

Задание 1. Пусть $f: (0, +\infty) \rightarrow \mathbb{R}$ — обратимая функция, а X — случайная величина.

Докажите, что если для

любого $t > 0$ $P[X > t] \leq f(t)$, то для любого $\delta > 0$ с вероятностью как минимум $1 - \delta$ выполняется, что $X \leq f^{-1}(\delta)$.

Док-во.

$P[X > t] \leq f(t) \Rightarrow P[X \leq t] \geq 1 - f(t)$.

Пусть $t = f^{-1}(\delta)$, тогда:

$P[X \leq f^{-1}(\delta)] \geq 1 - f(f^{-1}(\delta))$

$P[X \leq f^{-1}(\delta)] \geq 1 - \delta$

Следовательно для любого $\delta > 0$ с вероятностью как минимум $1 - \delta$ выполняется, что $X \leq f^{-1}(\delta)$.

Задание 2.

Док-во.

Пусть для $\forall i f(x_i) = 1$. Подберем полином так, чтобы доказать требуемое.

$(-1) * (x - x_1)^2 * \dots * (x - x_m)^2$. Данный полином принимает значение 0 на всех x_i из $S|_x$ и отрицательное значение на любом другом x , следовательно такой классификатор из h_p совпадает с классификатором из h_s .

Вывод: в классе пороговых полиномиальных классификаторов существует ERM-парадигма эмпирический риск, которой равен 0.

Задание 3.

Пусть $h = A(S)$, тогда для $\forall i \Rightarrow h(x_i) = y_i$. Т.е. любая гипотеза, порожденная алгоритмом A , не ошибается на $S|_x$. Следовательно $L_S(h) = 0$, т.е. A является реализацией ERM-алгоритма.

Ссылка на IPython-notebook:

https://github.com/AlexandrShestak/ml_mag/blob/master/ml/lab1/lab1.ipynb

Средний размер выборки, чтобы true risk был 10%: 13

Средний размер выборки, чтобы true risk был 1%: 75

Средний размер выборки, чтобы true risk был 0.1%: 458

Как ответ на предыдущий пункт должен зависеть от площади X (при неизменной площади Q)? Чем больше будет X тем меньше становится вероятность того, что алгоритм ошибется, следовательно размер выборок нужен будет меньше.

От относительной площади Q и X ?

Формула true risk : $\text{true_risk} = (S(Q) - S(H))/S(X)$. S - площадь, H - множество, которое покрывается алгоритмом A . Следовательно чем больше отношение площади Q и X , тем true risk будет больше, и нужен будет больший размер выборки для достижение нужного результата.

От размерности пространства X ? От размерности пространства зависит..

Например если X - это отрезок длины 1, Q - отрезок от нуля до 0 до $\sqrt{1/2}$, то

Средний размер выборки, чтобы true risk был 10%: 8

Средний размер выборки, чтобы true risk был 1%: 37

Средний размер выборки, чтобы true risk был 0.1%: 166

Это можно объяснить тем, что с уменьшением размерности уменьшается шанс попасть в ту область, в которой алгоритм ошибается.

Должен ли зависеть результат от D?

Зависимость от D будет. Например результат при нормальном распределении нужный результат будет достигнут быстрее

https://github.com/AlexandrShestak/ml_mag/blob/master/ml/lab1/lab1_normal_distribution.ipynb

Но тут, наверное, не выполняется условие одинаковой распределенности. При выполнении условий независимости и одинаковой распределенности, результат от D не должен зависеть.