

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

ЛАБОРАТОРНАЯ РАБОТА №6
по дисциплине «Искусственные нейронные сети»
Тема: «Прогноз успеха фильмов по обзорам»

Студентка гр. 8383

Ишанина Л.Н.

Преподаватель

Жангиров Т.Р.

Санкт-Петербург

2021

Цели.

Прогноз успеха фильмов по обзорам (Predict Sentiment From Movie Reviews)

Задачи.

- Ознакомиться с задачей регрессии
- Изучить способы представления текста для передачи в ИНС
- Достигнуть точность прогноза не менее 95%

Ход работы.

Были импортированы необходимые зависимости для предварительной обработки данных и построения модели. Листинг приведен ниже.

```
import matplotlib.pyplot as plt
import numpy as np
from keras import layers
from keras.datasets import imdb
from keras import Sequential
```

Далее был загружен датасет IMDb, который уже встроен в Keras. Поскольку мы не хотим иметь данные обучения и тестирования в пропорции 50/50, мы сразу же объединим эти данные после загрузки для последующего разделения в пропорции 80/20. Листинг представлен ниже:

```
from keras.datasets import imdb
(training_data, training_targets), (testing_data,
testing_targets) = imdb.load_data(num_words=10000)
data = np.concatenate((training_data, testing_data), axis=0)
targets = np.concatenate((training_targets, testing_targets),
axis=0)
```

Затем была построена ИНС следующей архитектуры:

```
model.add(layers.Dense(50, activation="relu", input_shape=dmns,
))
model.add(layers.Dropout(0.3, noise_shape=None, seed=None))
model.add(layers.Dense(50, activation="relu"))
model.add(layers.Dropout(0.2, noise_shape=None, seed=None))
model.add(layers.Dense(50, activation="relu"))
model.add(layers.Dense(1, activation="sigmoid"))
```

Результаты работы

1) Было выбрано число входных параметров нейронной сети: 500.

Результаты запуска представлены на графиках точности и ошибки на рис.1-2.

