrm{uc}} : (0,1)^2 \longrightarrow \overline{\mathbb{N}},$$

обладающая следующим свойством.

Для любых $\varepsilon, \delta \in (0,1)$ выполняется условие

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) \leqslant \delta \quad \text{при всех } n \geqslant n_{\mathcal{H},\mathcal{P},l}^{\mathrm{uc}}(\varepsilon, \delta).$$

В дальнейшем, если из контекста понятно, о каком семействе вероятностных мер и функции потерь идёт речь, то будут использоваться сокращённые обозначения $\Delta_{\mathcal{H}}$, $\Delta_{\mathcal{F}}$ и $n_{\mathcal{H}}^{\mathrm{uc}}$.

Равномерная сходимость эмпирического риска

Основные определения

Из приведённых определений следует, что класс гипотез $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска тогда и только тогда, когда функция $n_{\mathcal{H}}^{\mathrm{uc}}$ принимает только конечные значения.

Из приведённых определений следует, что класс гипотез $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска тогда и только тогда, когда функция $n_{\mathcal{H}}^{\mathrm{uc}}$ принимает только конечные значения.

Замечание

Распространим введённые понятия на класс функций $\mathcal{F} \subseteq [0, 1]^{\mathsf{Z}}$, $\mathcal{H} \simeq_{l} \mathcal{F}$. Будем говорить, что класс функций $\mathcal{F}$ обладает свойством равномерной сходимости эмпирического риска, если этим свойством обладает класс гипотез $\mathcal{H}$. При этом, по определению положим $n_{\mathcal{F}}^{\mathrm{uc}} := n_{\mathcal{H}}^{\mathrm{uc}}$.

Содержание

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Существует тесная связь между методом минимизации эмпирического риска и функциями $\Delta_{\mathcal{H}}$ и $\Delta_{\mathcal{F}}$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Существует тесная связь между методом минимизации эмпирического риска и функциями $\Delta_{\mathcal{H}}$ и $\Delta_{\mathcal{F}}$.

Если не делать дополнительных предположений, то анализ свойств метода минимизации эмпирического риска сводится к изучению поведения этих функций.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Лемма 4.1 (Вапник и Червоненкис, Девройе).

Пусть $\mathsf{P} \in \mathcal{M}^1_+(\mathsf{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$ и $\mathbf{z} :\longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathsf{Z}^*)$ – метод минимизации эмпирического риска для класса $\mathcal{F}$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Лемма 4.1 (Вапник и Червоненкис, Девройе).

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $\mathbf{z} :\longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathrm{Z}^*)$ – метод минимизации эмпирического риска для класса $\mathcal{F}$.

Тогда

$$R(\mathrm{P}; f_{\mathbf{z}}) \leqslant R(\mathrm{P}; \mathcal{F}) + 2\Delta_{\mathcal{F}}(\mathbf{z}) \qquad (\mathbf{z} \in \mathrm{Z}^*). \tag{14}$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Лемма 4.1 (Вапник и Червоненкис, Девройе).

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $\mathbf{z} :\longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathrm{Z}^*)$ – метод минимизации эмпирического риска для класса $\mathcal{F}$.

Тогда

$$R(\mathrm{P}; f_{\mathbf{z}}) \leqslant R(\mathrm{P}; \mathcal{F}) + 2\Delta_{\mathcal{F}}(\mathbf{z}) \qquad (\mathbf{z} \in \mathrm{Z}^*). \tag{14}$$

◀ Зафиксируем $f \in \mathcal{F}$ и $\mathbf{z} \in \mathrm{Z}^*$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Лемма 4.1 (Вапник и Червоненкис, Девройе).

Пусть $\mathrm{P} \in \mathcal{M}^1_+(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $\mathbf{z} :\longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathrm{Z}^*)$ – метод минимизации эмпирического риска для класса $\mathcal{F}$.

Тогда

$$R(\mathrm{P}; f_{\mathbf{z}}) \leqslant R(\mathrm{P}; \mathcal{F}) + 2\Delta_{\mathcal{F}}(\mathbf{z}) \qquad (\mathbf{z} \in \mathrm{Z}^*). \tag{14}$$

◀ Зафиксируем $f \in \mathcal{F}$ и $\mathbf{z} \in \mathrm{Z}^*$.

Из определения функции $\Delta_{\mathcal{F}}$ следует

$$|R(\mathrm{P}; f) - r(f, \mathbf{z})| \leqslant \Delta_{\mathcal{F}}(\mathbf{z}).$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Лемма 4.1 (Вапник и Червоненкис, Девройе).

