

Машинное обучение, ФКН ВШЭ

Теоретическое домашнее задание №18

Задача 1. Покажите, что расстояние Джаккарда является метрикой.

Задача 2. Пусть решается задача восстановления регрессии при помощи метода k ближайших соседей, при этом фиксирована некоторая обучающая выборка $X = \{(x_i, y_i)\}_{i=1}^{\ell}$, а ответ y на объекте x генерируется следующим образом:

$$y \sim \mathcal{N}(f(x), \sigma^2), \sigma > 0,$$

где $f(x)$ — истинная зависимость. Найдите шумовую компоненту, смещение и дисперсию в bias-variance разложении среднеквадратичного риска $L(\mu) = \mathbb{E}_{x,y} [(y - \mu(X)(x))^2]$ для произвольного k .