

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ
по лабораторной работе №3
по дисциплине «Искусственные нейронные сети»
Тема: Регрессионная модель изменения цен на дома в Бостоне

Студент гр. 8383

Колмыков В.Д.

Преподаватель

Санкт-Петербург

2020

Цель работы.

Реализовать предсказание медианной цены на дома в пригороде Бостона в середине 1970-х по таким данным, как уровень преступности, ставка местного имущественного налога и т. д.

Данный набор содержит относительно немного образцов данных: всего 506, разбитых на 404 обучающих и 102 контрольных образца. И каждый признак во входных данных (например, уровень преступности) имеет свой масштаб. Например, некоторые признаки являются пропорциями и имеют значения между 0 и 1, другие — между 1 и 12 и т. д.

Ход работы.

Был реализован код из методических указаний.

1. Различия задач классификации и регрессии.

- Классификация определяет принадлежность (вероятность принадлежности) входных данных к одному из конечного числа классов. Регрессия предсказывает числовое значение неизвестной функции.
- Соответственно, классификация на выходе дает вектор вероятностей (числа от 0 до 1), регрессия — число любой величины.
- Следовательно, обычно, модель ИНС для классификации заканчивается слоем с линейной активирующей функцией, для регрессии — сигмоидальной или softmax (бинарная или множественная классификация)

2. Изучение влияния кол-ва эпох.

Программа со 100 эпохами обучения выдала среднее абсолютное отклонение для валидационных данных в 2.5 (2.5 тыс. долларов для изучаемой системы).

Увеличение числа эпох до 120 не дало уменьшение ошибки.

Обучение со 150 эпохами выдала значение средней ошибки 2.7.

Кол-во эпох было уменьшено до 70. В результате средняя ошибка оказалась равна 2.3.

При 50 эпохах среднее значение ошибки оказалось 2.5.

При 60 эпохах среднее значение ошибки оказалось 2.4.

При 80 эпохах среднее значение ошибки 2.4.

Следовательно, переобучение происходит примерно при кол-ве эпох более 70.

Графики ошибки при обучении и валидации в 200 эпох для каждой из моделей приведены на рис. 1-4.

