|  |  |
| --- | --- |
| Группа | R32811 |
| Студент | Филиппов Александр |
| Преподаватель | Евстафьев Олег Александрович |

Отчёт по практической работе №2

Нейронные сети

## Прогнозирование цен на жилье с помощью нейросетевой регрессионной модели

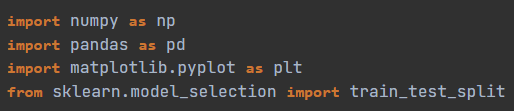
Необходимо по имеющимся данным о ценах на жильё предсказать окончательную цену каждого дома с учетом характеристик домов с использованием нейронной сети. Описание набора данных содержит 80 классов (набор переменых) классификации оценки типа жилья, и находится в файле data\_description.txt.

В работе требуется дополнить раздел «Моделирование» в подразделе

«Построение и обучение модели» создать и инициализировать последовательную модель нейронной сети с помощью фрэймворков тренировки нейронных сетей как: Torch или Tensorflow. Скомпилировать нейронную сеть выбрав функцию потерь и оптимизатор соответственно. Оценить точность полученных результатов. Вывести предсказанные данные о продаже.

## Импорт библиотек

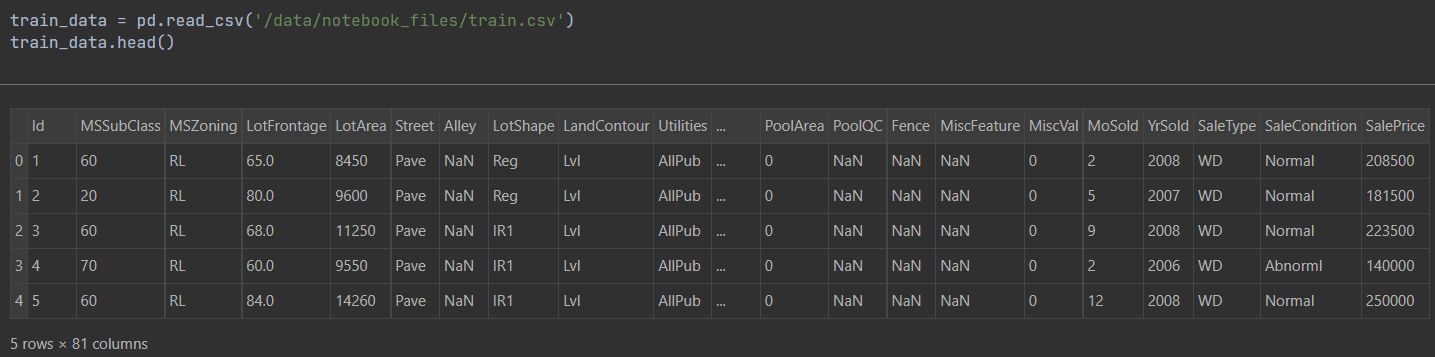
Импортируем необходимые библиотеки:



## Считаем набор данных и отобразим их

Загрузим набор данных и присвоим следующими переменные:

* train\_data: данные, используемые для обучения модели
* test\_data: данные, используемые для проверки модели

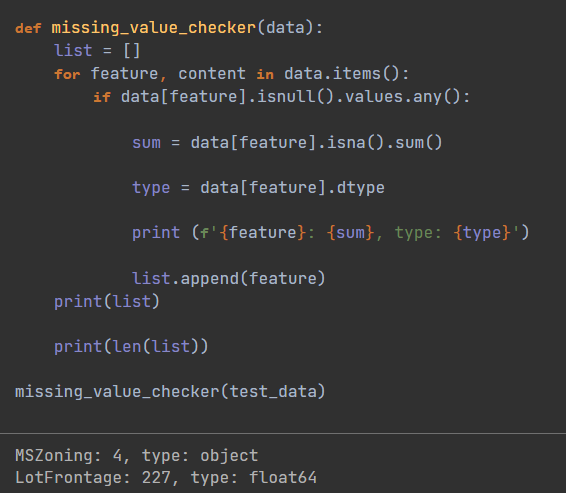


# Подготовка данных

Как можно видеть, train\_data имеет на один столбец больше, чем test\_data, это столбец SalePrice, для обучения модели перед применением ее для предсказания меток в test\_data.

