

Задачи для семинара по линейной регрессии

1. Предсказания в линейной регрессии (майнор ИАД). Какие предсказания сделает линейная модель

$$a(x) = w_1x_1 + w_2x_2 + w_3x_3 + b$$

для объекта $x = (7, -3, 2)$, вектора весов $w = (0.5, 1.5, -2)$ и сдвига $b = 10$?

2. Обучение константного алгоритма. Пусть мы обучаем константный алгоритм регрессии: $a(x) = c = \text{const}$. Функция потерь – квадратичная. Найдите оптимальное значение c . Необходимо найти минимум функции:

$$\sum_{i=1}^{\ell} (y_i - c)^2 \rightarrow \min_{c \in \mathbb{R}}$$

После решения задачи подумайте, каким будет ответ, если использовать абсолютную функцию потерь: $L(y, z) = |y - z|$.

3. Обучение одномерной линейной регрессии без сдвига. У вас есть выборка $\{x_1, \dots, x_\ell\}$, $x_i \in \mathbb{R}$ из ℓ объектов с одним признаком и вектор $\{y_1, \dots, y_\ell\}$, $y_i \in \mathbb{R}$ вещественных целевых переменных (задача регрессии). Вы обучаете на этих данных линейную регрессию без свободного коэффициента (настраивается только вес единственного признака) с использованием квадратичной функции потерь. Выведите формулу для оптимального значения этого веса.

4. Общая формула обучения линейной регрессии без сдвига В общем случае формула для поиска оптимальных весов в линейной регрессии с квадратичной функцией потерь выглядит так: $w = (x^T X)^{-1} X^T Y$. Формула выводится по аналогии с задачей 3, но в многомерном случае. Вопросы:

- Проверьте размерности в формуле (проверить, что во всех матричных умножениях и обращениях размерности корректны).
- Проверьте, что ваше решение в задаче 3 является частным случаем данной формулы.

Если вы решили все задачи, реализуйте линейную регрессию как класс языка Python с функциями `fit` и `predict`. Протестировать реализацию можно, используя данные Бостона. В качестве заготовки используйте материалы семинара по классам.