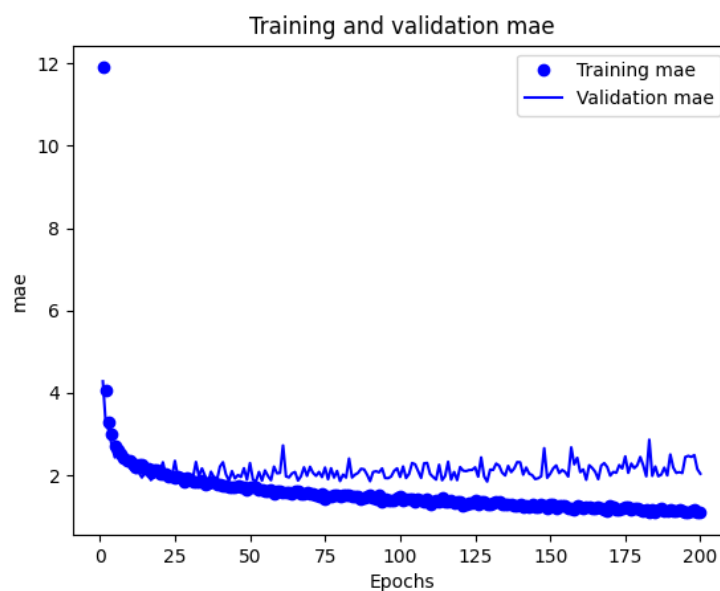


Рисунок 1 – Ошибки при обучении и валидации первой модели

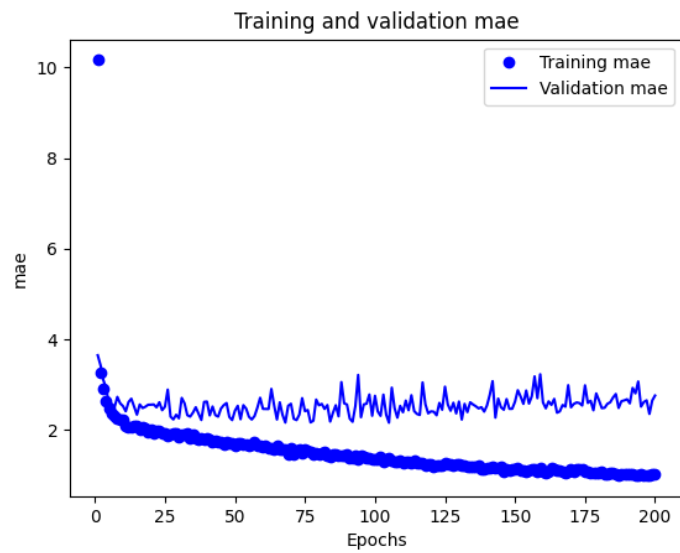


Рисунок 2 – Ошибки при обучении и валидации второй модели

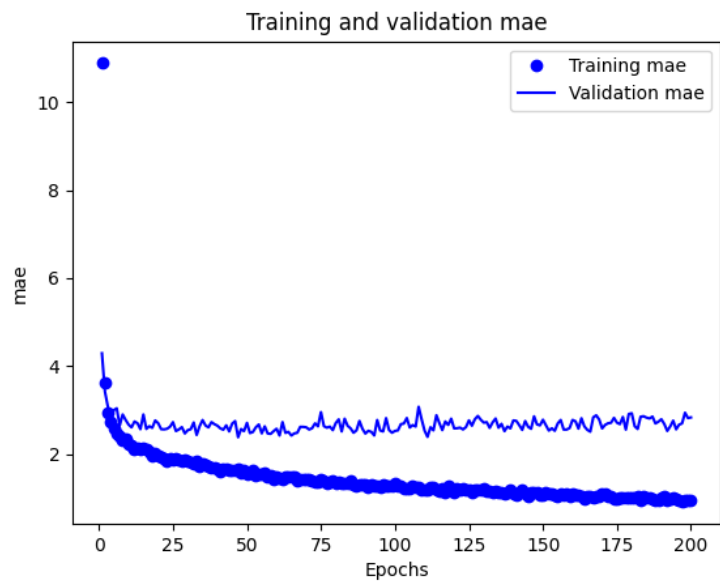


Рисунок 3 – Ошибки при обучении и валидации третьей модели

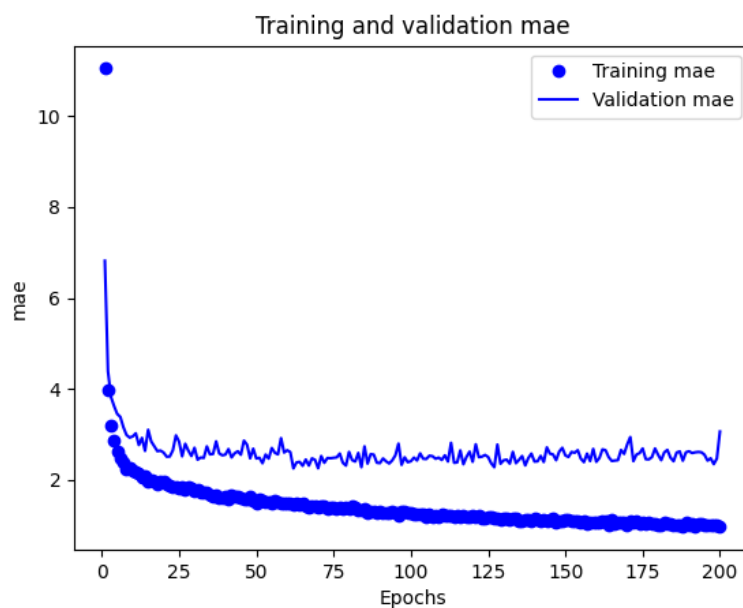


Рисунок 4 – Ошибки при обучении и валидации четвертой модели

Усредненный график по всем моделям приведен на рис. 5.

