Теория статистического обучения

Н. К. Животовский
nikita.zhivotovskiy@phystech.edu

31 марта 2016 г.

Задачи по курсу.

Часть первая.

Задача 1

- Привести пример случайной величины $X \geq 0$, для которой в неравенстве Маркова достигается точное равенство.
- Привести пример случайной величины Y, для которой в неравенстве Чебышева достигается точное равенство.

Задача 2 Пусть $\psi(x)$ — плотность распределения случайной величины $X \sim \mathcal{N}(0, 1)$.

- Доказать, что $\psi'(x) + x\psi(x) = 0$.
- Доказать, что для всех $x > 0$

$$\psi(x)\left(\frac{1}{x} - \frac{1}{x^3}\right) \leq P(X \geq x) \leq \psi(x)\left(\frac{1}{x} - \frac{1}{x^3} + \frac{1}{x^5}\right).$$

- Сравнить верхнюю оценку в последнем неравенстве с оценкой, получаемой с помощью метода Чернова.

Задача 3 Пусть F, F_n — соответственно функция распределения и эмпирическая функция распределения некоторой случайной величины X.

- Оценить с наперед заданной вероятностью для фиксированного n величину

$$\sup_{x \in \mathbb{R}} |F(x) - F_n(x)|.$$

- Построить неасимптотический аналог критерия Колмогорова–Смирнова.
- Можно ли в общем случае улучшить полученные по n порядки?

Задача 4 Пусть $\mathcal{F}^d$ — класс, состоящий из всех функций, представимых в виде конъюнкций наборов переменных $x_1, \ldots, x_d$ и их отрицаний. Пусть V_d — VC-размерность класса $\mathcal{F}^d$.

- Доказать, что $\mathcal{F}^d$ является агностически PAC–обучаемым.
- Доказать, что $V_d \leq d \ln(3)$
- Можно ли улучшить оценку до $V_d \leq d$?
- Какова выборочная сложность для минимизатора эмпирического риска по классу $\mathcal{F}^d$ в агностическом случае и в случае отсутствия шума?

Задача 5 Рассмотрим задачу классификации с классами $\{1, -1\}$ и бинарной функцией потерь. Обозначим $\eta(x) = \mathbb{E}[Y|X = x]$ и $f^*(x) = \operatorname{sign}(\eta(x))$.

- Доказать, что f^* минимизирует предсказательный риск, то есть
$$\inf_{f \in \mathcal{Y}^{\mathcal{X}}} L(f) = L(f^*).$$
- Доказать, что для всех $f \in \mathcal{Y}^{\mathcal{X}}$ имеет место соотношение:
$$L(f) - L(f^*) = \mathbb{E}\left(|\eta(X)|\mathbf{I}[f(X) \neq f^*(X)]\right).$$
- Доказать, что в данной задаче, если определять радемахеровские сложности без модулей, то есть как
$$\mathcal{R}_n(\ell \circ \mathcal{F}) = \frac{1}{n}\, \mathbb{E}_\sigma \sup_{g \in \ell \circ \mathcal{F}} \left(\sum_{i=1}^n \sigma_i g(X_i, Y_i) \right),$$
то условная радемахеровская сложность $\mathcal{R}_n(\ell \circ \mathcal{F})$ не зависит от наблюдаемых классов Y_i объектов обучающей выборки.
- Пусть выполнено условие малого шума, то есть $|\eta(X)| \geq \frac{1}{c}$ для некоторого $c \geq 1$. Оценить в этих условиях $L(f^*)$.

Задача 6 Доказать, что в случае квадратичной функции потерь избыточный риск любой функции $f \in \mathcal{F}$ задается с помощью $f, f^*_{\mathcal{F}}, f^*$, где $f^*_{\mathcal{F}}, f^*$ — соответственно минимизатор риска в классе $\mathcal{F}$ и глобальный минимизатор риска.

Задача 7 Пусть даны классы $\mathcal{F}_1, \ldots, \mathcal{F}_r$, такие что среди них наибольшая VC-размерность равняется d. Доказать, что VC-размерность объединения этих классов имеет порядок $O(d \log(d) + \log(r))$.

Задача 8

- Доказать, что для любого множества $A \subseteq \{0, 1\}^n$
$$\mathcal{R}_n(A) \leq \inf_{\alpha \in [0,1]} \left(\alpha + \sqrt{\frac{2 \log(2N(\varepsilon, A))}{n}} \right).$$
- Применить полученное неравенство для оценки избыточного риска в задаче классификации в случае конечной VC-размерности.

Часть вторая.