Пусть $\mathrm{P} \in \mathcal{M}_+^1(\mathrm{Z})$, $\mathcal{F} \subseteq [0,1]^{\mathrm{Z}}$ и $\mathbf{z} :\longmapsto f_{\mathbf{z}}$ $(\mathbf{z} \in \mathrm{Z}^*)$ – метод минимизации эмпирического риска для класса $\mathcal{F}$.

Тогда

$$R(\mathrm{P}; f_{\mathbf{z}}) \leqslant R(\mathrm{P}; \mathcal{F}) + 2\Delta_{\mathcal{F}}(\mathbf{z}) \qquad (\mathbf{z} \in \mathrm{Z}^*). \tag{14}$$

◀ Зафиксируем $f \in \mathcal{F}$ и $\mathbf{z} \in \mathrm{Z}^*$.

Из определения функции $\Delta_{\mathcal{F}}$ следует

$$|R(\mathrm{P}; f) - r(f, \mathbf{z})| \leqslant \Delta_{\mathcal{F}}(\mathbf{z}).$$

Используя это неравенство, получим

$$R(\mathrm{P}; f) \geqslant r(f, \mathbf{z}) - \Delta_{\mathcal{F}}(\mathbf{z}) \geqslant r(f_{\mathbf{z}}, \mathbf{z}) - \Delta_{\mathcal{F}}(\mathbf{z}) \geqslant R(\mathrm{P}; f_{\mathbf{z}}) - 2\Delta_{\mathcal{F}}(\mathbf{z}).$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Переходя к инфимуму по $f \in \mathcal{F}$ в левой части этого неравенства, получим (14).

$\blacksquare$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Теорема 4.6.

Предположим, что класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_+^1(\mathsf{Z})$ и функции потерь l.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Теорема 4.6.

Предположим, что класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l.

Тогда, если для $\mathcal{H}$ существует метод минимизации эмпирического риска, то этот класс гипотез является агностически PAC-обучаемым относительно $\mathcal{P}$ и l. При этом выполняется неравенство

$$n^{\mathrm{apac}}_{\mathcal{H}}(\varepsilon, \delta) \leqslant n^{\mathrm{uc}}_{\mathcal{H}}(\varepsilon/2, \delta) \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{15}$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Теорема 4.6.

Предположим, что класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l.

Тогда, если для $\mathcal{H}$ существует метод минимизации эмпирического риска, то этот класс гипотез является агностически PAC-обучаемым относительно $\mathcal{P}$ и l. При этом выполняется неравенство

$$n_{\mathcal{H}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon/2, \delta) \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{15}$$

◀ Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$, $\varepsilon \in (0, 1)$ и $n \in \mathbb{N}$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Теорема 4.6.

Предположим, что класс гипотез $\mathcal{H} \subseteq \mathrm{Y}^{\mathrm{X}}$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathrm{Z})$ и функции потерь l.

Тогда, если для $\mathcal{H}$ существует метод минимизации эмпирического риска, то этот класс гипотез является агностически PAC-обучаемым относительно $\mathcal{P}$ и l. При этом выполняется неравенство

$$n^{\mathrm{apac}}_{\mathcal{H}}(\varepsilon, \delta) \leqslant n^{\mathrm{uc}}_{\mathcal{H}}(\varepsilon/2, \delta) \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{15}$$

◀ Зафиксируем произвольные $\mathrm{P} \in \mathcal{P}$, $\varepsilon \in (0, 1)$ и $n \in \mathbb{N}$.

Обозначим через $\mathcal{A}$ метод минимизации эмпирического риска для класса гипотез $\mathcal{H}$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Определим два события

$$Q_\varepsilon := \left\{ \mathbf{z} \in \mathsf{Z}^n \ : \ \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) > \varepsilon \right\},$$
$$R_\varepsilon := \left\{ \mathbf{z} \in \mathsf{Z}^n \ : \ R(\mathsf{P}, l; h_{\mathbf{z}}) > R(\mathsf{P}, l; \mathcal{H}) + \varepsilon \right\}.$$

Согласно лемме 4.1 имеет место включение $Q^c_{\varepsilon/2} \subseteq R^c_\varepsilon$,

