**МИНОБРНАУКИ РОССИИ**

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**

**«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**

**Кафедра вычислительной техники**

**ОТЧЕТ**

**По лабораторной работе №3**

**По дисциплине «Машинное обучение» По теме «Регрессия»**

Студент группы 2308 Попов Н.А.

Преподаватель Турсуков Н.О.

Санкт-Петербург 2024

Оглавление

[Цель работы 2](#_Toc185606343)

[Задачи работы 3](#_Toc185606344)

[Используемые метрики 4](#_Toc185606345)

[Обучение моделей и подбор параметров 5](#_Toc185606346)

[1) Инициализация моделей: 5](#_Toc185606347)

[4) Обучение моделей: 5](#_Toc185606348)

[5) Оценка качества моделей: 6](#_Toc185606349)

[7) Хранение и вывод результатов: 7](#_Toc185606350)

[Анализ результатов 7](#_Toc185606351)

[1. Модель Linear Regression: 7](#_Toc185606352)

[2. Модель LASSO: 7](#_Toc185606353)

[3. Модель Ridge: 7](#_Toc185606354)

[Вывод 7](#_Toc185606355)

# Цель работы

Целью данной работы является получение и закрепления навыков предобработки данных и применения методов машинного обучения для решения задач регрессии.

# Задачи работы

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Предварительная обработка данных (при необходимости)
   1. Визуализация значимых признаков (диаграммы рассеяния, ящики с усами, гистограммы)
   2. Очистка данных (удаление пропусков, нормализация, удаление дубликатов)
2. Обучение моделей и подбор параметров (где применимо):
   1. Линейная регрессия
   2. LASSO
   3. Ридж-регрессия
3. Оценка моделей
   1. Вывод метрик
   2. Построение графиков

# Краткая теоретическая информация

**Модели регрессии**

**Linear Regression (Линейная регрессия)**

Линейная регрессия — это статистический метод, который моделирует зависимость между независимыми переменными (предикторами) и зависимой переменной (целью). Модель предполагает, что существует линейная зависимость, и находит коэффициенты, минимизируя ошибку предсказания.

**LASSO (Least Absolute Shrinkage and Selection Operator)**

LASSO — это метод линейной регрессии с добавлением L1-регуляризации, которая штрафует модель за большие коэффициенты. Это помогает

уменьшить переобучение и делает модель более интерпретируемой, удаляя незначимые признаки (через установку их коэффициентов в ноль).

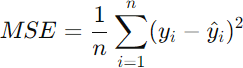
**Ridge Regression**

Ridge — это тоже метод линейной регрессии, но с L2-регуляризацией. В отличие от LASSO, который может полностью исключить признаки, Ridge лишь уменьшает их значения, что помогает избежать переобучения, но все признаки остаются в модели.

# Используемые метрики

**MSE (Mean Squared Error)**

Среднеквадратичная ошибка (MSE) — это метрика, которая измеряет среднюю квадратичную разницу между предсказанными и реальными значениями. MSE более чувствительна к крупным ошибкам, поскольку ошибки возводятся в квадрат



# Обучение моделей и подбор параметров

Для решения задачи классификации были выбраны три модели машинного обучения:

* **Linear Regression**
* **LASSO**
* **Ridge**

Краткое описание создания и обучения моделей регрессии:

# Инициализация моделей:

1. Для решения задачи предсказания цены применяются три модели:
   1. Линейная регрессия (Linear Regression)
   2. LASSO (L1 регуляризация)
   3. Ridge (L2 регуляризация)
2. Модели инициализируются с помощью соответствующих классов из библиотеки sklearn.