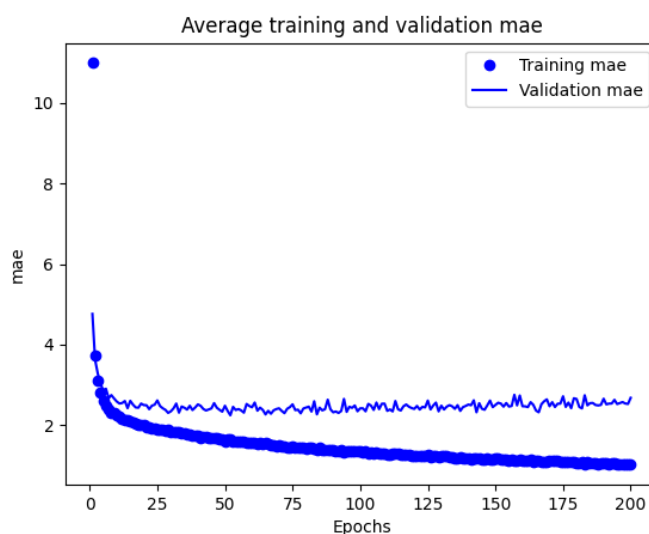


Рисунок 5 – Усредненный график ошибок по всем моделям

Из графиков видно, что серьезно уменьшаться ошибка перестает примерно после 10 эпох. Также можно заметить (хоть это и не очевидно по этим графикам), что ошибка на графиках становится минимальной примерно при 70 эпохах. Следовательно точка переобучения находится около 70.

3. Изучение влияния кол-ва блоков (K).

Результаты при $k = 4$ были приведены выше.

Число блоков было увеличено до 8. Разброс ошибок в прогонах увеличился: от 1.9 до 3. Среднее же значение осталось 2.3. Логично, что разница между графиками на разных итерациях довольно велика. На каждой итерации получались разные графики. Два из них приведены на рис. 6.

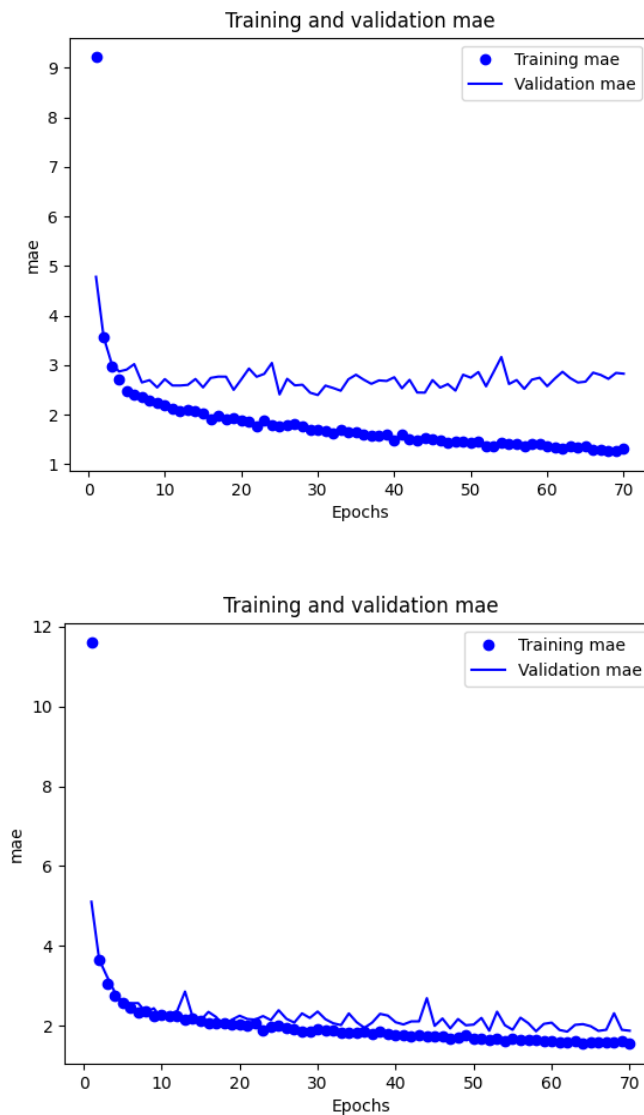


Рисунок 6 – Графики ошибок для двух из восьми моделей

Усредненный график приведен на рис. 7.

