Домашнее задание 1

Авласов Владислав

Задание 1

Возьмём неравенство из условия и перепишем, построив обратное событие:

$$\mathbb{P}[X \leqslant t] > 1 - f(t) \tag{1}$$

Пусть $f(t) = \delta$. Тогда область значений функции $f^{-1} : \mathbb{R} \to (0, +\infty), t = f^{-1}(\delta) > 0$. Подставим в неравенство, получаем:

$$\mathbb{P}\left[X \leqslant f^{-1}(\delta)\right] > 1 - \delta \tag{2}$$

ЧТД.

Задание 2

Для указанного классификатора h_S рассмотрим условие: $\exists i \in [m] : x_i = x$. Его можно представить в виде:

$$\prod_{i=1}^{m} (x - x_i) = 0 \tag{3}$$

Это полином степени m, принимающий неотрицательное значение для x_i. Для тренировочной выборки он всегда обращается в ноль, а для остальных элементов результат будет зависеть от знака произведения. Давайте сделаем так, чтобы для всех элементов не из тренировочной выборки полином был меньше нуля. Для этого возведём каждую скобку в квадрат и домножим весь полином на -1:

$$(-1) \cdot \prod (x - x_i)^2 \tag{4}$$

Это выражение меньше нуля всегда, кроме случаев попадания в выборку S, что соответсвует первому классификатору h_S
ЧТД.
Вывод: ERM-правило для класса пороговых полиномиальных классификаторов приводит к переобучению.

Задание 3

Т.к. у нас выполняется предположение о реализуемости, то существует гипотеза с $L_S(h) = 0$. Для того, чтобы эмпирический риск был минимальным, мы обязательно должны включить все точки положительного класса в прямоугольник. Предположение о реализуемости гарантирует нам, что мы сможем это сделать с нулевой ошибкой, т.е. не включив ни одной точки отрицательного класса в прямоугольник. Минимальный прямоугольник, содрежащий все точки положительного класса, входит в множество подходящих нам прямоугольников. Т.е. выбирая его, мы реализуем ERM-алгоритм. ЧТД.