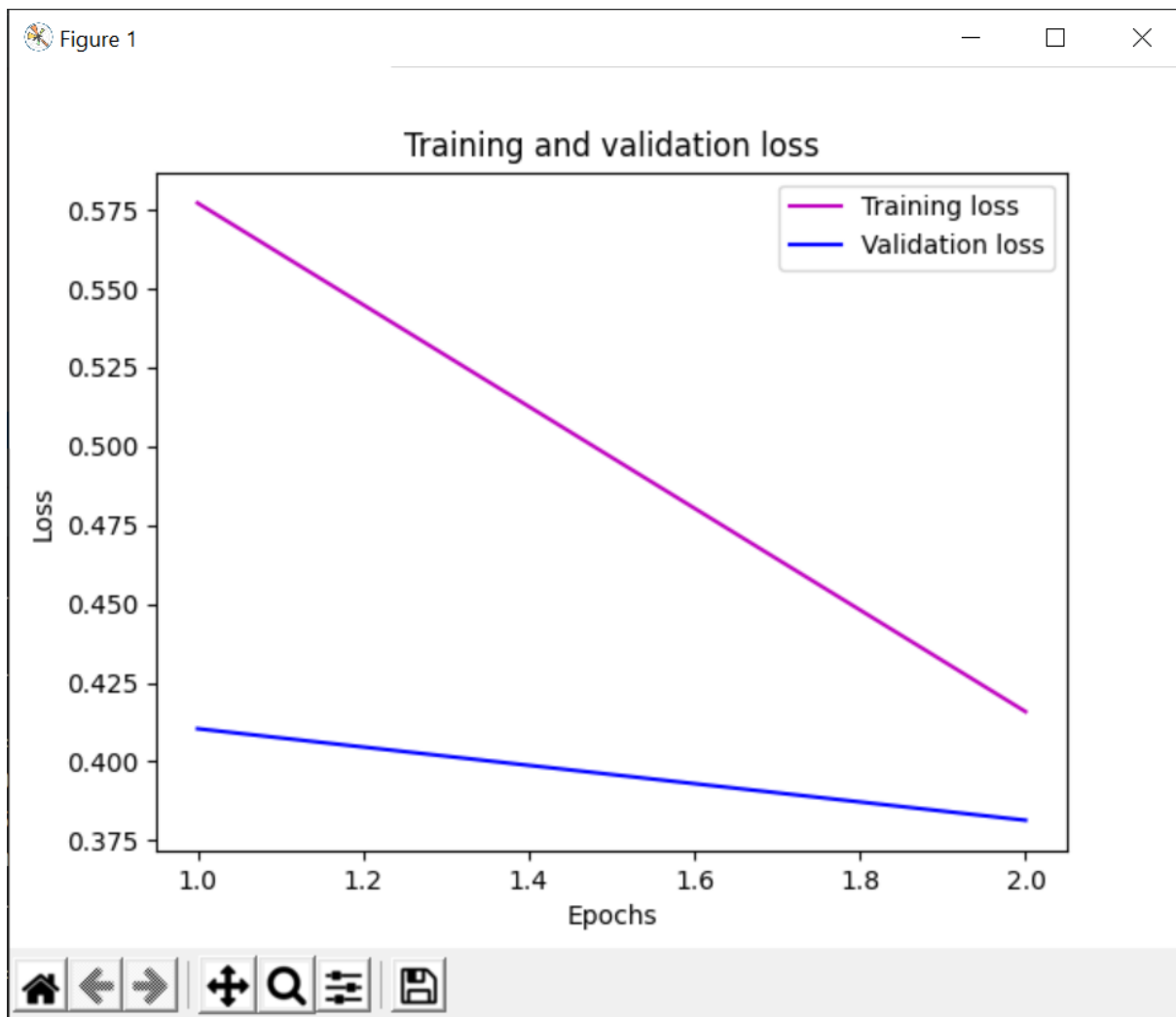


Рисунок 1 – графики потерь

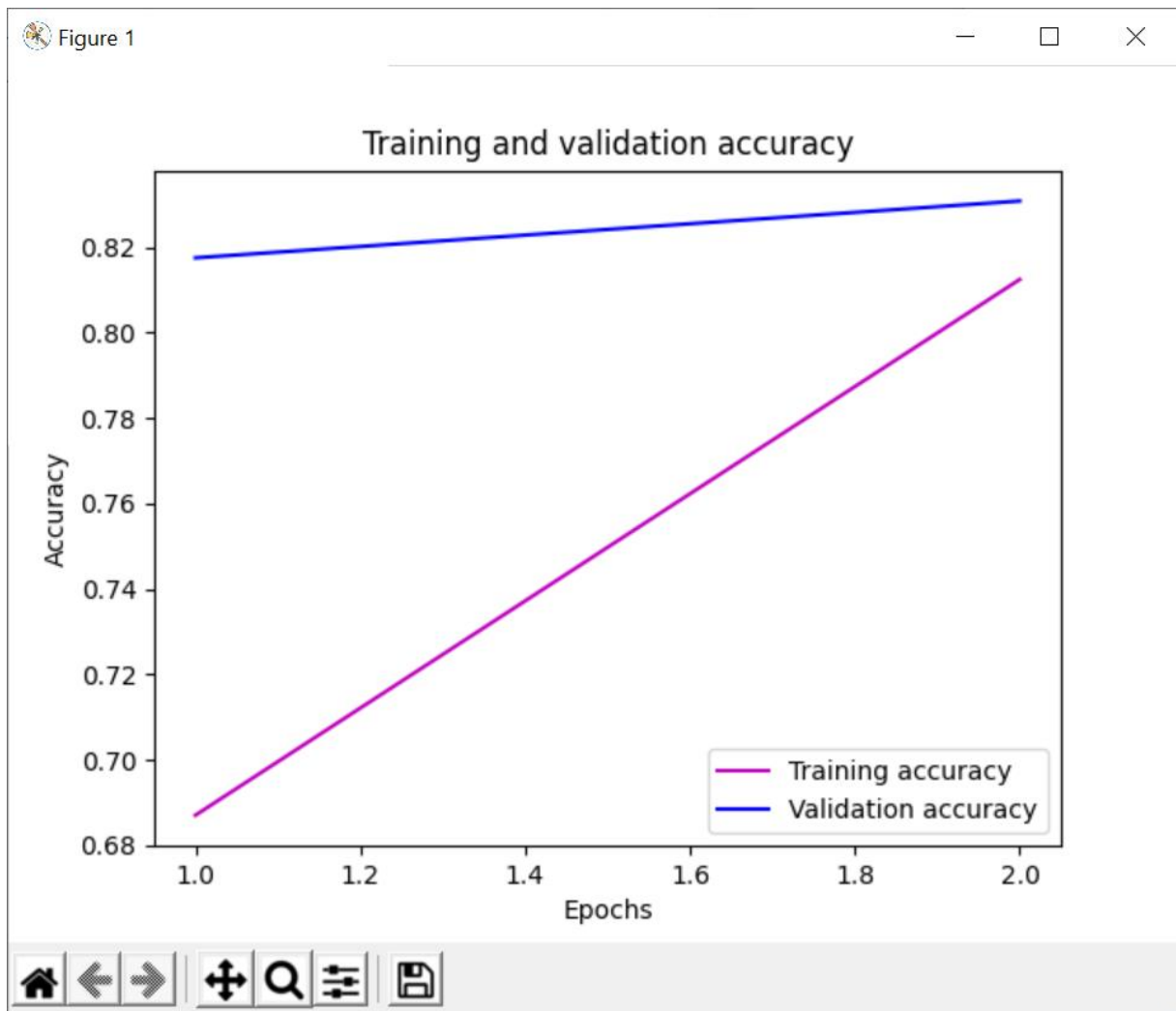


Рисунок 2 – графики точности

Исходя из графиков видно, что явного переобучения не наблюдается.

Значения потери и точности на последней эпохе:

loss: 0.3814 - accuracy: 0.8308

2) Далее было выбрано число входных параметров нейронной сети: 2500

Графики точности и потерь представлены на рис. 3-4.



