Домашнее задание №1 по курсу «Машинное обучение»: основы машинного обучения

Скавыш Максим

Задание 1

Пусть $f: (0, +\infty\ ) \to R$ обратимая функция, X – сл. величина.
Доказать: если для любого t > 0 $P[X > t] \leq f(t)$, то для любого $\delta > 0$ с вероятностью как минимум $1 - \ \delta$ выполняется $X \leq f^{-1}(\delta)$

Решение:

для любого t > 0 выполняется $P[X > t] \leq f(t)$.
Пусть $f(t)$ обратимая функция, тогда имеет место $f^{-1}\big(f(t)\big) = t$.
Обозначим для любого фиксированного $t \in (0, +\infty)$: $f(t) = \delta \ \Rightarrow f^{-1}(\delta) = t,\ \delta \in R$
Перепишем $P[X > t] \leq f(t)$ как $P[X > f^{-1}(\delta)] \leq \delta$ и так как вероятность не может быть меньше 0, то накладываем на δ ограничение $\delta > 0$.
Выражение $P[X > f^{-1}(\delta)] \leq \delta$ эквивалентно $P[X \leq f^{-1}(\delta)] > 1 - \delta$ из чего следует что если для любого t > 0, $P[X > t] \leq f(t)$, то для любого $\delta > 0$ с вероятностью как минимум $1 - \ \delta$ выполняется $X \leq f^{-1}(\delta)$.

Задание 2

Решение:

Рассмотрим класс полиномиальных классификаторов:
$$h_p(x) = \begin{cases} 1, & a_0 + a_1x + \cdots + a_nx^n \geq 0 \\ & 0, \text{иначе} \end{cases}$$
и построим в нем классификатор, совпадающий с $h_s(x)$ из условия.
Пусть $S = \big((x_1,\ y_1), \dots,\ (x_m,\ y_m)\big)$ – тренировачная выборка
Пусть $L = \{\ i \in (1, \dots, m) : y_i = 1\}$
Построим полином который будет принимать 0-вые значения только в $X_L = \{\ x_i,\ i \in L\ \}$, таким полиномом является полином с корнями $\{\ x_i,\ i \in L\ \}$ и его можно представить в виде: $\prod_{i \in L}(x - \ x_i)$.
Рассмотрим полином $P_L(x) = \ (-1)\prod_{i \in L}(x - \ x_i)^2$ он всегда ≤ 0 и принимает значение 0 только в точках X_L, тогда классификатор вида $\begin{cases} 1, & P_L(x) = \geq 0 \\ & 0, \text{иначе} \end{cases}$ будет принадлежать классу $h_p(x)$ и совпадать с классификатором $h_s(x)$.

Какой вывод можно сделать о ERM-парадигме в классе пороговых полиномиальных классификаторов?

В классе пороговых полиномиальных классификаторов обязательно найдется классификатором с полиномом степени не меньше $\sum_{i \in (1, \dots, m)} y_i$ имеющий ошибку на тренировочной выборке равной 0, однако это будет приводить к переобучению и тому, что на всем D такой классификатор будет иметь большую погрешность.

Задание 3

$$h_{a_1,b_1,a_2b_2}(x_1,x_2) = \begin{cases} 1, & a_1 \le x_1 \le b_1 \text{ и } a_2 \le x_2 \le b_2 \\ & 1, \text{иначе} \end{cases}$$

1. Пусть алгоритм А выбирает наименьший прямоугольник, содержащий все точки положительного класса. Докажите, что А является реализацией ERM-алгоритма

Выполнено предположение о реализуемости ⇒ найдётся такая гипотеза $h^*\in$ H, что $L_{D,f}(h^*) = 0$ ⇒ найдётся такая гипотеза $h_s \in$ H, что $L_s(h_s) = 0$
Пусть алгоритм А выбирает наименьший прямоугольник, содержащий все точки положительного класса. Тогда на тренировочной выборке гипотеза h_s полученная алгоритмом будет иметь минимально возможную эмпирическую ошибку $L_s(h_s) = 0$
ERM-алгоритм выбирает самую лучшую гипотезу по отношению к тренировочной выборке: $ERM_H(S) \in argmin_{h\in H}\ L_s(h) \Rightarrow$ А является реализацией ERM-алгоритма.

2. Реализуйте программу

3. 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%?
   https://github.com/MaksimSkavysh/Math/blob/master/ml-homework_1.py-Copy2.ipynb

   10% ~ m в среднем 67
   1% ~ m в среднем 550
   0.1% ~ m в среднем 4000