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Определим два события

$$Q_\varepsilon := \big\{\mathbf{z} \in \mathrm{Z}^n \ : \ \Delta_{\mathcal{H}}(\mathrm{P}, l; \mathbf{z}) > \varepsilon\big\},$$
$$R_\varepsilon := \big\{\mathbf{z} \in \mathrm{Z}^n \ : \ R(\mathrm{P}, l; h_{\mathbf{z}}) > R(\mathrm{P}, l; \mathcal{H}) + \varepsilon\big\}.$$

Согласно лемме 4.1 имеет место включение $Q^c_{\varepsilon/2} \subseteq R^c_\varepsilon$,

из которого в свою очередь следует включение $R_\varepsilon \subseteq Q_{\varepsilon/2}$,

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Определим два события

$$Q_\varepsilon := \left\{\mathbf{z} \in \mathsf{Z}^n \ : \ \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) > \varepsilon\right\},$$
$$R_\varepsilon := \left\{\mathbf{z} \in \mathsf{Z}^n \ : \ R(\mathsf{P}, l; h_{\mathbf{z}}) > R(\mathsf{P}, l; \mathcal{H}) + \varepsilon\right\}.$$

Согласно лемме 4.1 имеет место включение $Q^c_{\varepsilon/2} \subseteq R^c_\varepsilon$,

из которого в свою очередь следует включение $R_\varepsilon \subseteq Q_{\varepsilon/2}$,

а значит и неравенство $\mathsf{P}^n(R_\varepsilon) \leqslant \mathsf{P}^n(Q_{\varepsilon/2})$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Определим два события

$$Q_\varepsilon := \{\mathbf{z} \in \mathsf{Z}^n \, : \, \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) > \varepsilon\},$$
$$R_\varepsilon := \{\mathbf{z} \in \mathsf{Z}^n \, : \, R(\mathsf{P}, l; h_{\mathbf{z}}) > R(\mathsf{P}, l; \mathcal{H}) + \varepsilon\}.$$

Согласно лемме 4.1 имеет место включение $Q^c_{\varepsilon/2} \subseteq R^c_\varepsilon$,

из которого в свою очередь следует включение $R_\varepsilon \subseteq Q_{\varepsilon/2}$,

а значит и неравенство $\mathsf{P}^n(R_\varepsilon) \leqslant \mathsf{P}^n(Q_{\varepsilon/2})$.

Переходя в левой и правой части этого неравенства к супремуму по $\mathsf{P} \in \mathcal{P}$, получим

$$\bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) = \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n(R_\varepsilon) \leqslant \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n(Q_{\varepsilon/2}) = q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon/2). \tag{16}$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

По условию теоремы рассматриваемый класс гипотез $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

По условию теоремы рассматриваемый класс гипотез $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска.

Следовательно, переходя к пределу в левой и правой частях неравенства (16), можно записать

$$0 \leqslant \lim_{n\to\infty} \bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) \leqslant \lim_{n\to\infty} q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon/2) = 0,$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

По условию теоремы рассматриваемый класс гипотез $\mathcal{H}$ обладает свойством равномерной сходимости эмпирического риска.

Следовательно, переходя к пределу в левой и правой частях неравенства (16), можно записать

$$0 \leqslant \lim_{n\to\infty} \bar{r}_{\mathcal{A}}(\mathcal{P}, l; n, \varepsilon) \leqslant \lim_{n\to\infty} q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon/2) = 0,$$

что по определению означает агностическую PAC-обучаемость класса гипотез $\mathcal{H}$.

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Из (16) следует также неравенство

$$n_{\mathcal{A}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon/2, \delta) \qquad \big(\varepsilon, \delta \in (0, 1)\big),$$

Равномерная сходимость эмпирического риска

Достаточное условие агностической PAC-обучаемости

Из (16) следует также неравенство

$$n_{\mathcal{A}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon/2, \delta) \qquad \big(\varepsilon, \delta \in (0, 1)\big),$$

а значит в силу произвольности выбора метода минимизации эмпирического риска выполняется также и требуемое неравенство (15).

$\blacksquare$

Содержание

Равномерная сходимость эмпирического риска

Конечный случай

Установим свойство равномерной сходимости эмпирического риска и агностическую PAC-обучаемость конечного класса гипотез.

Равномерная сходимость эмпирического риска

Конечный случай

Установим свойство равномерной сходимости эмпирического риска и агностическую PAC-обучаемость конечного класса гипотез.

Теорема 4.7.

Конечный класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$, $|\mathcal{H}| < \infty$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}^1_+(\mathsf{Z})$ и функции потерь l.

