

Отчёт по лабораторной работе
«Полносвязные нейронные сети прямого распространения»
Вариант нечётный.

Выполнил студент гр. 932101

Печкина К. Е.

Преподаватель

Аксёнов С. В.

2024 г.

Бинарный классификатор

Набор данных:

Для работы с бинарными классификаторами использовался набор данных Diabetes Health Indicators Dataset. Оценка вероятности диагностики диабета у человека.

Значение признаков:

1. Diabetes_012 – целевой признак. Класс отрицательный – 0 (no diabetes – нет диабета), класс положительный – 1 & 2 (prediabetes – преддиабетическое состояние & diabetes - диабет)
2. HighBP – высокое артериальное давление. 0 – невысокое, 1 – высокое.
3. HighChol – высокий уровень холестерина. 0 – невысокий, 1 – высокий.
4. CholCheck – была ли проверка холестерина в течение 5 лет. 0 – была, 1 – не было.
5. BMI – индекс массы тела
6. Smoker – человек курил не менее 100 сигарет за всю жизнь
7. Stroke – был ли у человека инсульт
8. HeartDiseaseorAttack – наличие ишемической болезни сердца или инфаркта миокарда
9. PhysActivity – наличие физической активности за последние 30 дней
10. Fruits – употребление фруктов 1 и более раз в день
11. Veggies – употребление овощей 1 и более раз в день
12. HvyAlcoholConsump – употребление более 14 алкогольных напитков в неделю для мужчин или более 7 для женщин
13. AnyHealthcare – наличие любого медицинского страхования
14. NoDocbcCost – была ли необходимость обратиться к врачу за последние 12 месяцев, но человек отказался из-за стоимости
15. GenHlth – оценка здоровья по шкале 1-5
16. MentHlth – в течение скольких дней за последний месяц психическое здоровье человека было не очень хорошим. Шкала 1 – 30 дней.
17. PhysHlth – в течение скольких дней за последний месяц физическое здоровье человека было не очень хорошим. Шкала 1 – 30 дней.

18. DiffWalk – есть ли серьезные трудности с ходьбой или подъемом по лестнице

19. Sex – пол. 0 – женщина, 1 – мужчина.

20. Age – возраст. 13 категорий. 1 – от 18 до 24, 9 – от 60 до 64, 13 – от 80 и более

21. Education – образование. Шкала 1 – 6. 1- никогда не посещал школу или только детский сад, 2 – классы с 1 по 8, 3 – классы с 9 по 11, 4 – 12 класс или сдал экзамен GED, 5 – колледж от 1 года до 3 лет, 6 – колледж 4 года и более.

22. Income – доход. Шкала 1 – 8. 1 – менее \$10,000, 2 – менее \$15,000, 3 – менее \$20,000, 4 – менее \$25,000, 5 – менее \$35,000, 6 – менее \$50,000, 7 – менее \$75,000, 8 – \$75,000 и более.

Набор данных не содержал пропущенных значений, но в нем присутствовало несколько дубликатов, которые были удалены. Значения 2 в целевом столбце были преобразованы в 1, так как этого требовало условие задания.

Для анализа и обучения бинарного классификатора были использованы все признаки из набора данных, кроме целевого.

Датасет был разделен на 3 части: тренировочную (70%), тестовую (20%) и валидационную (10%).

Нейронные сети:

Всего было создано 4 бинарных классификатора, которые отличались своими параметрами:

1. Base Model:

Слой:

1. 4 нейрона, функция активации - ReLU
2. 1 нейрон, функция активации – Sigmoid

Функция потерь: Binary Crossentropy

Процедура оптимизации: Adam

Метрика: Ассигасу

Число объектов в батче: 10

Количество эпох: 100

Модель закончила обучение на 17-ой эпохе, достигнув лучшего значения ассигасу на валидационной выборке на 7-ой эпохе.

2. Second Model:

Слой:

1. 4 нейрона, функция активации - ReLU
2. 4 нейрона, функция активации - Sigmoid
3. 4 нейрона, функция активации - ReLU
4. 4 нейрона, функция активации - Sigmoid
5. 2 нейрона, функция активации - ReLU
6. 1 нейрон, функция активации - Sigmoid

Функция потерь: Binary Crossentropy

Процедура оптимизации: Adam

Метрика: Ассурасу

Число объектов в батче: 10

Количество эпох: 100

Модель закончила обучение на 15-ой эпохе, достигнув лучшего значения ассурасу на валидационной выборке на 5-ой эпохе.

3. Third Model:

Слои:

1. 64 нейрона, функция активации – ReLU
2. 64 нейрона, функция активации – Sigmoid
3. 32 нейрона, функция активации – ReLU
4. 16 нейронов, функция активации – Sigmoid
5. 8 нейронов, функция активации – ReLU
6. 1 нейрон, функция активации - Sigmoid

Функция потерь: Binary Crossentropy

Процедура оптимизации: Adam

Метрика: Ассурасу

Число объектов в батче: 100

Количество эпох: 50

Модель закончила обучение на 22-ой эпохе, достигнув лучшего значения ассурасу на валидационной выборке на 12-ой эпохе.

