Задание номер 2

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

12 марта 2017 г.

Задание принимается до 2.00 утра 26 марта по адресу slt.fupm.2017@gmail.com. В начале текста задания *обязательно* указывается:

- С кем вы делали это задание.
- Какие источники (кроме материалов лекций) вы использовали.

Задание оформляется в формате pdf (текст набирается в latex/Word) и в таком виде, чтобы ваши коллеги могли разобрать текст решения. Задания, оформленные не в соответствии с указанными правилами, не принимаются. Желательно оставлять зазоры между задачами для пометок.

Упражнение 1. Пусть класс $\mathcal{S}$ состоит всех тех функций на $[0, 1]$, которые принимают значение 1 не более чем на конечном числе точек, а на оставшихся точках равны нулю.

- Докажите, что для его Радемахеровской сложности выполнено $R(\mathcal{S}) \geq \frac{1}{2}$.
- Что можно сказать о выполнении равномерных законов больших чисел для $\mathcal{S}$?

Указание. Можно ссылаться на результат задачи 1.

■

Упражнение 2 [Линейные классы при $d \geq n$] Пусть $\mathcal{F} = \{x \to \mathrm{sign}((x, \theta))|\ \theta \in \mathbb{R}^d\}$ — класс линейных разделяющих правил.

- Предположим, что $d \geq n$ и точки $X_1, \ldots, X_n$ общего положения в $\mathbb{R}^d$. Покажите, что в этом случае для условной Радемахеровской сложности выполнено $R_n(\mathcal{F}) = 1$.
- Что можно сказать про обучаемость класса с помощью минимизации эмпирического риска?

Упражнение 3 [Гауссовские сложности] Определим Гауссовскую сложность класса $\mathcal{F}$ как $G(\mathcal{F}) = \frac{1}{n}\mathbb{E}_g\mathbb{E}_X \sup\limits_{f\in\mathcal{F}} |\sum\limits_{i=1}^{n} g_i f(X_i)|$, где $g_i \sim \mathcal{N}(0,1)$ и независимы. Докажите, $R(\mathcal{F}) \le CG(\mathcal{F})$ для некоторой универсальной константы $C > 0$.

Указание. Используйте, что для некоторого $c > 0$ имеет место $\mathbb{E}(|g|/c) = 1$, где $g \sim \mathcal{N}(0,1)$. Примените неравенство Йенсена, предварительно подставив эту единицу в нужную часть определения.

■

Упражнение 4 [Размерность Вапника-Червоненкиса]

- Привести пример семейства классификаторов, для которого для любого конечного набора из n различных точек из $\mathcal{X}$ верхняя оценка на функцию роста переходит в равенство. Такие классы называются *максимальными*.
- Покажите, что VC размерность семейства всех классификаторов в $\mathbb{R}^d$, образованных выпуклыми замкнутыми множествами, бесконечна.
- Найти VC размерность семейства классификаторов на плоскости, образованных всеми многоугольниками с не более чем 4-мя вершинами.
- Найти VC размерность семейства классификаторов на плоскости, образованных всеми возможными окружностями.
- Пусть семейство классификаторов $\mathcal{F}$ имеет VC размерность d. Рассмотрим семейство классификаторов $\mathcal{F}_k$, которые получаются голосованием большинства не более чем k классификаторов из $\mathcal{F}$. Доказать, что VC размерность $\mathcal{F}$ ограничена сверху $O(kd\log(kd))$.

Упражнение 5 [Contraction] Ограничьте $R(\ell \circ \mathcal{F})$ с помощью $R(\mathcal{F})$ в следующих задачах:

- Бинарная классификация с двумя классами $\mathcal{Y} = \{-1, 1\}$ и индикаторной функцией потерь.
- Регрессии с $|Y| \le a$, $\sup\limits_{f\in\mathcal{F}} \|f\|_\infty \le b$ и квадратичный функцией потерь.

Задача 1. [Десимметризация] Докажите что,

$$\mathbb{E}\sup_{f\in\mathcal{F}} |Pf - P_n f| \ge \frac{1}{2}R(\mathcal{F}) - \frac{C}{2\sqrt{n}},$$

где $C = \sup\limits_{f\in\mathcal{F}} \|f\|_\infty$.

Задача 2 Докажите эквивалентность трех утверждений для класса $\mathcal{F}$ бинарных классификаторов:

1. Класс $\mathcal{F}$ обучаем.
2. Класс $\mathcal{F}$ имеет конченую VC размерность.
3. Класс $\mathcal{F}$ является равномерным классом Гливенко-Кантелли.

Обратите внимание, что условие 3 выписано для класса $\mathcal{F}$, а не для класса $\ell \circ \mathcal{F}$.

Задача 3

- Докажите, что класс положительных линейных решающих правил (первая компонента направляющего вектора неотрицательна) является максимальным в том смысле, что будучи ограниченным на любой конечный набор точек в общем положении, верхняя оценка на функцию роста переходит в точное равенство.
- Является ли максимальным класс всех линейных классификаторов?