

Проект по дисциплине
Проориентационный семинар «Введение в специальность»

Решение задачи регрессии с помощью нейронных сетей.

Предсказание стоимости ноутбуков по
их характеристикам.



Студент бакалавриата
«Прикладная математика»
Морозов Денис

Цели и задачи проекта

→ Цель - разработать модель нейронной сети, с помощью которой возможно спрогнозировать цены на ноутбуки в зависимости от набора их характеристик с неплохой точностью.

Задачи:

1. Изучить принципы и современные методы регрессионного анализа с помощью нейросетей и ознакомиться со способами их реализации на языке Python.
2. Выбрать наиболее подходящий метод, основанный на нейросетевых технологиях, и подробнее изучить его.
3. Провести предварительный анализ и обработку выбранных данных для дальнейшей работы с ними.
4. Разработать нейронную сеть согласно выбранной модели для предсказания цен на ноутбуки на основе обработанных данных.
5. Оценить результат работы созданной модели.

Использование библиотек Python



01

Keras

Создание и настройка
модели нейронной сети

02

Scikit-learn

Обработка данных для
машинного обучения



03

Pandas

Предобработка данных

04

NumPy

Работа с данными

05

Matplotlib

Визуализация данных

База данных

Для обучения модели был выбран
датасет «Laptop Price» (размера 1303x13)
с сайта www.kaggle.com

	laptop_ID	Company	Product	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_euros
0	1	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	1339.69
1	2	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	898.94
2	3	HP	250 G6	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	575.00
3	4	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	2537.45
4	5	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	1803.60
5	6	Acer	Aspire 3	Notebook	15.6	1366x768	AMD A9-Series 9420 3GHz	4GB	500GB HDD	AMD Radeon R5	Windows 10	2.1kg	400.00
6	7	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.2GHz	16GB	256GB Flash Storage	Intel Iris Pro Graphics	Mac OS X	2.04kg	2139.97
7	8	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	256GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	1158.70
8	9	Asus	ZenBook UX430UN	Ultrabook	14.0	Full HD 1920x1080	Intel Core i7 8550U 1.8GHz	16GB	512GB SSD	Nvidia GeForce MX150	Windows 10	1.3kg	1495.00
9	10	Acer	Swift 3	Ultrabook	14.0	IPS Panel Full HD 1920x1080	Intel Core i5 8250U 1.6GHz	8GB	256GB SSD	Intel UHD Graphics 620	Windows 10	1.6kg	770.00

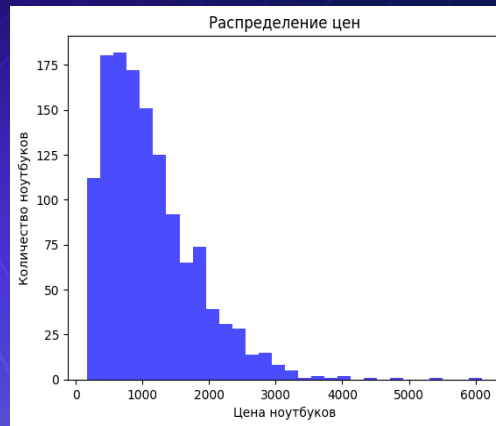
Первые 10 строк датасета



Обработка данных

Для обработки данных были приняты следующие решения:

- Разбить столбец ScreenResolution на столбцы IPS (0/1), Touchscreen (0/1), res_X и res_Y (разрешение экрана по горизонтали и вертикали)
- Разбить столбец Cpu на столбцы Cpu_name и Cpu_freq (частота)
- В столбце Memory перевести всё в GB и разбить на четыре столбца SSD, HDD, Hybrid, Flash Storage
- Убрать GB и kg в столбцах Ram и Weight
- Удалить столбцы Product и Gpu, так как они содержат слишком много уникальных значений относительно общего количества строк в датасете
- Закодировать категориальные значения с помощью OneHotEncoder из sklearn



Средняя цена – 1123 евро
Минимальная – 174 евро
Максимальная – 6099 евро

	Company	TypeName	Inches	Ram	OpSys	Weight	Price_euros	Ips	Touchscreen	res_X	res_Y	Cpu_name	Cpu_freq	SSD	HDD	Flash Storage	Hybrid
0	Apple	Ultrabook	13.3	8	macOS	1.37	1339.69	1	0	1600	1600	Intel Core i5	2.3	128	0	0	0
1	Apple	Ultrabook	13.3	8	macOS	1.34	898.94	0	0	900	900	Intel Core i5	1.8	0	0	128	0
2	HP	Notebook	15.6	8	No OS	1.86	575.00	0	0	1080	1080	Intel Core i5	2.5	256	0	0	0
3	Apple	Ultrabook	15.4	1	macOS	1.83	2537.45	1	0	1800	1800	Intel Core i7	2.7	512	0	0	0
4	Apple	Ultrabook	13.3	8	macOS	1.37	1803.60	1	0	1600	1600	Intel Core i5	3.1	256	0	0	0