4. Fourth Model:

Слои:

1. 64 нейрона, функция активации – ReLU
2. Dropout(0.1)
3. 64 нейрона, функция активации – Sigmoid
4. Dropout(0.1)
5. 32 нейрона, функция активации – ReLU

6. Dropout(0.1)
7. 16 нейронов, функция активации – Sigmoid
8. 8 нейронов, функция активации – ReLU
9. 1 нейрон, функция активации - Sigmoid

Функция потерь: Binary Crossentropy

Процедура оптимизации: SGD

Метрика: Ассурасу

Число объектов в батче: 100

Количество эпох: 50

Модель закончила обучение на 40-ой эпохе, достигнув лучшего значения ассурасу на валидационной выборке на 30-ой эпохе.

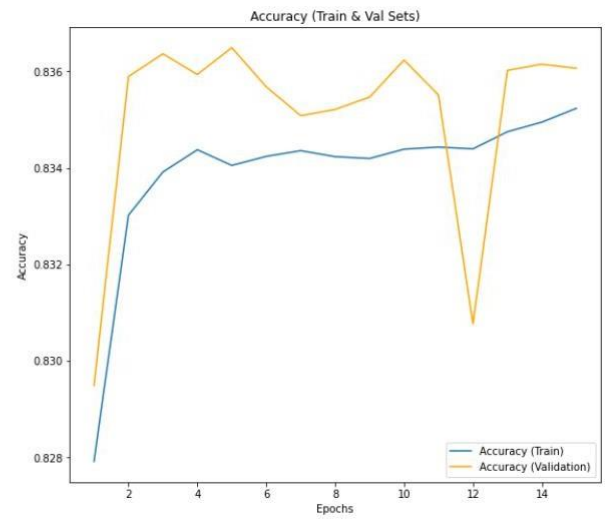
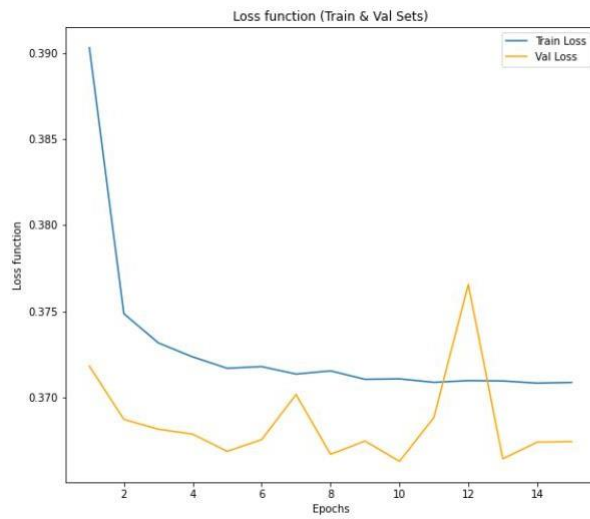
Метрики качества:

Ниже представлена таблица с метриками качества каждой модели. Оценка качества бинарных классификаторов происходила по следующим метрикам: Recall, Precision, Weighted Accuracy и AUC.

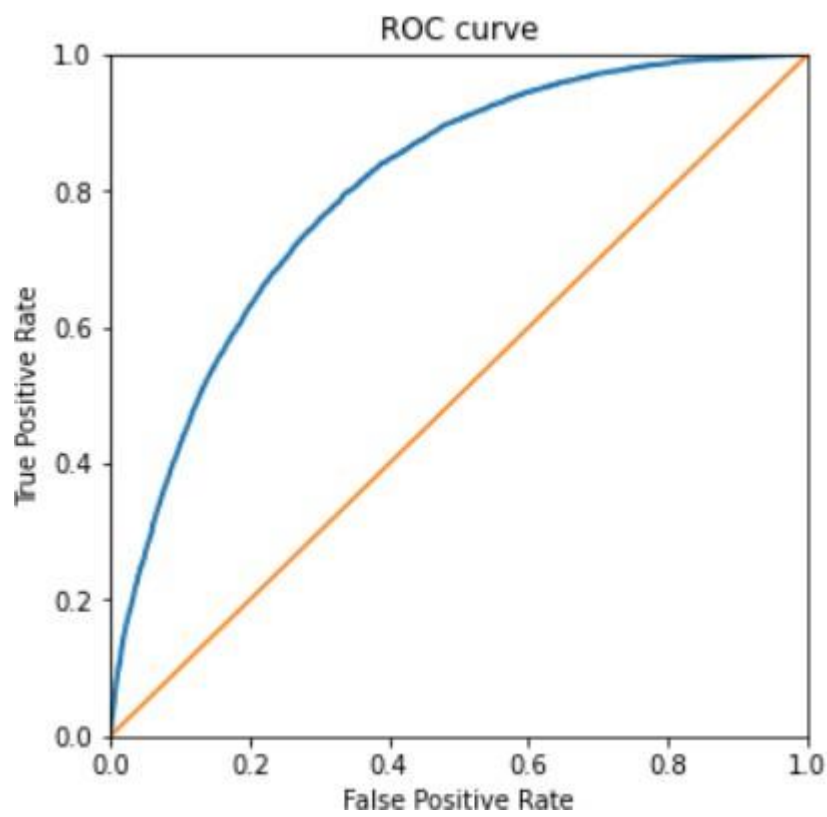
	model	recall	precision	weighted_accuracy	auc
0	base_model	0.1277	0.6110	0.5556	0.8023
1	second_model	0.1958	0.5608	0.5823	0.8032
2	third_model	0.1830	0.5674	0.5774	0.8030
3	fourth_model	0.1476	0.5904	0.5634	0.7990

Метрики качества у Second Model в среднем выше, чем у остальных моделей.

