

Домашнее задание №1 по курсу «Машинное обучение»: ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ

Колесов Алексей

4 сентября 2017 г.

Вам необходимо прислать pdf-файл с доказательствами, графиками и прочим. Отдельно нужно прислать код программы и способ её запустить (в идеале, ссылку на IPython-notebook).

1 Задания

1. Пусть $f : (0, +\infty) \rightarrow \mathbb{R}$ — обратимая функция, а X — случайная величина. Докажите, что если для любого $t > 0$ $\mathbb{P}[X > t] \leq f(t)$, то для любого $\delta > 0$ с вероятностью как минимум $1 - \delta$ выполняется, что $X \leq f^{-1}(\delta)$.
2. На лекции мы рассмотрели вот такой классификатор:

$$h_S(x) = \begin{cases} y_i & \text{если } \exists i \in [m] \text{ такой что } x_i = x \\ 0 & \text{иначе} \end{cases}$$

Пусть $X = \mathbb{R}$. Рассмотрим класс **пороговых полиномиальных классификаторов**:

$$h_p(x) = \begin{cases} 1 & \text{если } a_0 + a_1x + \dots + a_nx^n \geq 0 \\ 0 & \text{иначе} \end{cases}$$

Докажите, что в классе h_p найдётся классификатор, совпадающий с h_S . Какой вывод можно сделать о ERM-парадигме в классе пороговых полиномиальных классификаторов?

3. Это задание о **прямоугольниках со сторонами, параллельными осям координат**. Такой классификатор выглядит следующим образом:

$$h_{a_1, b_1, a_2, b_2}((x_1, x_2)) = \begin{cases} 1 & \text{если } a_1 \leq x_1 \leq b_1 \text{ и } a_2 \leq x_2 \leq b_2 \\ 0 & \text{иначе} \end{cases}$$

Будем считать, что выполнено предположение о реализуемости.

- Пусть алгоритм A выбирает наименьший прямоугольник, содержащий все точки положительного класса. Докажите, что A является реализацией ERM-алгоритма
- Реализуйте программу, которая принимает число m и генерирует выборку из m точек: пусть X — квадрат с противоположными углами в $(0, 0)$ и $(1, 1)$, D — равномерное распределение над X . $f(x) = [(x, y) \in Q]$, где Q — квадрат с противоположными углами в $(0, 0)$ и $(\sqrt{1/2}, \sqrt{1/2})$. Затем программа должна реализовать алгоритм A из предыдущего пункта и вычислить true risk полученной гипотезы
- Постройте график true risk в зависимости от m

- Запустите программу для всех m от 1 до n (n выберите в зависимости от показателей алгоритма). Какой в среднем понадобился размер выборки, чтоб true risk 10%? 1%? 0.1%?
- Как ответ на предыдущий пункт должен зависеть от площади X (при неизменной площади Q)? От относительной площади Q и X ? От размерности пространства X ? Должен ли зависеть результат от D ? Если нет, то почему? Если да, то как объяснить тот факт, что зависимость есть, хотя в определении PAC-learnable класса выборочная сложность не зависит от D ?