

Задание 1 по теме “Линейная регрессия.”

**1.1 Многомерная линейная регрессия.** Пусть имеется набор данных  $(x_\alpha^{(i)}, y^{(i)})$ ,  $i = 1 \dots m$ ,  $\alpha = 1 \dots k$ . Здесь индекс  $(i)$  нумерует события в наборе, а индекс  $\alpha$  – переменные. Линейная регрессия – аппроксимация данных функцией первого порядка  $y = b_0 + \sum_{\alpha=1}^m b_\alpha x_\alpha$ .

- i) Используя квадратичную функцию стоимости и считая ошибки измерения  $y^{(i)}$  одинаковыми и равными  $\sigma$ , получить выражения для коэффициентов линейной регрессии  $b_0, b_\alpha$ .
- ii) Повторить вывод, считая ошибки измерения  $y^{(i)}$  разными и равными  $\sigma^{(i)}$ .

**1.2 Линейная аппроксимация функций.** Для перечисленных ниже функций

- i)  $\sin(x)$ ,  $x \in [0, \pi/2]$
- ii)  $\log(1+x)$ ,  $x \in [0, 1]$
- iii)  $|x|$ ,  $x \in [-1, 1]$
- iv) ступенька  $\theta(x-1)$ ,  $x \in [0, 2]$
- v)  $\exp(-1/x^2)$ ,  $x \in [0, 3]$

выполнить следующие упражнения:

- a) Сгенерировать набор  $N$  равномерно распределенных случайных значений  $x$  и вычислить для них значение исследуемой функции. Этот набор назовем тренировочным. Аналогично построить второй (тестовый) набор такого же объема. Рекомендуется использовать для этого `numpy.random` и другие функции из библиотеки NumPy или `torch.random` из библиотеки PyTorch.
- б) Используя тренировочный набор, построить линейную регрессию.
- в) Получить аналитически ответ для линейной регрессии в пределе  $N \rightarrow \infty$ .
- г) Используя тестовый набор, построить долю объясненной дисперсии как функцию  $N$ .

**1.3 Упрощенная классификация гамма-всплесков.** В этом задании начнем рассматривать задачу классификации гамма-всплесков, стартуя с предельно упрощенной модели. Будем считать, что для гамма-всплеска известна лишь длительность, причем короткие и длинные гамма-всплески равномерно распределены по длительности.

- a) Смоделировать тренировочный набор  $3N$  коротких гамма-всплесков с длительностью  $\tau \in (0, 2)$  секунд и  $7N$  длинных гамма-всплесков с длительностью  $\tau \in (2, 1000)$  секунд. Аналогично построить тестовый набор.
- б) Считаем, что искомая функция равна 0 для коротких гамма-всплесков и 1 для длинных гамма-всплесков.

- в) Построить классификатор гамма-всплесков, основанный на методе линейной регрессии.
- г) Построить кривую обучения.

**1.4 Применение линейной регрессии к реальным данным.** Вам предлагается построить линейную регрессию на реальных данных и получить коэффициенты полуэмпирической формулы Вайцзеккера. Для этого будут использованы данные Центра данных фотоядерных экспериментов НИИЯФ МГУ <http://cdfe.sinp.msu.ru/services/gsp.ru.html>.

- а) Скачайте файл с данными по ссылке: [http://ppc.inr.ac.ru/grisha\\_data/A\\_Z\\_deltaE.dat](http://ppc.inr.ac.ru/grisha_data/A_Z_deltaE.dat).
- б) Загрузите таблицу в Питон, используя, например, `numpy.loadtxt(filename, options)` (NumPy) или `pandas.read_table(filename, options)` (Pandas).
- в) Добавьте в таблицу дополнительные величины, используемые капельной моделью ядра, используя столбцы  $A$  и  $Z$ , а также доступные в Питоне математические операции. Слагаемое, зависящее от чётности ядер, можно получить, используя, например, `*(column%2==0)` (такой множитель будет работать как символ Кронекера).
- г) Получите коэффициенты формулы Вайцзеккера, используя многомерную линейную регрессию, и сравните результат со справочными данными.