

1 Зважена лінійна регресія

У цьому завданні ви працюватимете з реальними даними з [зарплатного опитування DOU.ua за травень 2016р.](#) Ви реалізуєте зважену лінійну регресію, яка передбачає зарплати Java-інженерів, та навчите свою модель за допомогою градієнтного спуску.

У записнику **salary-prediction-wlr.ipynb** уже реалізована підготовка даних, візуалізація та оцінка результатів моделі. Вам залишається реалізувати саму логіку зваженої лінійної регресії, заповнивши пропущені місця в коді.

Після того, як закінчите, запустіть останню комірку — вона містить автоматичні тести, що перевіряють правильність ваших обчислень. Ви повинні побачити повідомлення «ОК», якщо все працює вірно.

- a. Реалізуйте функцію гіпотези зваженої лінійної регресії.
- b. Реалізуйте функцію зважування навчальних прикладів.
- c. Реалізуйте функцію втрат зваженої лінійної регресії.
- d. Обчисліть градієнт функції втрат зваженої лінійної регресії.
- e. Реалізуйте правило оновлення ваг при градієнтному спуску.

2 Класифікація спаму методом Баєсса

Ви застосуєте наївний Баєсівський класифікатор зі згладжуванням Лапласа для навчання спам-фільтру (на основі даних [SpamAssassin Public Corpus](#)).

Потрібно реалізувати мультиноміальну модель подій для класифікації текстів — [multinomial event model](#) та порівняти з [multivariate Bernoulli event model](#). Прочитати про деталі реалізації моделей можна у [статті](#).

Заповніть пропущений код у записнику **spam-bayes.ipynb**. У вас повинен вийти кращий результат, ніж при multivariate Bernoulli (> 95% точності).

- a. Закодуйте лист у вигляді вектора ознак.
- b. Підрахуйте кількість слів у ham- та spam- листах.
- c. Обчисліть апіорні ймовірності для класів ham та spam.
- d. Обчисліть ймовірності появи слів в рамках кожного класу.
- e. Реалізуйте функцію класифікації нового листа.
- f. Порівняйте точність вашого класифікатору із бібліотечними (BernoulliNB та MultinomialNB).