Рисунок 3 – графики потерь

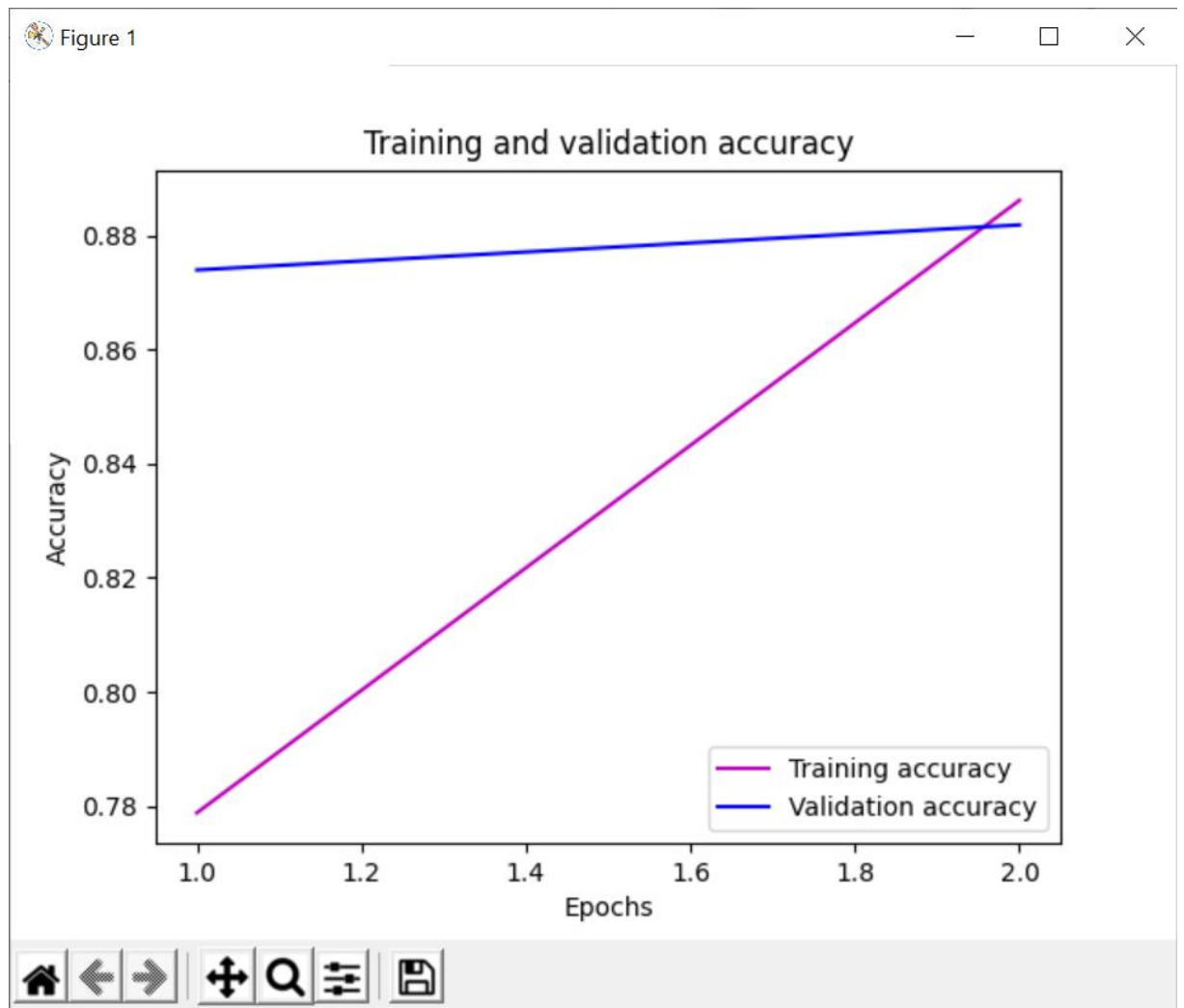


Рисунок 4 – графики точности

Исходя из графиков видно, что явного переобучения не наблюдается.

Значения потери и точности на последней эпохе:

loss: 0.2798 - accuracy: 0.8819

Значение точности значительно увеличилось, а значение потерь уменьшилось.

3) Далее было выбрано число входных параметров нейронной сети: 5000

Графики точности и потерь представлены на рис. 5-6.

