

Семинар №1 по курсу «Машинное обучение»

Колесов Алексей

4 сентября 2017 г.

1 Задания

1. **Лемма о ERM в случае $\frac{\epsilon}{2}$ -репрезентативной выборки** Пусть выборка S является $\frac{\epsilon}{2}$ -репрезентативной. Тогда если h_S — ERM-гипотеза, то выполняется:

$$L_D(h_S) \leq \min_{h \in H} L_D(h) + \epsilon$$

2. Пусть D — распределение над X и $f : X \rightarrow \{-1; +1\}$ — размечающая функция. Допустим, мы хотим найти насколько несбалансированы метки классов, а именно найти такую вероятность:

$$p_+ = \mathbb{P}_{x \sim D}[f(x) = 1]$$

Пусть у вас есть конечная выборка S размера m независимо полученная из D . Предложите оценку \hat{p}_+ для величины p_+ . Покажите, что для любого $\delta > 0$ с вероятностью как минимум $1 - \delta$, $|p_+ - \hat{p}_+| \leq \sqrt{\frac{\log(2/\delta)}{2m}}$