Теория статистического обучения

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

21 мая 2015 г.

Задачи по курсу

Задача 1 Пусть F, F_n — соответственно функция распределения и эмпирическая функция распределения некоторой случайной величины X.

- Оценить с наперед заданной вероятностью для фиксированного n величину

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

- Можно ли в общем случае улучшить полученные по n порядки?

Задача 2 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 — размерность Вапника–Червоненкиса класса $\mathcal{F}^d$.

- Доказать, что $\mathcal{F}^d$ является агностически PAC–обучаемым.
- Доказать, что $V_d \le d\ln(3)$

  Указание: сосчитайте мощность класса.
- Можно ли улучшить оценку до $V_d \le d$?
- Какова выборочная сложность для минимизатора эмпирического риска по классу $\mathcal{F}^d$ в агностическом и реализуемом случаях?

  Указание: для анализа реализуемого случая понадобятся идеи из 5-ой лекции.

Задача 3 Рассмотрим задачу классификации с классами $\{1, -1\}$ и бинарной функцией потерь. Обозначим $\eta(x) = \mathbb{E}[Y|X = x]$ и $f^*(x) = \mathrm{sign}(\eta(x))$.

- 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:

$$L(f) - L(f^*) = \mathbb{E}\left(|\eta(X)|\mathbf{I}[f(X) \neq f^*(X)]\right).$$

Указание: доказательство есть в рекомендованной книге A Probabilistic Theory of Pattern Recognition.

- 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.

Указание: убедитесь, что $\mathcal{R}_n(\ell \circ \mathcal{F}) = \frac{1}{2}\mathcal{R}_n(\mathcal{F})$.

Задача 4 [Задача по лекциям К. В. Воронцова]

- Выписать формулу полного скользящего контроля $C(\mu, \mathbb{X})$ для унимодальной цепи. Возможно ли её упростить?

- Предлагается вывести формулу вероятности переобучения Q_ε для модельного семейства алгоритмов $A = \{a_1, \ldots, a_D\}$, $n(a_d, \mathbb{X}) = m + 1 - (d \bmod 2)$ при $D = 3, 4, 5$. Метод μ — пессимистичная минимизация эмпирического риска.

Задача 5 Пусть класс $\mathcal{F}$ состоит из линейных решающих правил, причем евклидовы нормы всех векторов в нем ограничены некоторой константой B. Доказать агностическую PAC–обучаемость данного класса при условии, что функция потерь является выпуклой и липшицевой по первому аргументу.

Задача 6 Пусть $\mathbf{x}_1, \ldots, \mathbf{x}_m$ — векторы в $\mathbb{R}^d$ и $\mathbf{x}$ — случайный вектор, распределенный равномерно на $\mathbf{x}_1, \ldots, \mathbf{x}_m$. Предположим также, что $\mathbb{E}\mathbf{x} = \mathbf{0}$.

- Рассмотрим задачу поиска единичного вектора ω такого, что дисперсия $\omega^T\mathbf{x}$ максимальна. Доказать, что решением этой задачи будет первая главная компонента системы $\mathbf{x}_1, \ldots, \mathbf{x}_m$.

- Пусть первой главной компоненте соответствует вектор ω_1. Доказать, что среди всех единичных векторов ω, таких что $\omega^T\mathbf{x}$ и $\omega_1^T\mathbf{x}$, некоррелированы максимальную дисперсию $\omega^T\mathbf{x}$ дает вторая главная компонента.