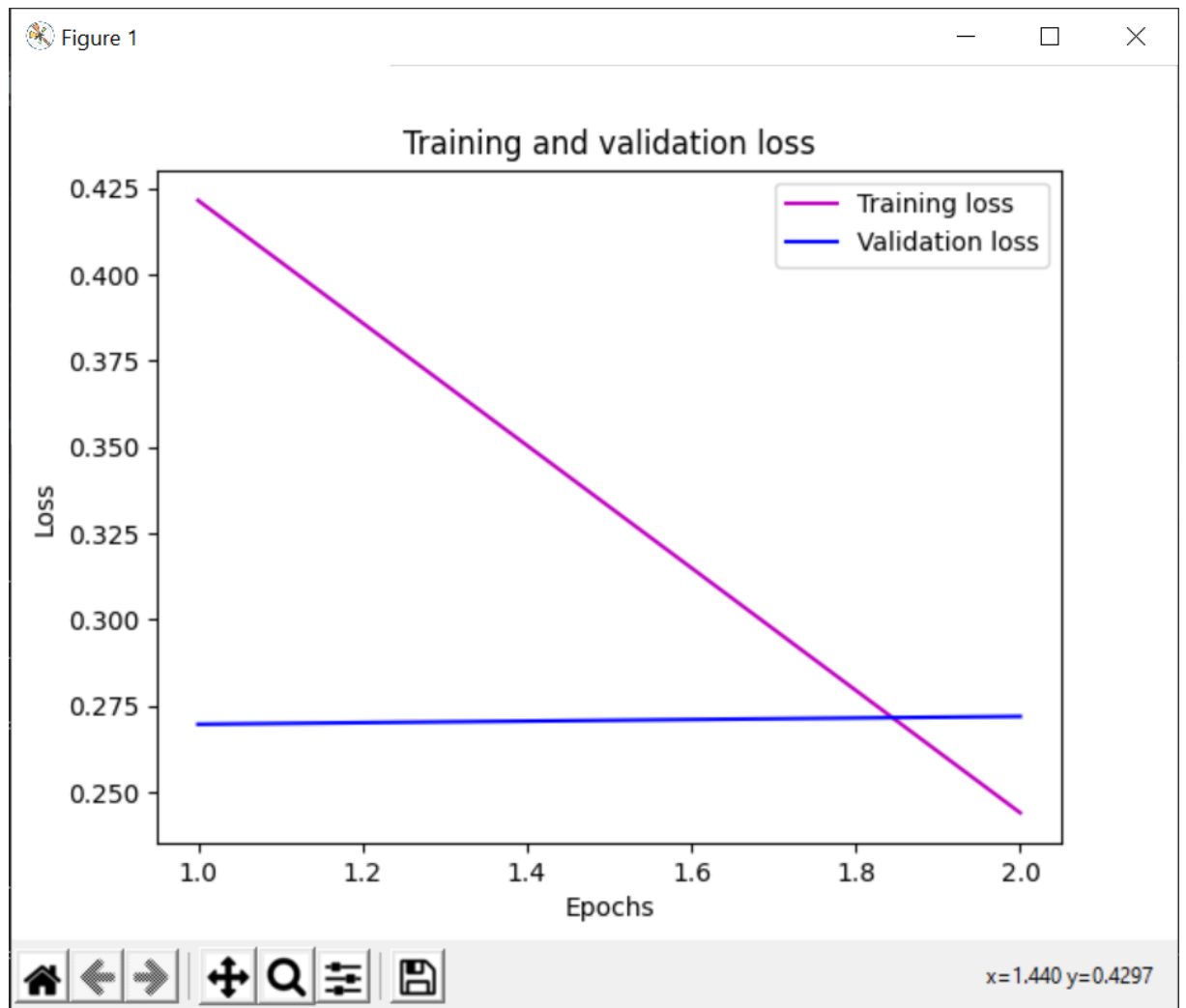


Рисунок 5 – графики потерь

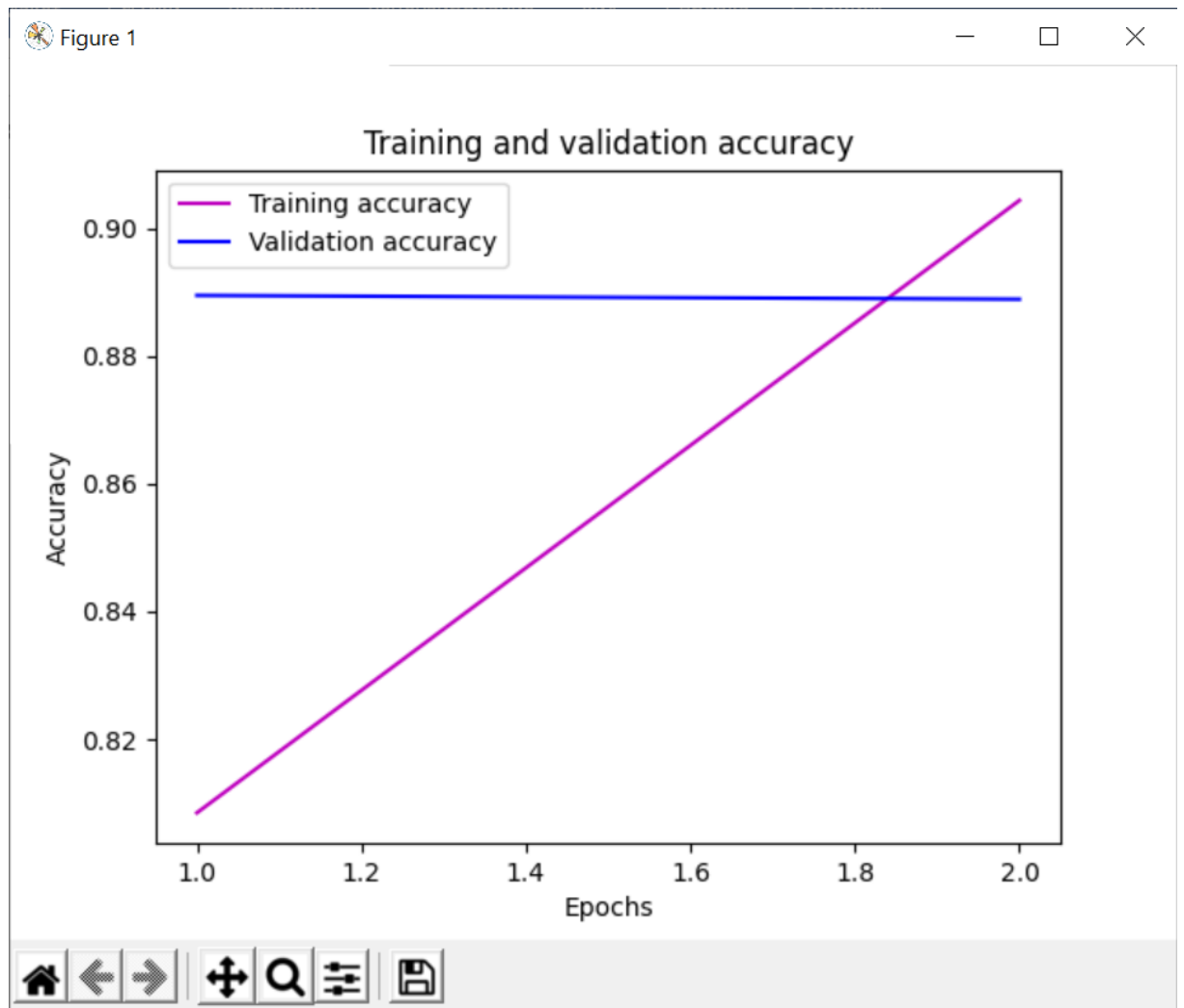


Рисунок 6 – графики точности

Исходя из графиков видно, что происходит незначительное переобучение. Значения потери и точности на последней эпохе:

loss: 0.2719 - accuracy: 0.8889

Значение точности немного увеличилось, а значение потерь немного уменьшилось.

4) Далее было выбрано число входных параметров нейронной сети:
10000

Графики точности и потерь представлены на рис. 7-8.