Обработанный датасет (до кодирования OneHotEncoder-ом)

Архитектура нейронной сети

Количество нейронов

normalizer

Слой
нормализации



Модель многослойного персептрона
(Sequential() из Keras)



300

Входной
слой

Функция активации ReLU, L2-регуляризация

Dropout(0.5)

1200

Первый
скрытый
слой

Функция активации ReLU, L2-регуляризация

Dropout(0.5)

200

Второй
скрытый
слой

Функция активации ReLU, L2-регуляризация

1

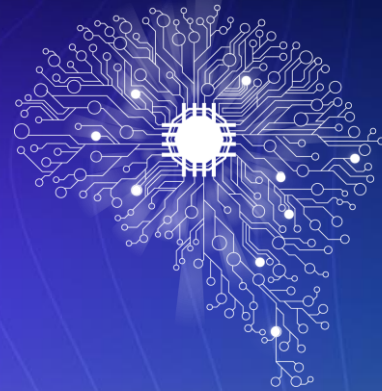
Выходной
слой

Линейная функция активации



Предсказанное
значение

Обучение модели



Тренировочная
выборка: 85%

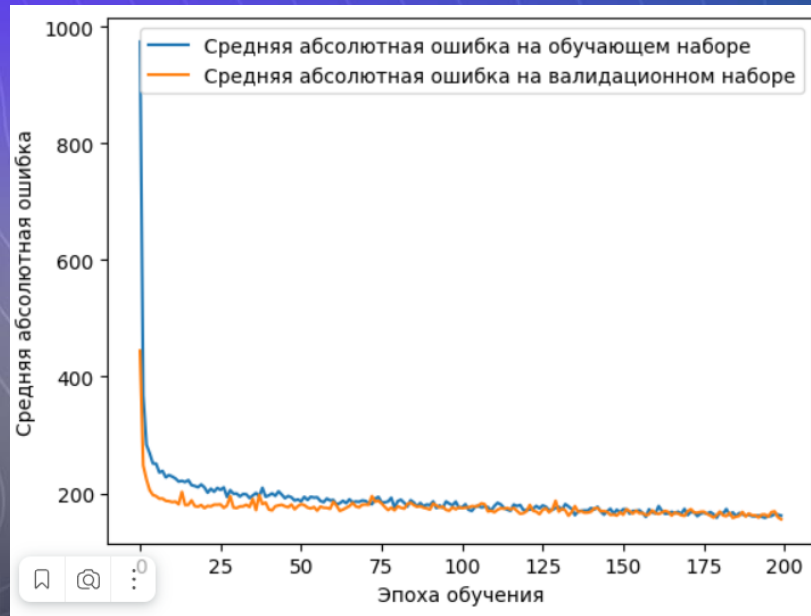
Тестовая
выборка: 15%

Количество эпох
обучения: 200

Оптимизатор:
Adam

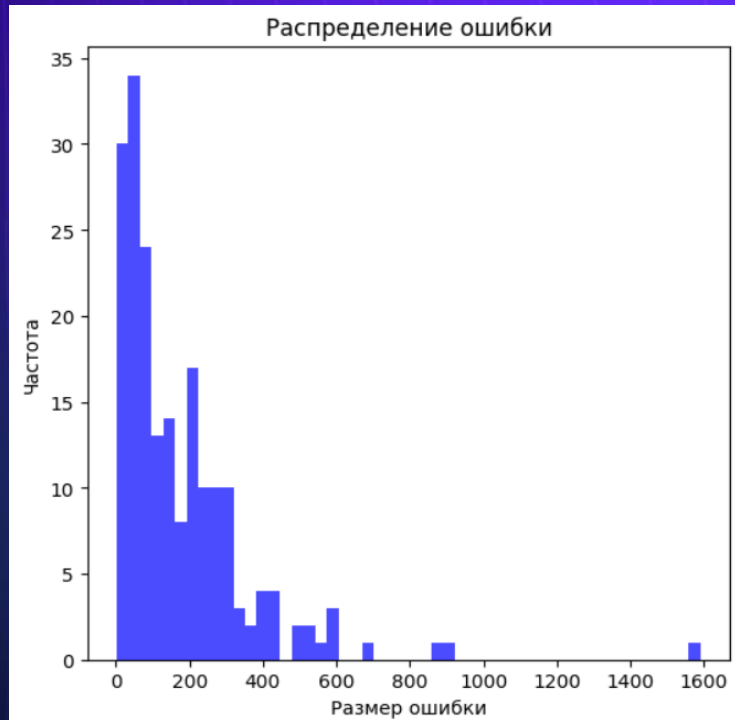
Функция потерь:
MSE

Метрика: MAE



Оценка результата

Гистограмма распределения
абсолютной ошибки на
тестовом наборе:



175 евро

Средняя абсолютная
ошибка на тестовом
наборе данных





Спасибо за внимание!

