Инструкция по линейной регрессии

Для того, чтобы обучить и использовать линейную регрессию необходимо выполнить следующие шаги:

 Создать виртуальное пространство, написав в терминале команду:

python -m venv venv

2. Установить датасет домов, выполнив в терминале команду:

```
pip install scikit-learn==1.1.3
```

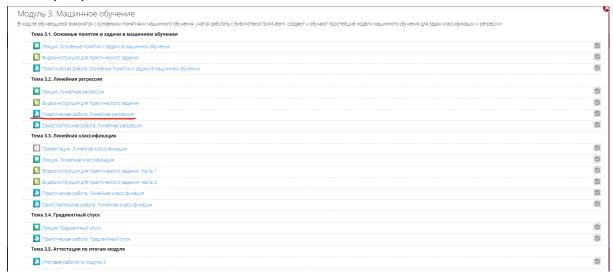
3. Создать файл main.py и написать в нём следующий код:

```
from sklearn.datasets import load_boston #импорт датасета в программу boston = load_boston() #объявление переменной с данными # print(boston.DESCR)

from sklearn.model_selection import train_test_split #импорт функции для разделения датасета на части X_train, X_test, y_train, y_test = train_test_split(boston.data, boston.target, test_size=0.2, random_state=42) #разделение датасета на тестовую и обучающую части #в соотношении 80%: 20%

from sklearn.linear_model import LinearRegression #импорт алгоритма линейной регрессии model = LinearRegression() #создание линейной регрессии model.fit(X_train, y_train) #обучение линейной регрессии housel = [[0.00632, 18.00 , 2.310, 0, 0.5380 , 6.5750 , 65.20 , 4.0900 , 1 , 296.0 , 15.30 ,396.90 , 4.98]] #определяем параметры дома для предсказания цены cost_housel = model.predict(housel) #предсказываем цену дома
```

- 4. Запустить выполнение программы
- 5. Загрузить файл main.py в МЭО -> 3 модуль -> Тема 3.2 Линейная регрессия -> Практическая работа. Линейная регрессия



В датасете содержатся данные о домах в Бостоне, США. На их основе формируется цена жил. площади. Задача линейной регрессии - попытаться предсказать цену дома, на основе данных о доме:

<u>CRIM</u> - Уровень преступности на душу населения по городам.

<u>ZN</u> - Доля земель под жилую застройку, зонированных под участки площадью более 25 000 кв. футов.

<u>INDUS</u> - Доля акров неторгового бизнеса на город.

<u>CHAS</u> - Charles River фиктивная переменная (= 1, если участок граничит с рекой; 0 в противном случае).

<u>NOX</u> - Концентрация оксидов азота (частей на 10 миллионов).

<u>RM</u> - Среднее количество комнат на жилище.

<u>AGE</u> - Пропорция квартир, занимаемых владельцами, построенных до 1940 г.

<u>DIS</u> - Взвешенные расстояния DIS до пяти центров занятости Бостона.

RAD - Индекс доступности радиальных автомагистралей.

<u>ТАХ</u> - Полная ставка налога на имущество за 10 000 долларов США.

<u>PTRATIO</u> - Соотношение учеников и учителей по городам. <u>B</u> - 1000 (Bk - 0,63) $^{\circ}$ 2, где Bk - доля чернокожих по городам. <u>LSTAT</u> - $^{\circ}$ более низкий статус населения.