Равномерная сходимость эмпирического риска
Конечный случай

Установим свойство равномерной сходимости эмпирического риска и агностическую PAC-обучаемость конечного класса гипотез.

Теорема 4.7.

Конечный класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$, $|\mathcal{H}| < \infty$ обладает свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l.

При этом выполняется неравенство

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon, \delta) \leqslant \left\lceil \frac{1}{2\varepsilon^2} \ln\left(\frac{2|\mathcal{H}|}{\delta}\right) \right\rceil \quad (\varepsilon, \delta \in (0, 1)). \tag{17}$$

Равномерная сходимость эмпирического риска

Конечный случай

- Будем рассматривать класс функций $\mathcal{F} \subseteq [0,1]^Z$, $\mathcal{H} \simeq_l \mathcal{F}$.

Равномерная сходимость эмпирического риска

Конечный случай

◀ Будем рассматривать класс функций $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\mathcal{H} \simeq_l \mathcal{F}$.

Сразу заметим, что $|\mathcal{F}| < |\mathcal{H}|$.

Равномерная сходимость эмпирического риска

Конечный случай

◀ Будем рассматривать класс функций $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\mathcal{H} \simeq_l \mathcal{F}$.

Сразу заметим, что $|\mathcal{F}| < |\mathcal{H}|$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$, $\varepsilon \in (0,1)$ и $n \in \mathbb{N}$. Для каждого $f \in \mathcal{F}$ определим событие

$$Q_{\varepsilon,f} := \left\{ \mathbf{z} \in \mathsf{Z}^n \ : \ \left| R(\mathsf{P}; f) - r(f, \mathbf{z}) \right| > \varepsilon \right\}.$$

Равномерная сходимость эмпирического риска

Конечный случай

◀ Будем рассматривать класс функций $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\mathcal{H} \simeq_l \mathcal{F}$.

Сразу заметим, что $|\mathcal{F}| < |\mathcal{H}|$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$, $\varepsilon \in (0,1)$ и $n \in \mathbb{N}$. Для каждого $f \in \mathcal{F}$ определим событие

$$Q_{\varepsilon,f} := \left\{\mathbf{z} \in \mathsf{Z}^n \ : \ \left|R(\mathsf{P}; f) - r(f, \mathbf{z})\right| > \varepsilon\right\}.$$

Имеет место представление

$$Q_\varepsilon = \bigcup_{f \in \mathcal{F}} Q_{\varepsilon,f}, \tag{18}$$

◀ Будем рассматривать класс функций $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\mathcal{H} \simeq_l \mathcal{F}$.

Сразу заметим, что $|\mathcal{F}| < |\mathcal{H}|$.

Зафиксируем произвольные $\mathsf{P} \in \mathcal{P}$, $\varepsilon \in (0,1)$ и $n \in \mathbb{N}$. Для каждого $f \in \mathcal{F}$ определим событие

$$Q_{\varepsilon,f} := \left\{\mathbf{z} \in \mathsf{Z}^n \; : \; \left|R(\mathsf{P}; f) - r(f, \mathbf{z})\right| > \varepsilon\right\}.$$

Имеет место представление

$$Q_\varepsilon = \bigcup_{f \in \mathcal{F}} Q_{\varepsilon,f}, \tag{18}$$

где событие Q_ε определено при доказательстве предыдущей теоремы 4.6.

Равномерная сходимость эмпирического риска

Конечный случай

Функции координатных проекций

$$\mathsf{Z}^n \ni \mathbf{z} = (z_1, \ldots, z_n) \longmapsto z_i \quad (i = 1, \ldots, n)$$

являются независимыми случайными величинами с распределением P.

Равномерная сходимость эмпирического риска

Конечный случай

Функции координатных проекций

$$Z^n \ni \mathbf{z} = (z_1, \ldots, z_n) \longmapsto z_i \quad (i = 1, \ldots, n)$$

являются независимыми случайными величинами с распределением P.

Но тогда для каждой функции $f \in \mathcal{F}$ независимыми и одинаково распределёнными будут также случайные величины

$$Z^n \ni \mathbf{z} = (z_1, \ldots, z_n) \longmapsto f(z_i) \quad (i = 1, \ldots, n).$$

Равномерная сходимость эмпирического риска

Конечный случай

Функции координатных проекций

$$Z^n \ni \mathbf{z} = (z_1, \ldots, z_n) \longmapsto z_i \quad (i = 1, \ldots, n)$$

являются независимыми случайными величинами с распределением P.

