Лабораторная работа 09-11. Линейные модели: Регрессия

# Основная часть

## Задание 1: Простая линейная регрессия.

Используйте набор данных "Boston Housing" из sklearn.datasets. Постройте модель линейной регрессии, сделайте предсказания и вычислите MSE (Mean Squared Error).

## Задание 2: Использование кросс-валидации.

С использованием того же набора данных проведите k-fold кросс-валидацию (k=10) для своей модели и сравните среднее значения MSE на всех фолдах.

## Задание 3: Построение Ridge регрессии.

Примените Ridge регрессию к набору данных "Boston Housing". Подберите гиперпараметр `alpha` через кросс-валидацию.

## Задание 4: Построение Lasso регрессии.

Также примените Lasso регрессию к тем же данным. При подборе гиперпараметра `alpha` через кросс-валидацию сравните количество нулевых весов в модели с результатами Ridge регрессии.

## Задание 5: Использование других функций потерь.

Используйте набор данных diabetes из sklearn.datasets и постройте модель HuberRegressor - линейную модель с функцией потерь Хьюбера, которая менее чувствительна к выбросам по сравнению с MSE.

# Сравнение моделей

## Задание 6: Исследование эффекта масштабирования признаков.

Используйте любой набор данных с числовыми признаками. Тренируйте модели Ridge и Lasso регрессии на исходных данных и предобработанных данных (используйте стандартизацию и нормализацию). Сравните коэффициенты моделей, полученных для исходных и предобработанных данных.

## Задание 7: Изучение влияния регуляризации.

Выберите набор данных с высокой размерностью признаков. Создайте модели Lasso и Ridge регрессии. Проведите эксперименты с различными степенями регуляризации и установите, как они влияют на производительность модели и распределение весов признаков.

## Задание 8: Регрессия с использованием метода эластичной сети.

На том же наборе данных обучите ElasticNet, который объединяет L1 и L2 регуляризацию. Экспериментируйте с разными соотношениями L1 и L2 регуляризации и установите, как это влияет на производительность модели.

## Задание 9: Комбинирование методов отбора признаков и регуляризации.

Выберите подмножество признаков с помощью любого метода отбора признаков, а затем обучите модели с Lasso и Ridge регуляризацией. Сравнивай модели между собой и с моделью, построенной на всех признаках.

## Задание 10: Использование метрик, устойчивых к выбросам.

Используйте набор данных diabetes из sklearn. и обучите модель HuberRegressor, которая менее чувствительна к выбросам по сравнению с MSE-метрикой. Сравните эту модель с базовой моделью линейной регрессии.

В качестве датасета можно использовать данные из

* [репозитория UCI](<http://archive.ics.uci.edu/ml/index.php>),
* [Kaggle Datasets](https://www.kaggle.com/datasets) или
* [Google's Dataset Search](https://datasetsearch.research.google.com/).