

Машинное обучение, ФКН ВШЭ

Домашнее задание №6

Задача 1. Пусть $x \in \mathbb{R}^d$, и значение каждого признака на объекте x независимо генерируется из равномерного распределения: $x_i \sim U[0, 1]$, $i = 1, \dots, d$. Будем считать, что объекты в выборке независимы, а $\mathbb{E}[y|x] = x^T x$. Найдите смещение константного алгоритма: $\mu(X)(x) = C = \text{const}$.

Задача 2. Предположим, что объекты описываются единственным категориальным признаком, принимающим значения $x = 1, \dots, K$ с равной вероятностью, объекты независимы. Для каждой категории x определено истинное целевое значение f_x , а наблюдаемое целевое значение для объекта x определяется как $y = f_x + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Алгоритм обучения $\mu(X)$ запоминает среднее значение для каждой категории следующим образом: $\hat{f}_x = \frac{1}{\ell} \sum_{i=1}^{\ell} [x_i = x] y_i$, а затем для объекта x выдаёт предсказание \hat{f}_x . Найдите смещение такого алгоритма.

Задача 3. Предположим, что мы решаем задачу бинарной классификации и что у нас есть три алгоритма $b_1(x)$, $b_2(x)$ и $b_3(x)$, каждый из которых ошибается с вероятностью p . Мы строим композицию взвешенным голосованием: алгоритмам присвоены значимости w_1 , w_2 и w_3 , и для вынесения вердикта суммируются значимости алгоритмов, проголосовавших за каждый из классов:

$$a_0 = \sum_{i=1}^3 w_i [b_i(x) = 0],$$
$$a_1 = \sum_{i=1}^3 w_i [b_i(x) = 1].$$

Объект x относится к классу, для которого такая сумма оказалась максимальной. Например, если первые два алгоритма голосуют за класс 0, а третий — за класс 1, то выбирается класс 0, если $w_1 + w_2 > w_3$, и класс 1 в противном случае. Какова вероятность ошибки такой композиции этих трех алгоритмов, если:

1. $w_1 = 0.2$, $w_2 = 0.3$, $w_3 = 0.2$;
2. $w_1 = 0.2$, $w_2 = 0.5$, $w_3 = 0.2$?

Задача 4. Убедитесь, что можете ответить на следующие вопросы:

1. Как измеряется качество метода обучения, если известно распределение $p(x, y)$?
2. Что такое $\mu(X)$ — число, функция, что-то еще? Что такое $\mathbb{E}_X[\mu(X)]$?

3. Какие формулы у шума, смещения, разброса? Какой смысл у этих компонент?
4. Приведите пример семейства с маленьким смещением и большим разбросом. Приведите пример семейства с большим смещением и маленьким разбросом.
5. Как сгенерировать подвыборку с помощью бутстрапа?
6. Что такое бэггинг?
7. Как соотносятся смещение разброс композиции, построенной с помощью бэггинга, со смещением и разбросом одного базового алгоритма?
8. Как обучается случайный лес? В чем отличия от обычной процедуры построения решающих деревьев?
9. Почему хорошими базовыми алгоритмами для бэггинга являются именно деревья?
10. Как оценить качество случайного леса с помощью out-of-bag-процедуры?
11. Как с помощью случайного леса можно измерить сходство между объектами?