Но тогда для каждой функции $f \in \mathcal{F}$ независимыми и одинаково распределёнными будут также случайные величины

$$Z^n \ni \mathbf{z} = (z_1, \ldots, z_n) \longmapsto f(z_i) \quad (i = 1, \ldots, n).$$

Применяя к этим случайным величинам неравенство Хёффдинга, а точнее следствие 2.2, получим

$$\mathrm{P}^n(Q_{\varepsilon,f}) \leqslant 2e^{-2n\varepsilon^2}. \tag{19}$$

Равномерная сходимость эмпирического риска

Конечный случай

Из (18) и (19) следует, что

$$\mathrm{P}^n(Q_\varepsilon) \leqslant \sum_{f \in \mathcal{F}} \mathrm{P}^n(Q_{\varepsilon,f}) \leqslant \sum_{f \in \mathcal{F}} 2e^{-2n\varepsilon^2} = 2|\mathcal{F}|e^{-2n\varepsilon^2} \leqslant 2|\mathcal{H}|e^{-2n\varepsilon^2}.$$

Равномерная сходимость эмпирического риска

Конечный случай

Из (18) и (19) следует, что

$$\mathrm{P}^n(Q_\varepsilon) \leqslant \sum_{f \in \mathcal{F}} \mathrm{P}^n(Q_{\varepsilon,f}) \leqslant \sum_{f \in \mathcal{F}} 2e^{-2n\varepsilon^2} = 2|\mathcal{F}|e^{-2n\varepsilon^2} \leqslant 2|\mathcal{H}|e^{-2n\varepsilon^2}.$$

Для каждого $\delta \in (0,1)$ решением неравенства

$$2|\mathcal{H}|e^{-2n\varepsilon^2} \leqslant \delta$$

Равномерная сходимость эмпирического риска

Конечный случай

Из (18) и (19) следует, что

$$\mathrm{P}^n(Q_\varepsilon) \leqslant \sum_{f \in \mathcal{F}} \mathrm{P}^n(Q_{\varepsilon,f}) \leqslant \sum_{f \in \mathcal{F}} 2e^{-2n\varepsilon^2} = 2|\mathcal{F}|e^{-2n\varepsilon^2} \leqslant 2|\mathcal{H}|e^{-2n\varepsilon^2}.$$

Для каждого $\delta \in (0, 1)$ решением неравенства

$$2|\mathcal{H}|e^{-2n\varepsilon^2} \leqslant \delta$$

будет

$$n \geqslant \left\lceil \frac{1}{2\varepsilon^2} \ln\left(\frac{2|\mathcal{H}|}{\delta}\right) \right\rceil. \tag{20}$$

Равномерная сходимость эмпирического риска

Конечный случай

Но тогда (20) будет также удовлетворять неравенству

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) = \sup_{\mathrm{P} \in \mathcal{P}} \mathrm{P}^n(Q_\varepsilon) \leqslant \delta. \tag{21}$$

Равномерная сходимость эмпирического риска

Конечный случай

Но тогда (20) будет также удовлетворять неравенству

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) = \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n(Q_\varepsilon) \leqslant \delta. \tag{21}$$

В виду произвольности выбора значений $\varepsilon, \delta \in (0, 1)$ из (20) и (21) будет следовать равномерная сходимость эмпирического риска для класса гипотез $\mathcal{H}$ и справедливость оценки (17).

■

Равномерная сходимость эмпирического риска

Конечный случай

Но тогда (20) будет также удовлетворять неравенству

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) = \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n(Q_\varepsilon) \leqslant \delta. \tag{21}$$

В виду произвольности выбора значений $\varepsilon, \delta \in (0, 1)$ из (20) и (21) будет следовать равномерная сходимость эмпирического риска для класса гипотез $\mathcal{H}$ и справедливость оценки (17).

■

В случае конечного класса гипотез функция эмпирического риска на фиксированной обучающей выборке принимает конечное число значений.

Равномерная сходимость эмпирического риска

Конечный случай

Но тогда (20) будет также удовлетворять неравенству

$$q_{\mathcal{H}}(\mathcal{P}, l; n, \varepsilon) = \sup_{\mathsf{P} \in \mathcal{P}} \mathsf{P}^n(Q_\varepsilon) \leqslant \delta. \tag{21}$$

В виду произвольности выбора значений $\varepsilon, \delta \in (0, 1)$ из (20) и (21) будет следовать равномерная сходимость эмпирического риска для класса гипотез $\mathcal{H}$ и справедливость оценки (17).