Графики обучения для лучшего бинарного классификатора:



ROC-кривая для лучшего бинарного классификатора:



Многоклассовый классификатор

Набор данных:

Для работы с многоклассовыми классификаторами использовался набор данных Body Performance Data. Оценка уровня физического развития людей разного возраста.

Значение признаков:

1. age – возраст
2. gender – пол
3. height_cm – рост в сантиметрах
4. weight_kg – вес в килограммах
5. body fat_% – процент жира
6. diastolic – нижнее артериальное давление
7. systolic – верхнее артериальное давление
8. gripForce – сила захвата
9. sit and bend forward_cm – наклон в сантиметрах
10. sit-ups counts – количество приседаний
11. broad jump_cm – длина прыжка в сантиметрах
12. class – целевой признак. A, B, C, D

Набор данных не содержал пропущенных значений, но в нем присутствовал дубликат, который был удален. Значения в столбцах gender и class были закодированы.

Для анализа и обучения многоклассового классификатора были использованы все признаки из набора данных, кроме целевого.

Датасет был разделен на 3 части: тренировочную (70%), тестовую (20%) и валидационную (10%).

Нейронные сети:

Всего было создано 4 многоклассовых классификатора, которые отличались своими параметрами:

1. Base Model:

Слои:

1. 8 нейронов, функция активации - ReLU
2. 4 нейрона, функция активации – Softmax

Функция потерь: Categorical Crossentropy

Процедура оптимизации: Adam

Метрика: Accuracy

Число объектов в батче: 10

Количество эпох: 100

Модель закончила обучение на 58-ой эпохе, достигнув лучшего значения accuracy на валидационной выборке на 48-ой эпохе.

2. Second Model:

Слои:

1. 8 нейронов, функция активации - ReLU
2. 8 нейронов, функция активации - Sigmoid
3. 8 нейронов, функция активации - ReLU
4. 8 нейронов, функция активации - Sigmoid
5. 8 нейронов, функция активации - ReLU
6. 4 нейрона, функция активации - Softmax

Функция потерь: Categorical Crossentropy

Процедура оптимизации: Adam

Метрика: Accuracy

Число объектов в батче: 10

Количество эпох: 100

Модель закончила обучение на 29-ой эпохе, достигнув лучшего значения accuracy на валидационной выборке на 19-ой эпохе.

3. Third Model:

Слои:

1. 64 нейрона, функция активации – ReLU
2. 64 нейрона, функция активации – Sigmoid
3. 32 нейрона, функция активации – ReLU
4. 16 нейронов, функция активации – Sigmoid
5. 8 нейронов, функция активации – ReLU
6. 1 нейрон, функция активации - Softmax

Функция потерь: Categorical Crossentropy

Процедура оптимизации: Adam

Метрика: Accuracy

Число объектов в батче: 100

Количество эпох: 50

Модель закончила обучение на 45-ой эпохе, достигнув лучшего значения ассурасу на валидационной выборке на 35-ой эпохе.

4. Fourth Model:

Слои:

1. 128 нейронов, функция активации – ReLU
2. Dropout(0.1)
3. 64 нейрона, функция активации – Sigmoid
4. Dropout(0.1)
5. 32 нейрона, функция активации – ReLU
6. 16 нейронов, функция активации – Sigmoid
7. 8 нейронов, функция активации – ReLU
8. 1 нейрон, функция активации - Softmax

Функция потерь: Categorical Crossentropy

Процедура оптимизации: RMSprop

Метрика: Ассурасу

Число объектов в батче: 100

Количество эпох: 100

Модель закончила обучение на 100-ой эпохе, достигнув лучшего значения ассурасу на валидационной выборке на 100-ой эпохе.

