

Лабораторная работа 2. ОБУЧЕНИЕ С УЧИТЕЛЕМ. ЗАДАЧА РЕГРЕССИИ

ЗАДАНИЕ.

1. Изучение примеров. Изучите примеры: [Lab2_Ex1_MultiR+PolinomR.ipynb](#), [Lab2_Ex2 PolinomR+Pipeline.ipynb](#), [Lab2_Ex3 ElasticNet+GridSearch.ipynb](#), [Lab2_Ex4 Optuna.ipynb](#)

2. Загрузка и подготовка данных.

- В соответствии с индивидуальным вариантом загрузите предобработанный датасет в формате CSV для решения задачи регрессии.
- Выделите целевой признак и предикторы.
- Разбейте данные на обучающую и тестовую выборки (например, 80% на обучение и 20% на тестирование).

3. Решение задачи регрессии.

- Используйте модели из библиотеки Scikit-learn для решения задачи регрессии с применением регуляризации.
- Для всех моделей регрессии с регуляризацией подберите гиперпараметры тремя способами: GridSearchCV; RandomizedSearchCV; фреймворк Optuna.

Изучите самостоятельно:

1) *Онлайн-документация по фреймворку Optuna.* – <https://optuna.readthedocs.io/en/stable/>

2) *Optuna: подбор гиперпараметров для вашей модели.* – <https://practicum.yandex.ru/blog/optuna-podbor-giperparametrov/>

- Реализуйте один пайплайн для построения любой модели регрессии (см. пример в файле [Lab2_Ex2 PolinomR+Pipeline.ipynb](#)).

3.1. Простая линейная регрессия. Lasso-регрессия (L1-регуляризация). Ridge-регрессия (L2-регуляризация). Elastic Net.

3.2. Множественная линейная регрессия. Lasso-регрессия (L1-регуляризация). Ridge-регрессия (L2-регуляризация). Elastic Net.

3.3. Полиномиальная регрессия.

4. Визуализация. Постройте графики для визуализации результатов решения задачи регрессии (например, графики предсказанных значений против истинных значений).

5. Оценка качества моделей. Вычислите значения метрик оценки качества: R^2 , MAE, MSE, RMSE, MAPE для всех обученных моделей регрессии.

6. Реализация пользовательских функций.

- Самостоятельно реализуйте вычисление всех используемых метрик оценки качества модели регрессии.
- Поместите пользовательские функции в отдельный файл и подключите его к основной программе как библиотеку функций.

7. Создание таблицы результатов.

- Создайте две таблицы (объект DataFrame) и выведите в них наименования используемых регрессоров и значения вычисленных метрик оценки качества как с использованием библиотеки Scikit-learn, так и пользовательских функций (Образец 1 и Образец 2).

- Убедитесь, что при выводе значений предусмотрено необходимое количество знаков после запятой (см. Образец 1 и Образец 2).

Образец 1

| | Train Data | | | | | Test Data | | | | |
|-------------------|----------------|--------|--------|--------|--------|----------------|--------|--------|--------|--------|
| Перепеппор | R ² | MSE | RMSE | MAE | MAPE | R ² | MSE | RMSE | MAE | MAPE |
| Linear Regression | 0.XX | 0.XXXX | 0.XXXX | 0.XXXX | 0.XXXX | 0.XX | 0.XXXX | 0.XXXX | 0.XXXX | 0.XXXX |
| ... | | | | | | | | | | |

Образец 2

| | Scikit-learn | | | | | Кастомные функции | | | | |
|-------------------|----------------|--------|--------|--------|--------|-------------------|--------|--------|--------|--------|
| Перепеппор | R ² | MSE | RMSE | MAE | MAPE | R ² | MSE | RMSE | MAE | MAPE |
| Linear Regression | 0.XX | 0.XXXX | 0.XXXX | 0.XXXX | 0.XXXX | 0.XX | 0.XXXX | 0.XXXX | 0.XXXX | 0.XXXX |
| ... | | | | | | | | | | |

8. Вывод. Напишите вывод о выполненной Лабораторной работе №2, в котором выберите лучшую модель регрессии и обоснуйте свое решение.