■

В случае конечного класса гипотез функция эмпирического риска на фиксированной обучающей выборке принимает конечное число значений.

Поэтому для такого класса гипотез всегда существует метод минимизации эмпирического риска.

Равномерная сходимость эмпирического риска

Конечный случай

Сформулируем общее следствие из теорем 4.6 и 4.7.

Равномерная сходимость эмпирического риска

Конечный случай

Сформулируем общее следствие из теорем 4.6 и 4.7.

Следствие 4.1.

Конечный класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$, $|\mathcal{H}| < \infty$ является агностически PAC-обучаемым с помощью метода минимизации эмпирического риска и

$$n_{\mathcal{H}}^{\mathrm{apac}}(\varepsilon, \delta) \leqslant \left\lceil \frac{2}{\varepsilon^2} \ln\left(\frac{2|\mathcal{H}|}{\delta}\right) \right\rceil \qquad \bigl(\varepsilon, \delta \in (0,1)\bigr). \tag{22}$$

Содержание

Равномерная сходимость эмпирического риска

Общий случай

Как и в случае модели PAC-обучения, вполне ожидаемо оказалось, что конечный класс гипотез будет обладать свойством агностической PAC-обучаемости.

Равномерная сходимость эмпирического риска

Общий случай

Как и в случае модели PAC-обучения, вполне ожидаемо оказалось, что конечный класс гипотез будет обладать свойством агностической PAC-обучаемости.

Существование бесконечных не PAC-обучаемых классов бинарных классификаторов говорит о том, что эти классы также не будут обладать свойством агностической PAC-обучаемости.

Равномерная сходимость эмпирического риска

Общий случай

Как и в случае модели PAC-обучения, вполне ожидаемо оказалось, что конечный класс гипотез будет обладать свойством агностической PAC-обучаемости.

Существование бесконечных не PAC-обучаемых классов бинарных классификаторов говорит о том, что эти классы также не будут обладать свойством агностической PAC-обучаемости.

Приводимый ниже критерий совместно с теоремой 4.6 может использоваться для проверки агностической PAC-обучаемости в бесконечном случае.

Равномерная сходимость эмпирического риска

Общий случай

Как и в случае модели PAC-обучения, вполне ожидаемо оказалось, что конечный класс гипотез будет обладать свойством агностической PAC-обучаемости.

Существование бесконечных не PAC-обучаемых классов бинарных классификаторов говорит о том, что эти классы также не будут обладать свойством агностической PAC-обучаемости.

Приводимый ниже критерий совместно с теоремой 4.6 может использоваться для проверки агностической PAC-обучаемости в бесконечном случае.

Прежде, чем перейти к формулировке и доказательству этого критерия, сделаем одно замечание.

Равномерная сходимость эмпирического риска

Общий случай

В дальнейшем будет дано необходимое и достаточное условие агностической PAC-обучаемости для частного случая задачи бинарной классификации.

Равномерная сходимость эмпирического риска

Общий случай

В дальнейшем будет дано необходимое и достаточное условие агностической PAC-обучаемости для частного случая задачи бинарной классификации.

Существенную роль при этом будет играть следующее свойство функции потерь l_{01}.

Равномерная сходимость эмпирического риска

Общий случай

В дальнейшем будет дано необходимое и достаточное условие агностической PAC-обучаемости для частного случая задачи бинарной классификации.

Существенную роль при этом будет играть следующее свойство функции потерь l_{01}.

В случае её использования всегда существует метод минимизации эмпирического риска.

Равномерная сходимость эмпирического риска

Общий случай

В дальнейшем будет дано необходимое и достаточное условие агностической PAC-обучаемости для частного случая задачи бинарной классификации.

Существенную роль при этом будет играть следующее свойство функции потерь l_{01}.

В случае её использования всегда существует метод минимизации эмпирического риска.

Для произвольной функции потерь заранее расчитывать на выполнение такого предположения не приходится.

Равномерная сходимость эмпирического риска

Общий случай

В дальнейшем будет дано необходимое и достаточное условие агностической PAC-обучаемости для частного случая задачи бинарной классификации.

Существенную роль при этом будет играть следующее свойство функции потерь l_{01}.

В случае её использования всегда существует метод минимизации эмпирического риска.