## Проверяем нет ли тестовые данные пустых значений значений (Nan)

Построим функцию def missing\_value\_checker для проверки и подсчёта пропущеных значений в test\_data. А также выведем тип данных этих значений.



Проверяем какие признаки в таблице можно оставить, а какие удалить. Если пропущенных значений слишком много, то удалим признак. Если их

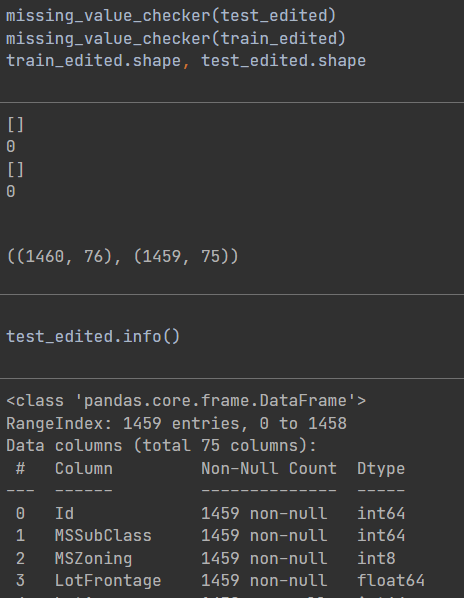
небольшое количество, то заполним mean или median для чисел, новая категория missing для строковых объектов.

В соответствии с этим:

* удалим ['Alley', 'FireplaceQu', 'PoolQC', 'Fence', 'MiscFeature'];
* заполним числовое отсутствующее значение значением mean;
* заполним строковое отсутствующее значение значением missing.

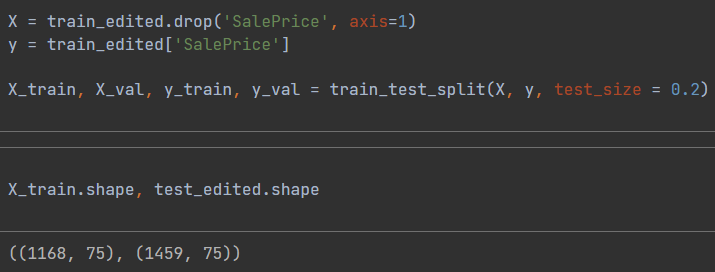


## Перепроверим наши данные:

****

**Разделим данные**

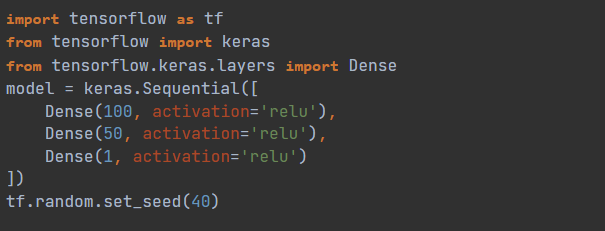
Поскольку мы не знаем метку (Цена) тестовых данных, для оценки модели, чтобы получить лучшую модель перед прогнозированием тестовых данных, разделим данные в файле train.scv на обучающие и проверочные данные, соотношение составляет 20%.



# Моделирование

## Построение и обучение модели

Создаем последовательную модель нейронной сети с помощью фрэймворков тренировки нейронных сетей как: Torch или Tensorflow.



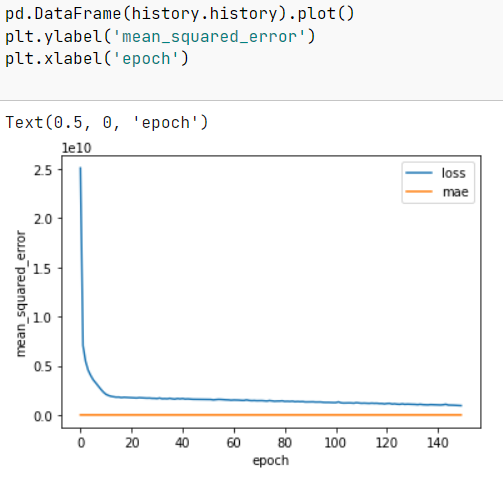
Скомпилируем нейронную сеть, выбрав функцию потерь и оптимизатор соответственно.