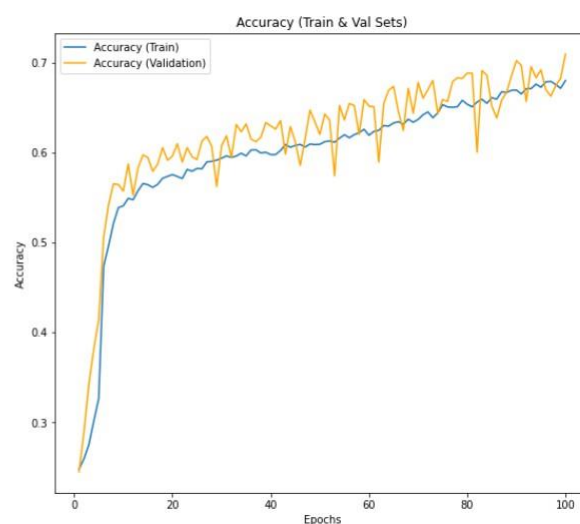
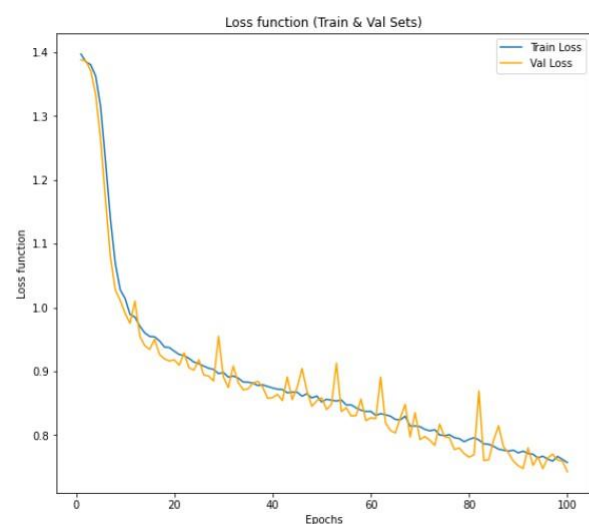
Метрики качества:

Ниже представлена таблица с метриками качества каждой модели. Оценка качества многоклассовых классификаторов происходила по следующим метрикам: Recall, Precision, Weighted Accuracy и AUC.

	model	recall	precision	weighted_accuracy	auc
0	base_model	0.6025	0.5948	0.6025	0.8500
1	second_model	0.6354	0.6503	0.6354	0.8594
2	third_model	0.6826	0.6988	0.6826	0.8872
3	fourth_model	0.7032	0.7200	0.7032	0.8967

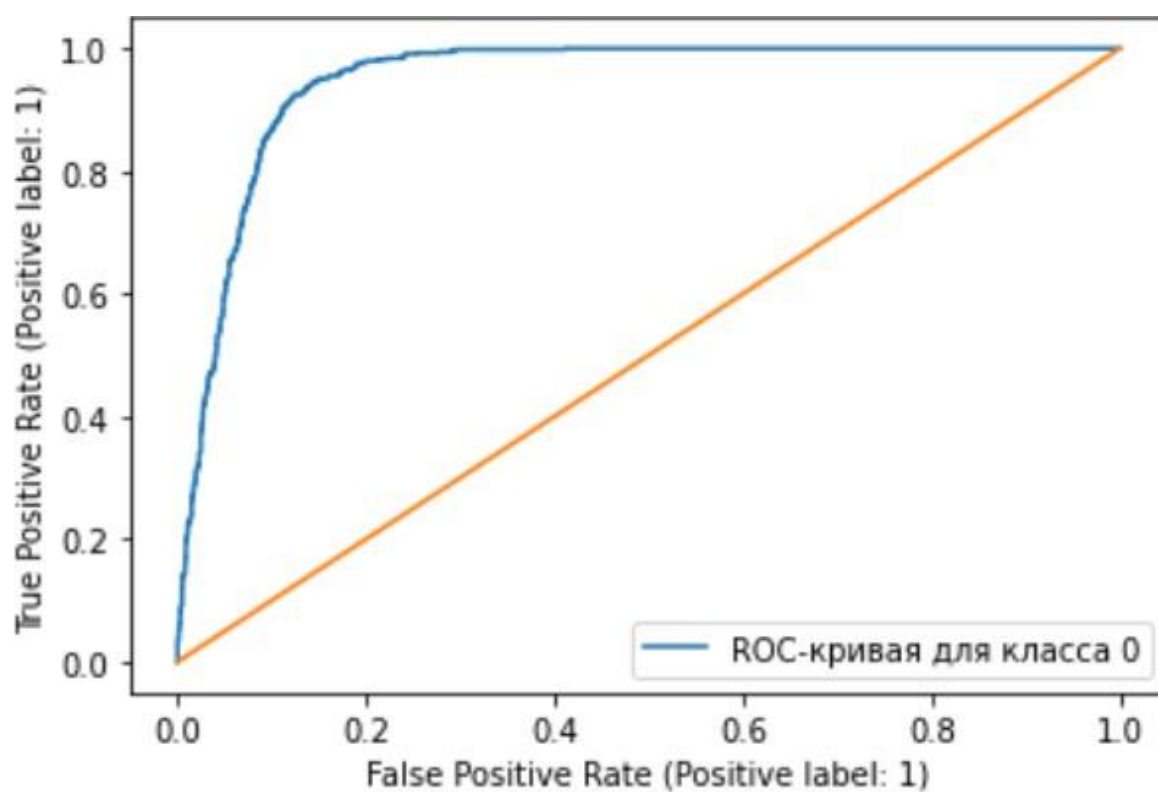
Метрики качества у Fourth Model выше, чем у остальных моделей.