Для произвольной функции потерь заранее расчитывать на выполнение такого предположения не приходится.

Поэтому для агностической PAC-обучаемости бесконечного класса гипотез не достаточно только требовать выполнение свойством равномерной сходимости эмпирического риска.

Равномерная сходимость эмпирического риска

Общий случай

Для такого класса должно быть отдельно проверено существование метода минимизации эмпирического риска.

Равномерная сходимость эмпирического риска

Общий случай

Теорема 4.8.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ будет обладать свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, если существует константа $C > 0$ такая, что для любых $\mathsf{P} \in \mathcal{P}$ и $n \in \mathbb{N}$ выполняется неравенство

$$\mathop{\mathbf{E}}_{\mathbf{z} \sim \mathsf{P}^n} \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) \leqslant \frac{C}{\sqrt{n}}. \tag{23}$$

Равномерная сходимость эмпирического риска

Общий случай

Теорема 4.8.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ будет обладать свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, если существует константа $C > 0$ такая, что для любых $\mathsf{P} \in \mathcal{P}$ и $n \in \mathbb{N}$ выполняется неравенство

$$\mathop{\mathbf{E}}_{\mathbf{z} \sim \mathsf{P}^n} \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) \leqslant \frac{C}{\sqrt{n}}. \tag{23}$$

Кроме того, в этом случае имеет место оценка

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon, \delta) \leqslant \left\lceil \max\left\{ \frac{4C^2}{\varepsilon^2}, \frac{2}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right) \right\} \right\rceil \qquad \big(\varepsilon, \delta \in (0, 1)\big). \tag{24}$$

Равномерная сходимость эмпирического риска

Общий случай

Теорема 4.8.

Класс гипотез $\mathcal{H} \subseteq \mathsf{Y}^{\mathsf{X}}$ будет обладать свойством равномерной сходимости эмпирического риска относительно семейства вероятностных мер $\mathcal{P} \subseteq \mathcal{M}_{+}^{1}(\mathsf{Z})$ и функции потерь l, если существует константа $C > 0$ такая, что для любых $\mathsf{P} \in \mathcal{P}$ и $n \in \mathbb{N}$ выполняется неравенство

$$\mathop{\mathbf{E}}_{\mathbf{z} \sim \mathsf{P}^n} \Delta_{\mathcal{H}}(\mathsf{P}, l; \mathbf{z}) \leqslant \frac{C}{\sqrt{n}}. \tag{23}$$

Кроме того, в этом случае имеет место оценка

$$n_{\mathcal{H}}^{\mathrm{uc}}(\varepsilon, \delta) \leqslant \left\lceil \max\left\{ \frac{4C^2}{\varepsilon^2}, \frac{2}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right) \right\} \right\rceil \qquad (\varepsilon, \delta \in (0,1)). \tag{24}$$

◀ Будем рассматривать класс функций $\mathcal{F} \subseteq [0,1]^{\mathsf{Z}}$, $\mathcal{H} \simeq_l \mathcal{F}$.

Равномерная сходимость эмпирического риска

Общий случай

Основная идея доказательства состоит в использовании неравенства МакДиармида.

Равномерная сходимость эмпирического риска

Общий случай

Основная идея доказательства состоит в использовании неравенства МакДиармида.

Для произвольных $\mathbf{z} = (z_1, \ldots, z_i, \ldots z_n)$, $\mathbf{z}' = (z_1, \ldots, z_i', \ldots z_n) \in \mathsf{Z}^n$, отличающихся только в i-ой координате ($1 \leqslant i \leqslant n$), запишем

$$|R(\mathsf{P}; f) - r(f, \mathbf{z})| - |R(\mathsf{P}; f) - r(f, \mathbf{z}')| \leqslant |r(f, \mathbf{z}) - r(f, \mathbf{z}')| = \frac{1}{n}|f(z_i) - f(z_i')| \leqslant \frac{1}{n},$$

Равномерная сходимость эмпирического риска

Общий случай

Основная идея доказательства состоит в использовании неравенства МакДиармида.