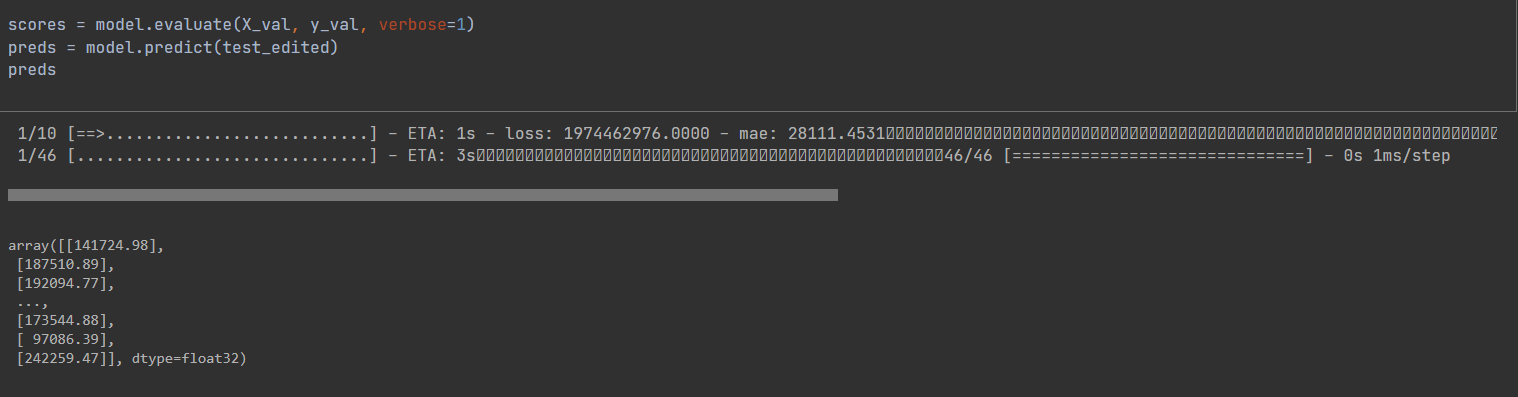
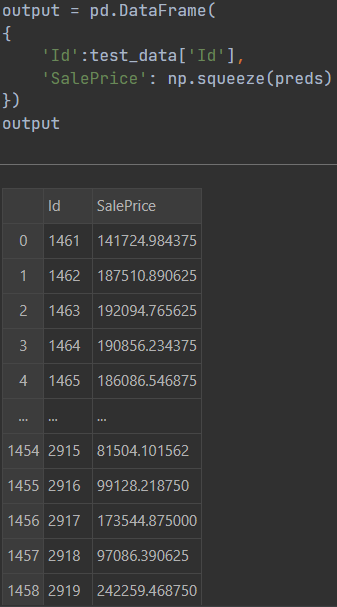
Обучаем модель на обучающих данных X\_train и y\_train задав гиперпараметры нашей модели нейронной сети, например количество эпох (epochs), размер мини-выборки (batch\_size) и другие.



## Оценим полученные результаты

****

**Предсказание**

**  
**

## Проведем тест для других значений нейронов на входном слое :

При увеличении

**Вывод :** с увеличением количества нейронов на входном слое уменьшается прогнозируемая цена.

## Проведем тест для других значений эпох :

**Вывод :** с увеличение эпох график становится более сглаженным, также изменяются прогнозируемые цены, уменьшается среднеквадратическая ошибка.

## Проведем тест для других значений размеров выборки :

**Вывод :** с увеличение количества размеров выборки изменяются цены, меняется изменение среднеквадратической ошибки.

**Ответы на вопросы :**

1. Изменение параметров влияет на результат обучения. Недостаточное количество эпох приводит к недостатку обучения, а слишком большое количество нейронов приводит к переобучению.
2. Эпоха – это прохождение всего набора данных через нейронную сеть в прямом и обратном направлении 1 раз, итерация – количество пакетов, необходимых для завершения 1ой эпохи.
3. Функция активации определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения. Благодаря функциям активации нейронные сети способны порождать более информативные признаковые описания, преобразуя данные нелинейным образом. В процессе работы удалось применить softmax и relu.
4. MSE – способ измерения качества модели, чем ниже, тем лучше. MAE – тоже способ измерения качества модели, более чувствителен к выбросам.