hw_theor02

February 8, 2021

0.1 Машинное обучение

0.1.1 Факультет математики НИУ ВШЭ, 2020-21 учебный год

Илья Щуров, Соня Дымченко, Руслан Хайдуров, Максим Бекетов, Павел Егоров

Страница курса

0.2 Домашнее задание №2

Фамилия и имя студента: *(впишите свои)*

0.2.1 Задача 1 (15 баллов)

Маша, Катя и Люба изучают выборку $x_1, \ldots, x_n$ из нормального распределения с неизвестным средним μ и дисперсией 1. Они хотят оценить μ по этой выборке. Маша в качестве оценки использует выборочное среднее Ave (то есть просто среднее арифметическое), Катя использует [медиану выборки](https://ru.wikipedia.org/wiki/Медиана_(статистика%29), а Люба функцию midrange:

$$\mathrm{midrange}(x_1, \ldots, x_n) = \frac{1}{2}(\max(x_1, \ldots, x_n) + \min(x_1, \ldots, x_n)).$$

1. Являются ли эти оценки несмещенными? Ответьте с помощью численного эксперимента: зафиксируйте какое-нибудь μ (например, $\mu = 0$) и n (например, $n = 10$), сгенерируйте много (например, 10 000) выборок (это можно сделать с помощью функции `np.random.normal`, в качестве `size` нужно передать пару `(число_выборок, n)` — получится матрица указанного размера, заполненная случайными числами из данного распределения), для каждой найдите значение соответствующей функции (нужно использовать функции `np.mean`, `np.random`, `np.max`, `np.min` — все они принимают параметр `axis` — изучите, как он работает) и усредните их. Получается ли число, близкое к μ? Становится ли оно ближе с увеличением числа выборок (при фиксированном n)?

2. Оцените дисперсию каждой оценки для различных n. Зафиксируйте число выборок (допустим, 1000) и в цикле по `n` от 2 до 100 выполните следующее. Сгенерируйте 1000 выборок длиной n, для каждой выборки найдите значение соответствущей оценки (аналогично предыдущему пункту) и посчитайте выборочную дисперсию для полученных оценок (с помощью `.var()`). Постройте график, показывающий зависимость дисперсии каждой из оценок от n. Какая оценка имеет наименьшую дисперсию? Какая наибольшую? Какую из этих оценок вы бы стали использовать, если бы хотели минимизировать квадратичную ошибку предсказания?

3. Выполните пункт 2 для равномерного распределения на отрезке $[-1, 1]$. Какая теперь оценка имеет наименьшую дисперсию? Какая наибольшую? Как вы можете объяснить разницу с предыдущим пунктом? Какую из этих оценок вы бы стали использовать в этом случае, если бы хотели минимизировать квадратичную ошибку предсказания?

[1]:
```
# ваше решение тут
```

0.2.2 Задача 2 (10 баллов)

Рассмотрим случайную величину Y, имеющую плотность $p(y)$, которую мы будем считать известной функцией. Пусть p непрерывна; рассмотрим носитель меры, заданной плотностью p, то есть множество $\{y \mid p(y) > 0\}$. Предположим, что носителем является связное подмножество прямой (быть может, вся прямая).

Мы хотим подобрать такую величину $\hat{y} \in \mathbb{R}$, чтобы матожидание функции потерь $\mathbb{E}_{y \sim Y} L(y, \hat{y})$ было минимальным. Пусть $L(y, \hat{y}) = |y - \hat{y}|$. Выразить оптимальное $\hat{y}$ через функцию p. Докажите, что оптимальное $\hat{y}$ определяется однозначно.

(впишите решение сюда)

0.2.3 Задача 3 (12 баллов)

Пусть дана выборка $x_1, \ldots, x_n$, все $x_i \in \mathbb{R}$ распределены как случайная величина X и независимы в совокупности. $\mathbb{E}[X] < \infty$, $\mathbb{D}[X] < \infty$. Для фиксированного вектора $w \in \mathbb{R}^n$ рассмотрим функцию

$$\varphi_w(x) = \langle w, x \rangle,$$

где $\langle w, x \rangle$ — стандартное скалярное произведение (скалярное произведение, записанное в ортонормированном базисе).

1. При каком условии на w эта функция будет несмещённой оценкой для $\mathbb{E}[X]$?
2. Среди всех w, при которых $\varphi_w(x)$ является несмещённой оценкой для $\mathbb{E}[X]$, найти такое, при котором дисперсия $\varphi_w(x)$ будет наименьшей. (Подсказка: вам понадобятся множители Лагранжа.)

(впишите решение сюда)

0.2.4 Задача 4 (10 баллов)

Рассмотрим задачу регрессии с одномерным пространством признаков. Пусть истинный закон генерирования данных описывается следующим образом: все x_i фиксированы и заданы так: $x_i = i - 3$, $i = 1, \ldots, 5$, а y_i являются случайными величинами:

$$y_i = |x_i| + \varepsilon_i,$$

где все ε_i независимы, $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{D}[\varepsilon_i] = 4$. Пусть $f_k(x)$ — предсказание метода k ближайших соседей в точке x. Найти ожидаемую квадратичную ошибку предсказания в точке $x = 0$ для $k = 3$. Представить её в виде суммы шума, смещения и разброса.

(впишите решение сюда)

0.2.5 Задача 5 (15 баллов)

Пусть числа $y_1, \ldots, y_n$ получены как выборка из нормального распределения $\mathcal{N}(\mu, \sigma^2)$ с неизвестными параметрами μ и σ^2. Мы хотим найти оценку наибольшего правдоподобия для μ и σ^2, то есть такие значения этих параметров, при которых функция правдоподобия

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2)$$

будет максимальной. На лекциях была найдена функция правдоподобия и показано, что оптимальное μ можно найти независимо от σ^2 и оно равно выборочному среднему.

1. Завершите нахождение оценки наибольшего правдоподобия: найдите теперь оптимальное σ^2.
2. Является ли полученная оценка несмещённой? Докажите.

(впишите решение сюда)

0.2.6 Задача 6 (10 баллов)

Гарри Поттер хочет найти философский камень, расположенный в точке минимума функции $f(x_1, x_2) = x_1^2 + x_2^2$. В момент времени 0 он стартует из точки $x^{(0)} = (2, 2)$. На i-й минуте Гарри мгновенно перемещается (аппарирует) из точки $x^{(i)}$ в точку

$$x^{(i+1)} = x^{(i)} - \eta \nabla f(x^{(i)}),$$

где $\nabla f(x^{(i)})$ — градиент f в точке $x^{(i)}$, $\eta \geq 0$ — фиксированное число. Опишите судьбу Гарри в зависимости от значения η. При каких значениях η Гарри подойдёт к философскому камню сколь угодно близко? Сколько времени ему понадобится, чтобы подойти к философскому камню на расстояние не больше ε?

(впишите решение сюда)