Для произвольных $\mathbf{z} = (z_1, \ldots, z_i, \ldots z_n)$, $\mathbf{z}' = (z_1, \ldots, z_i', \ldots z_n) \in \mathrm{Z}^n$, отличающихся только в i-ой координате ($1 \leqslant i \leqslant n$), запишем

$$|R(\mathrm{P}; f) - r(f, \mathbf{z})| - |R(\mathrm{P}; f) - r(f, \mathbf{z}')| \leqslant |r(f, \mathbf{z}) - r(f, \mathbf{z}')| = \frac{1}{n}|f(z_i) - f(z_i')| \leqslant \frac{1}{n},$$

а значит

$$|R(\mathrm{P}; f) - r(f, \mathbf{z})| \leqslant |R(\mathrm{P}; f) - r(f, \mathbf{z}')| + \frac{1}{n}, \tag{25}$$

где $\mathrm{P} \in \mathcal{P}$, $f \in \mathcal{F}$.

Равномерная сходимость эмпирического риска

Общий случай

Переходя к супремуму по $f \in \mathcal{F}$ в левой и правой частях неравенства (25), получим

$$\Delta_{\mathcal{F}}(\mathrm{P}; \mathbf{z}) \leqslant \Delta_{\mathcal{F}}(\mathrm{P}; \mathbf{z}') + \frac{1}{n}.$$

Равномерная сходимость эмпирического риска
Общий случай

Переходя к супремуму по $f \in \mathcal{F}$ в левой и правой частях неравенства (25), получим
$$\Delta_{\mathcal{F}}(\mathsf{P}; \mathbf{z}) \leqslant \Delta_{\mathcal{F}}(\mathsf{P}; \mathbf{z}') + \frac{1}{n}.$$

Следовательно, функция $g : \mathbf{z} \longmapsto \Delta_{\mathcal{F}}(\mathsf{P}; \mathbf{z})$ $(\mathbf{z} \in \mathsf{Z}^n)$ обладает свойством ограниченности покоординатных приращений.

Равномерная сходимость эмпирического риска

Общий случай

Переходя к супремуму по $f \in \mathcal{F}$ в левой и правой частях неравенства (25), получим

$$\Delta_{\mathcal{F}}(\mathrm{P}; \mathbf{z}) \leqslant \Delta_{\mathcal{F}}(\mathrm{P}; \mathbf{z}') + \frac{1}{n}.$$

Следовательно, функция $g : \mathbf{z} \longmapsto \Delta_{\mathcal{F}}(\mathrm{P}; \mathbf{z})$ $(\mathbf{z} \in \mathrm{Z}^n)$ обладает свойством ограниченности покоординатных приращений.

Запишем для неё неравенство МакДиармида

$$\mathrm{P}^n\{g - \mathbf{E}g \geqslant s\} \leqslant e^{-2ns^2} \quad (s > 0). \tag{26}$$

Равномерная сходимость эмпирического риска

Общий случай

Зафиксируем произвольные $\varepsilon, \delta \in (0, 1)$ и запишем

$$
\begin{aligned}
\mathsf{P}^n\{g \geqslant \varepsilon\} &= \mathsf{P}^n\{g - \mathbf{E}g \geqslant \varepsilon - \mathbf{E}g\} \leqslant \left| \begin{array}{c} \text{Неравенство} \\ \text{(23)} \end{array} \right| \\
&\leqslant \mathsf{P}^n\left\{g - \mathbf{E}g \geqslant \varepsilon - \frac{C}{\sqrt{n}}\right\} \leqslant \left| \text{При } n \geqslant \frac{4C^2}{\varepsilon^2} \right| \\
&\leqslant \mathsf{P}^n\left\{g - \mathbf{E}g \geqslant \frac{\varepsilon}{2}\right\} \leqslant \left| \begin{array}{c} \text{Неравенство} \\ \text{(26)} \end{array} \right| \\
&\leqslant e^{-n\varepsilon^2/2}.
\end{aligned}
\tag{27}
$$

Равномерная сходимость эмпирического риска

Общий случай

Из (27) следует, что при

$$n \geqslant \max\left\{\frac{4C^2}{\varepsilon^2}, \frac{2}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right)\right\}$$

выполняется неравенство

$$\mathsf{P}^n\{g \geqslant \varepsilon\} \leqslant \delta.$$

Равномерная сходимость эмпирического риска

Общий случай

Из (27) следует, что при

$$n \geqslant \max\left\{\frac{4C^2}{\varepsilon^2}, \frac{2}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right)\right\}$$

выполняется неравенство

$$\mathsf{P}^n\{g \geqslant \varepsilon\} \leqslant \delta.$$

В виду произвольности выбора $\mathsf{P} \in \mathcal{P}$ будет справедлива оценка (24), из которой в свою очередь вытекает равномерная сходимость эмпирического риска класса гипотез $\mathcal{H}$.

■