

# Лабораторная работа №3 - Регрессия

Наборы данных: lab3\_lin4.csv, lab3\_poly1.csv, Student\_Performance.csv

## 1. Линейная регрессия

- 1.1. Загрузите набор данных соответствующей цифре вашего варианта. Убедитесь, что загрузка прошла корректно.
- 1.2. Используя `train_test_split` разбейте выборку на обучающую и тестовую. Проверьте, что тестовая выборка соответствует обучающей. *Можно использовать диаграммы рассеяния/нормированные гистограммы/boxplot/violin plot.*
- 1.3. Проведите линейную регрессию используя `LinearRegression`. Получите коэффициенты регрессии и объясните полученные результаты.
- 1.4. Для обучающей и тестовой выборки рассчитайте коэффициент детерминации, MAPE, MAE. Объясните полученные значения метрик. Сравните метрики для обучающей и тестовой выборки, сделайте выводы о качестве обобщения полученной модели.
- 1.5. Для модели, которая дала лучшие результаты, постройте диаграмму рассеяния между предикторами и откликом. На диаграмме изобразите какое значение должно быть, и какое предсказывается. Визуально оцените качество построенного регрессора.

## 2. Нелинейная регрессия.

- 2.1. Загрузите набор данных соответствующей букве вашего варианта. Убедитесь, что загрузка прошла корректно.
- 2.2. Используя `train_test_split` разбейте выборку на обучающую и тестовую. Проверьте, что тестовая выборка соответствует обучающей.
- 2.3. Проверьте работу стандартной линейной регрессии на загруженных данных. Постройте диаграмму рассеяния данных с выделенной полученной линией регрессии. Объясните полученный результат.
- 2.4. Конструируя полиномиальные признаки для разных степеней полинома найдите степень полинома наилучшим образом аппроксимирующая данные. Постройте график зависимости коэффициента детерминации от степени полинома (на одном графике изобразите линии для обучающей и тестовой выборки отдельно). Сделайте вывод о том, при какой степени полинома модель начинает переобучаться.
- 2.5. Для выбранной степени полинома, рассчитайте и проанализируйте полученные коэффициенты. Рассчитайте значение метрик коэффициент детерминации, MAPE, MAE.

- 2.6. Для выбранной степени полинома, постройте диаграмму рассеяния данных с линией соответствующей полученному полиному. Сделайте выводы о качестве аппроксимации.

**3. Оценка модели регрессии**

- 3.1. Загрузите набор данных Student\_Performance.csv . Данный набор данных содержит информацию о характеристиках студента, а также качестве его обучения.
- 3.2. Проведите предобработку набора данных - замена текстовых данных, удаление null значений, удаление дубликатов. Разделите на обучающую и тестовую выборку.
- 3.3. Постройте модель, которая будет предсказывать значение признака Performance Index на основе остальных признаков. *Модель выберите самостоятельно.*
- 3.4. Проанализируйте полученную модель. Сделайте выводы о значимости/информативности признаков. Опишите какие проблемы могут возникнуть при применении модели.