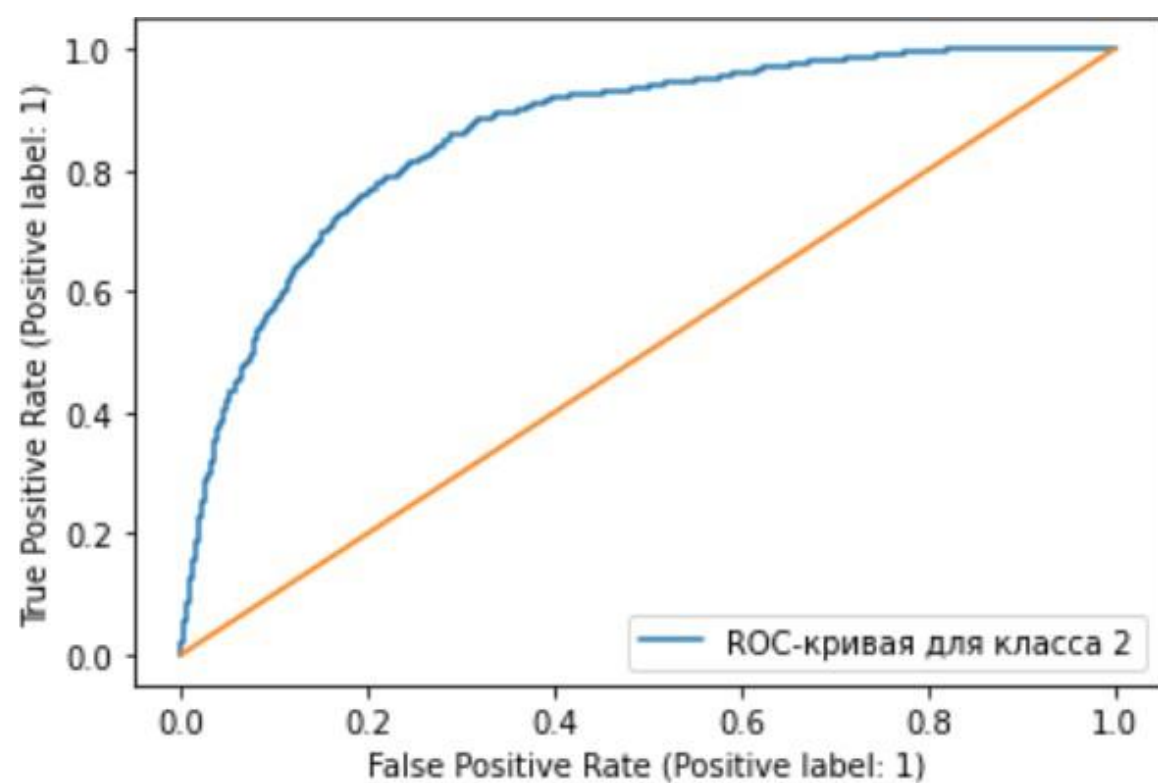
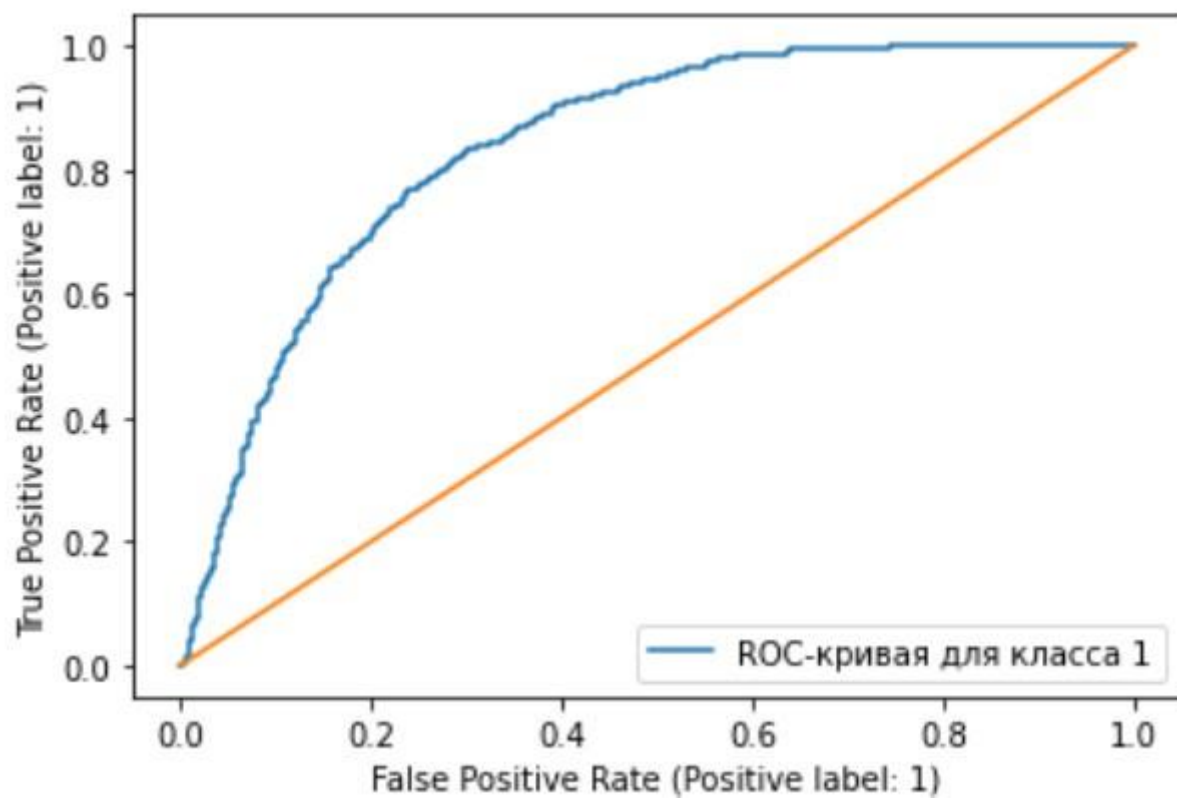
Графики обучения для лучшего многоклассового классификатора:

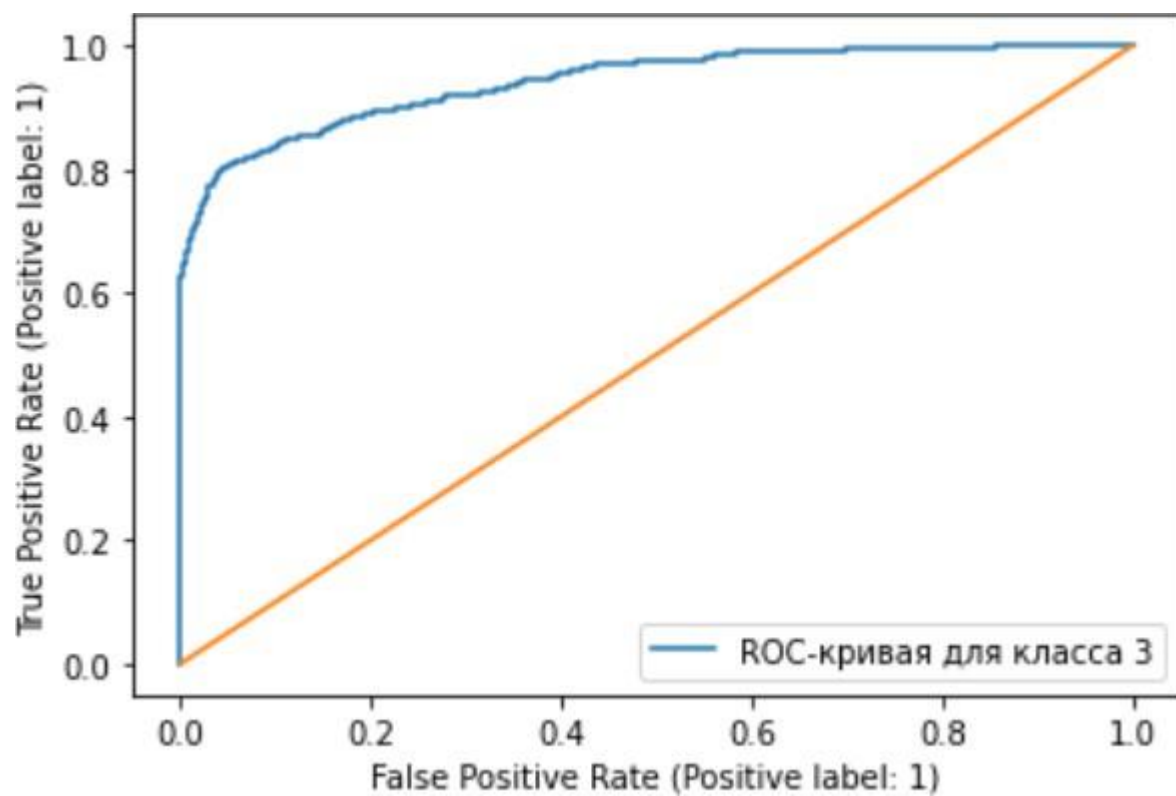


ROC-кривая классов для лучшего многоклассового классификатора:

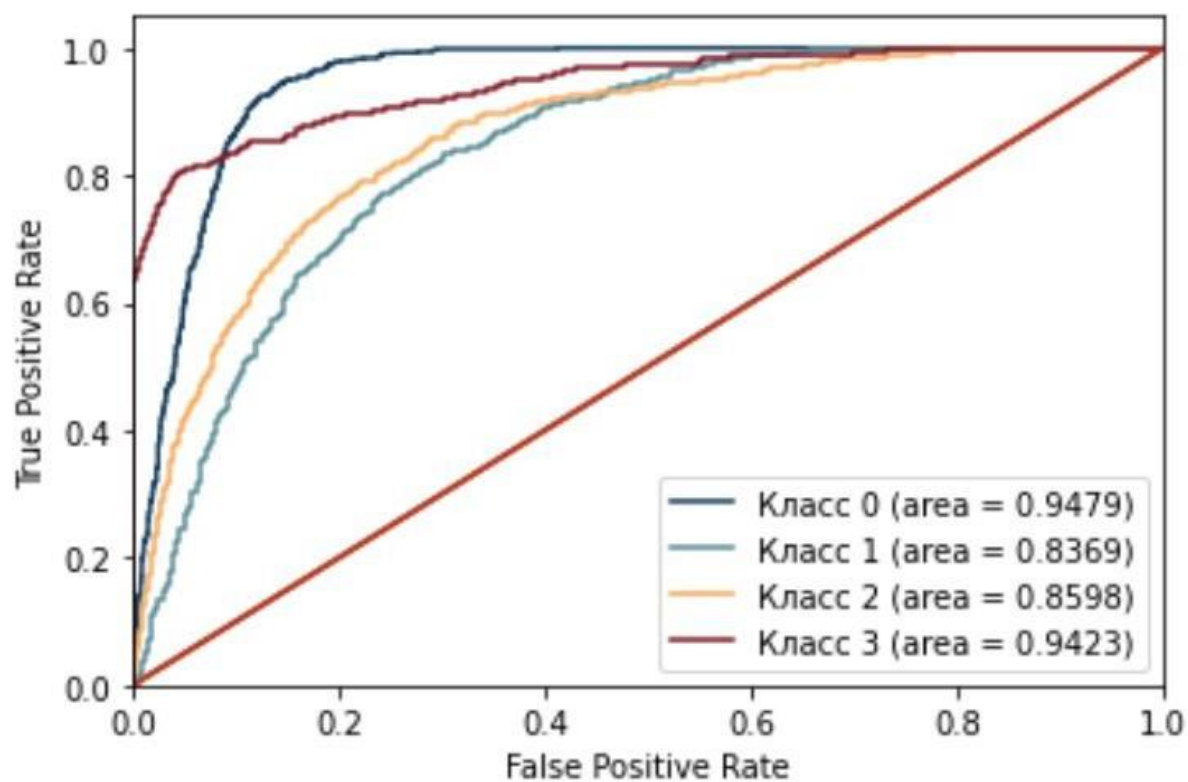
Графики ROC-кривой отдельного для каждого класса:







ROC-кривые для каждого класса на одном графике:



Регрессор

Набор данных:

Для работы с регрессорами использовался набор данных DS_2019_public. Оценка расходов домохозяйства на один из видов потребляемых ресурсов.

Датасет состоит из 10875 записей и 121 столбцов. Набор данных изначально не содержал пропущенных значений. Столбцы типа object состояли из числовых значений, поэтому мы перевели данные в них в тип float, а при невозможности такого преобразования удаляли строки с выбросами. В некоторых столбцах были выбросы в виде отрицательных значений. После удаления этих значений появились столбцы с NaN. Столбцы с процентом NaN больше 5% мы удалили, а в остальных столбцах удалили строки с пропущенными значениями. Дубликатов нет. После преобразований в наборе данных осталось 10785 записей и 113 столбцов.

Для анализа и обучения бинарного классификатора были использованы все признаки из набора данных, кроме целевого (TOTALDOL) и тех, в которых процент пропущенных значений был больше 5% (CELLAR, NUMTHERM, CONCRETE, COOLTYPE, STORIES, LGT1EE, ROOFTYPE, DRYRFUEL).

Датасет был разделен на 3 части: тренировочную (70%), тестовую (20%) и валидационную (10%).

Нейронные сети:

Всего было создано 4 регрессора, которые отличались своими параметрами:

1. Base Model:

Слой:

1. 16 нейронов, функция активации - ReLU
2. 1 нейрон

Функция потерь: Mean Squared Error

Процедура оптимизации: Adam

Метрика: —

Число объектов в батче: 10

Количество эпох: 200

Модель закончила обучение на 200-ой эпохе, достигнув лучшего значения ассурасу на валидационной выборке на 200-ой эпохе.

2. Second Model:

Слой:

1. 16 нейронов, функция активации - ReLU

2. 16 нейронов, функция активации - Sigmoid
3. 16 нейронов, функция активации - SeLU
4. 16 нейронов, функция активации - Sigmoid
5. 16 нейронов, функция активации - ReLU
6. 1 нейрон

Функция потерь: Mean Squared Error

Процедура оптимизации: Adam

Метрика: –

Число объектов в батче: 10

Количество эпох: 200

Модель закончила обучение на 74-ой эпохе, достигнув лучшего значения ассигасы на валидационной выборке на 64-ой эпохе.

3. Third Model:

Слой:

1. 64 нейрона, функция активации – ReLU
2. 64 нейрона, функция активации – Sigmoid
3. 32 нейрона, функция активации – SeLU
4. 32 нейронов, функция активации – Sigmoid
5. 16 нейронов, функция активации – ReLU
6. 1 нейрон

Функция потерь: Mean Squared Error

Процедура оптимизации: Adam

Метрика: –

Число объектов в батче: 100

Количество эпох: 160

Модель закончила обучение на 160-ой эпохе, достигнув лучшего значения ассигасы на валидационной выборке на 160-ой эпохе.