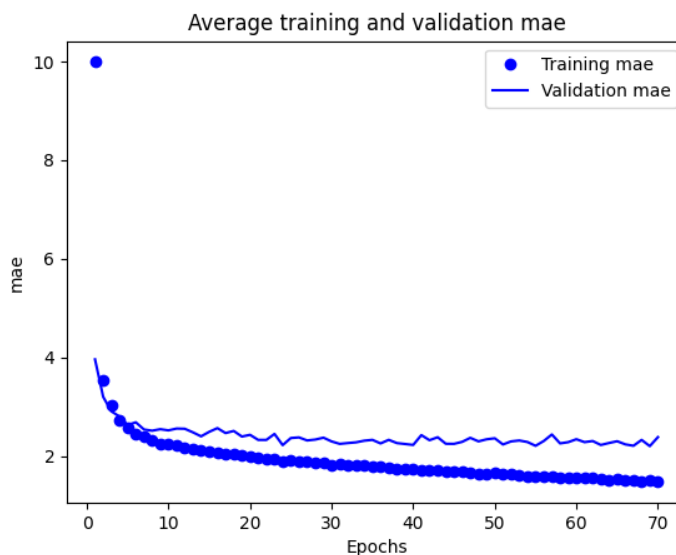


Рисунок 7 – Усредненный график для восьми блоков

При 6 блоках средняя ошибка все еще оказалась равной 2.3, разброс был от 1.8 до 2.8. Усредненный график представлен на рис. 8.

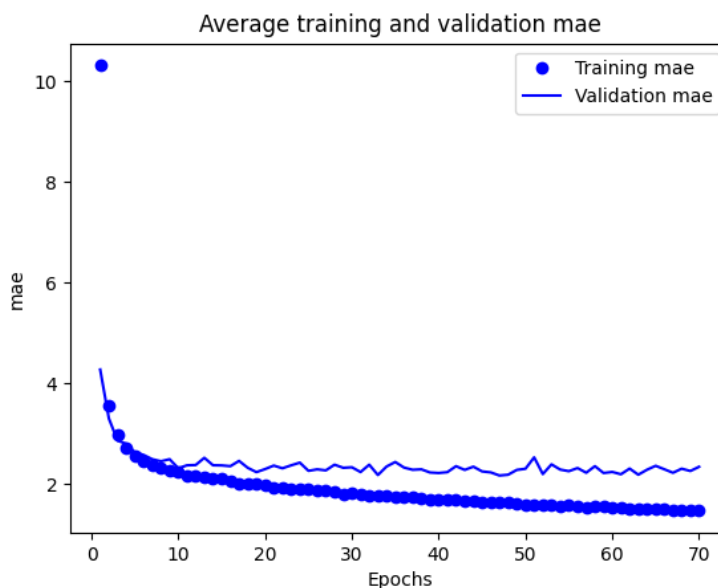


Рисунок 8 – Усредненный график для шести блоков

Число блоков было уменьшено до 3. Разброс ошибок в прогонах оказался небольшим: от 2.44 до 2.57. Среднее значение ошибки увеличилось до 2.5. Это может быть связано с увеличением размера валидационных выборок (треть от

общего числа данных) при небольшом кол-ве данных. Усредненный график представлен на рис. 9.

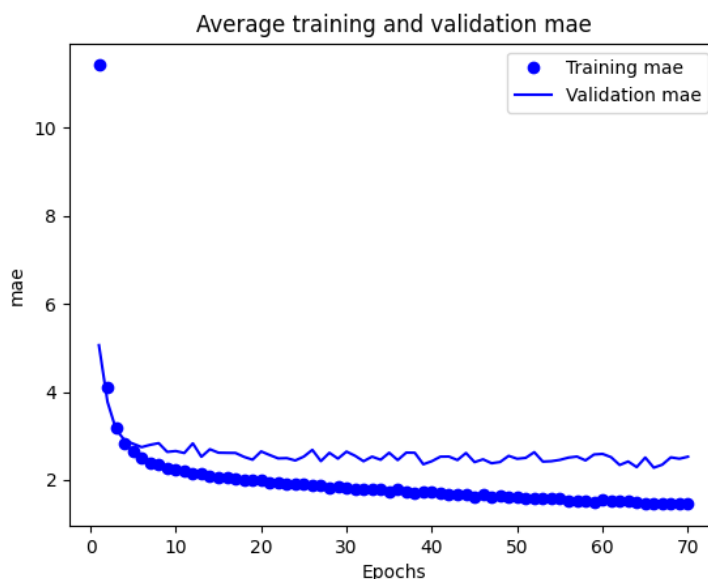


Рисунок 9 – Усредненный график для трех блоков

Из результатов исследования можно сделать вывод, что для данных заданного размера число блоков в 4 является оптимальным: достигается наибольший уровень точности при наименьшем времени обучения.

Выводы.

В ходе выполнения работы было реализовано предсказание медианной цены на дома в пригороде Бостона в середине 1970-х по разным данным. Был изучен метод перекрестной проверки по K блокам, а также влияние на него числа эпох и кол-ва блоков. Исследование было подкреплено результатами работы моделей, графиками ошибок моделей, а также усредненными графиками.