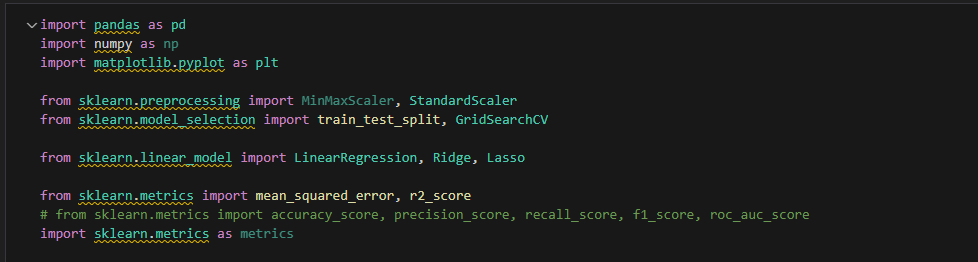


Рисунок 1 – Инициализация моделей

# Обучение моделей:

* 1. Каждая модель обучается на обучающих данных (x\_train и y\_train) с помощью метода .fit().

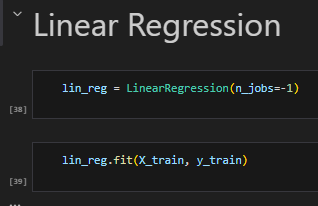


Рисунок 2 – Пример обучения моделей

# Оценка качества моделей:

1. Для каждой модели и каждого набора данных (обучающий и тестовый) вычисляются следующие метрики:

**MSE (среднеквадратическая ошибка)** — среднее значение квадратов разностей между предсказанными и реальными значениями.

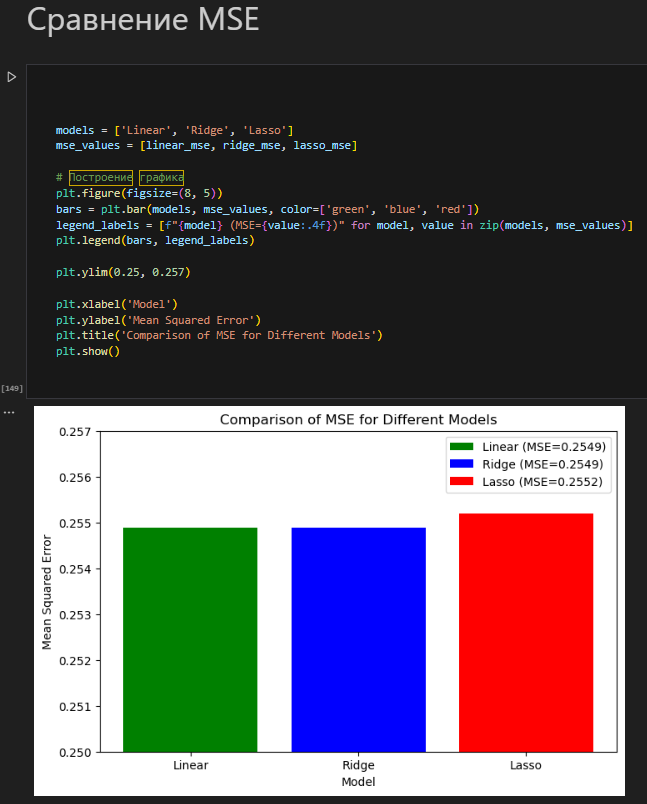


Рисунок 3 – Вывод результатов

# Хранение и вывод результатов:

1. Результаты для каждой модели и каждого набора данных сохраняются в словарь results.
2. Для каждого набора данных (обучающий и тестовый) выводятся значения всех метрик для каждой модели.

# Анализ результатов

# Модель Linear Regression:

* + Средняя абсолютная ошибка (MSE) составила 0.0376 на обучающих данных и 0.0373 на тестовых, что подтверждает высокую точность предсказаний модели.

# Модель LASSO:

* + Ошибка MSE существенно выше по сравнению с другими моделями, что подтверждает плохую прогностическую способность этой модели в данном контексте.

# Модель Ridge:

* + Результаты для модели Ridge аналогичны линейной регрессии.

# Вывод

Модель **линейной регрессии** показала наилучшие результаты как на обучающих, так и на тестовых данных, обеспечивая высокий коэффициент детерминации и низкие значения ошибок. Модель **LASSO** не продемонстрировала хорошую производительность, скорее всего из-за

чрезмерного регуляризующего воздействия, которое оказалось чрезмерным для данного набора данных. Модель **Ridge** продемонстрировала схожие результаты с линейной регрессией, что позволяет сделать вывод о ее

адекватности для задачи предсказания цен в рамках данного анализа