4. Fourth Model:

Слой:

1. 128 нейронов, функция активации – ReLU
2. Dropout(0.1)
3. 64 нейрона, функция активации – Sigmoid
4. Dropout(0.1)
5. 32 нейрона, функция активации – SeLU
6. 32 нейрона, функция активации – Sigmoid
7. 16 нейронов, функция активации – ReLU

8. 1 нейрон

Функция потерь: Mean Squared Error

Процедура оптимизации: Adam

Метрика: —

Число объектов в батче: 100

Количество эпох: 200

Модель закончила обучение на 184-ой эпохе, достигнув лучшего значения ассигасы на валидационной выборке на 174-ой эпохе.

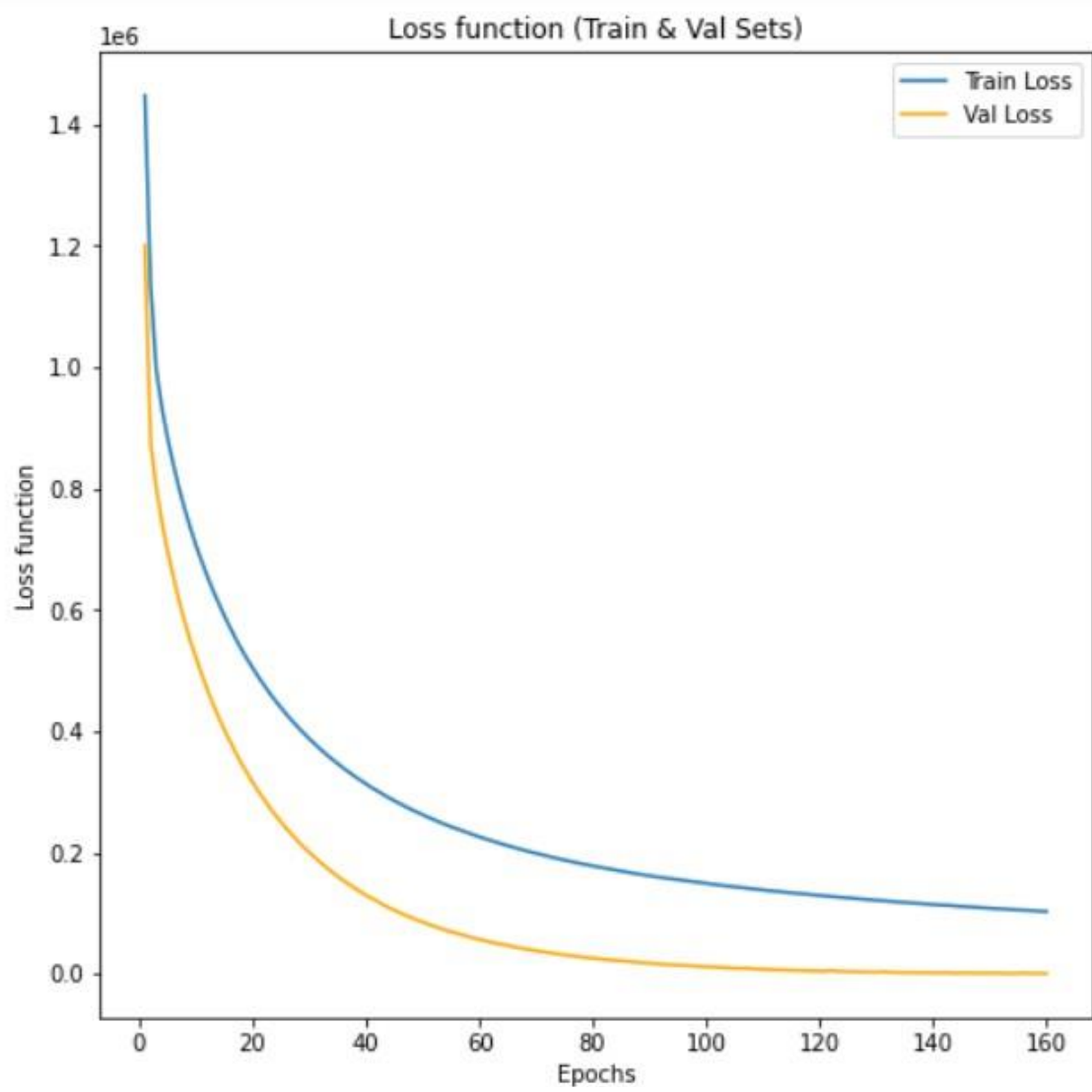
Метрики качества:

Ниже представлена таблица с метриками качества каждой модели. Оценка качества регрессоров происходила по следующим метрикам: MSE, MAE и R2.

	model	mse	mae	r2
0	base_model	6439.512	51.051	0.99453
1	second_model	1658.268	28.584	0.99859
2	third_model	1169.393	6.070	0.99900
3	fourth_model	2050.751	30.916	0.99823

Метрики качества у Third Model выше, чем у остальных моделей.

Графики обучения для лучшего регрессора:



Программный код

Программный код доступен по ссылке:

<https://github.com/pkvizavi/MachineLearning>