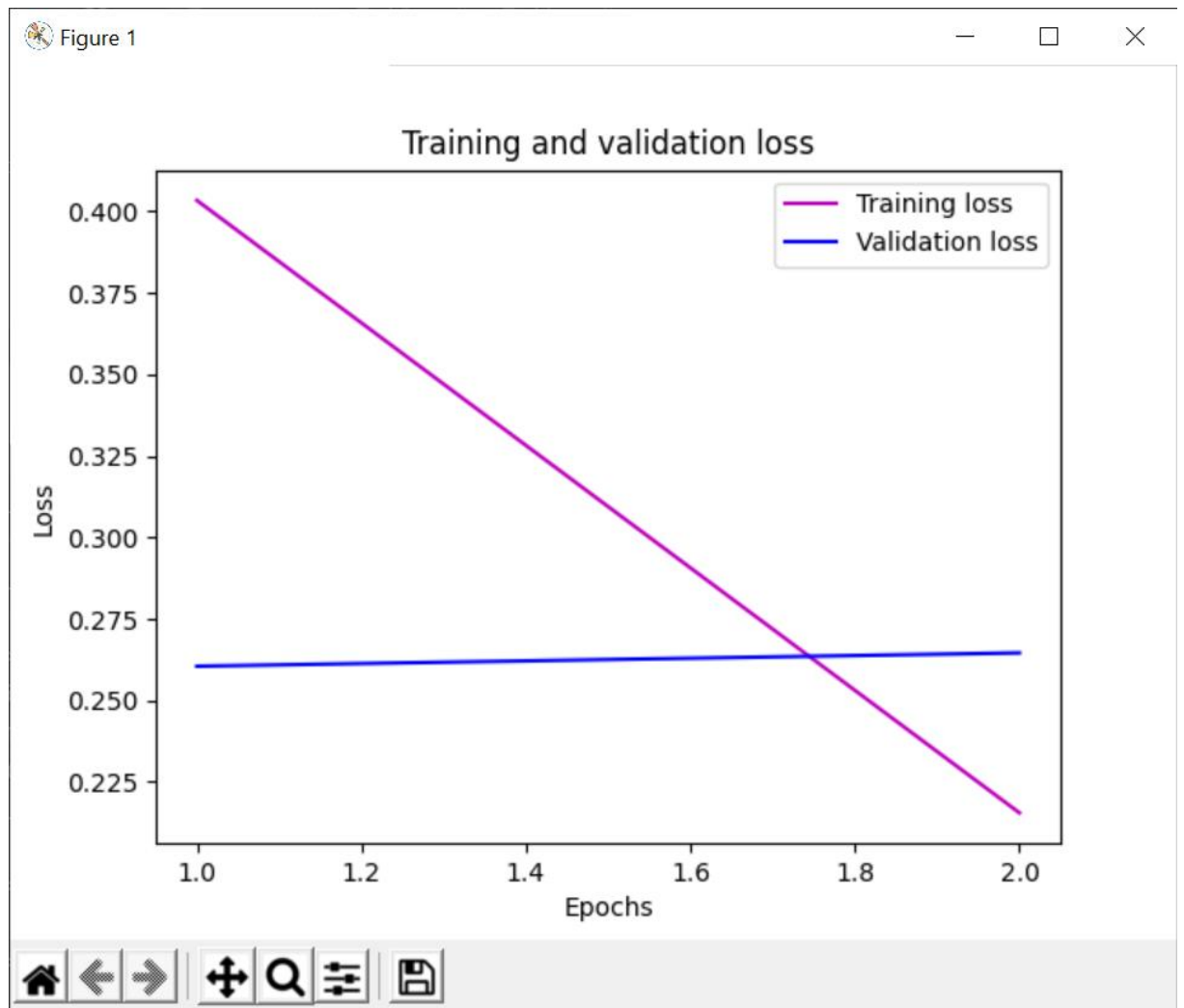


Рисунок 7 – графики потери

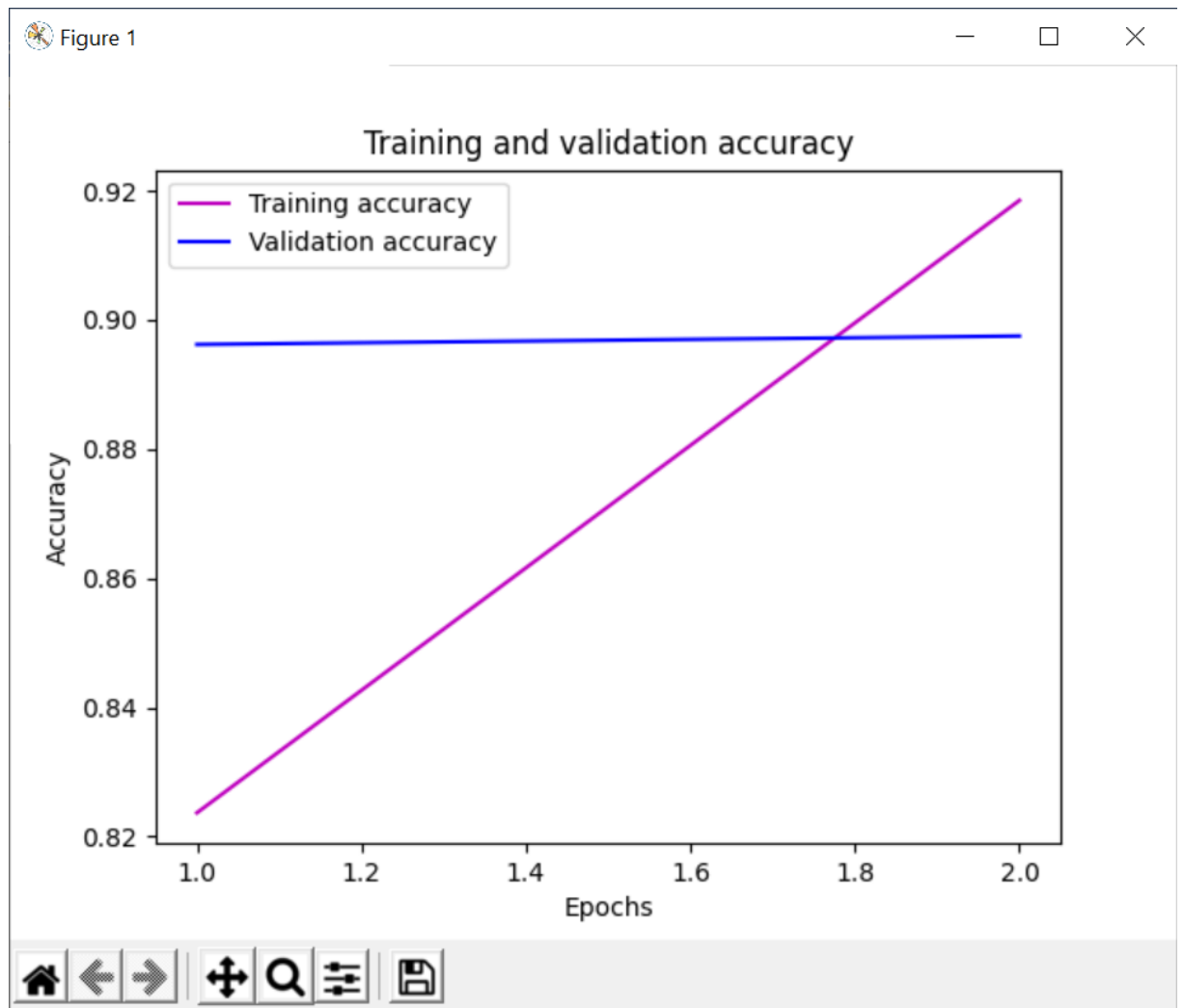


Рисунок 8 – графики точности

Исходя из графиков видно, что происходит незначительное переобучение. Значения потери и точности на последней эпохе:
loss: 0.2646 - accuracy: 0.8975

Значение точности улучшилось по сравнению со всеми предыдущими результатами.

Таблица 1 – Точности

Число параметров	500	2500	5000	10000
Точность	0.8308	0.8819	0.8889	0.8975

Таким образом, увеличение числа параметров приводят к увеличению точности. Но так же при большом числе параметров появляется переобучение после первой эпохе.

Далее была написана функция `test()`, в которой оценивается тестовый отзыв.

```
feedback = ["From the very first shots, we are shown idiotic  
graphics, which will be inserted in places throughout the film. It is  
so disgusting that your disgust will not end, throughout the entire  
timekeeping. Bottom line: worst movie I've seen, terrible acting, bad  
script, and disgusting humor."]
```

Результатом работы сети было число 0.30013126, что соответствует содержанию отзыва и характеризует его больше как отрицательный.

Вывод.

В ходе выполнения лабораторной работы было проведено ознакомление с задачей регрессии, изучена зависимость точности нейронной сети в зависимости от числа параметров. Таким образом, при увеличении размера числа параметров точность увеличивается.