

## **Assignment #5: Deep Learning COP4630**

Jonathan Masino(Architect), Ava Perrone(Developer), Steven Fernandez(Reporter)

### **Overview**

We formed our group days after A1 was released, we have taken some classes in the past so it was easy for us to assign roles based on that. We read the book and watch the different videos professor Marques has posted on YouTube where he explains each chapter and the code related to them. Our professor videos showing different hints for A5 were very helpful and gave us an idea on how to start this assignment.

After watching the video explanation for this assignment we started working with was given like the starting code and multiple hints in the video. It took us various attempts to find the proper solution for this assignment but we succeeded. These are some screenshots showing the final result of our A5.

### **Documentation of the solution**

#### **PART 1:**

(1) What type of preprocessing was applied to the raw input data? Why was it necessary to do so?

Ans: The integer sequence was encoded via multi-hot encoding. This was necessary because it sets up the data for the training set.

(2) Does your model suffer from overfitting? If so, what would you suggest doing about it?

Ans: Yes, the model suffered from overfitting because validation loss is increasing while training loss is decreasing. As a team we concluded that the best suggestion to correct the overfitting in this case is to correct some of the code.

(3) Is accuracy a good metric of success in this case? Why (not)?

Ans: Yes, as you can see from the graph, the validation accuracy line was quite accurate. This is good because it shows that the data was validated correctly.

#### **PART 2:**

(1) What type of preprocessing was applied to the raw input data? Why was it necessary to do so?

Ans: The data and the labels were encoded using the vectorize sequence. This was necessary to prepare the data for the training set.

(2) How many categories are there in this case?

Ans: There are 2 categories in this case loss and accuracy.

(3) Does your model suffer from overfitting? If so, what would you suggest doing about it?

Ans: Yes, the model suffered from overfitting because validation loss is decreasing at the beginning but after some time it pretty much stays stable while training loss is still decreasing. As a team we concluded that the best suggestion to correct the overfitting in this case is to correct some of the code.

(4) Is accuracy a good metric of success in this case? Why (not)?

Ans: Yes, as you can see from the graph, the validation accuracy line was quite accurate. This is good because it shows that the data was validated correctly.

### PART 3:

(1) What type of preprocessing was applied to the raw input data? Why was it necessary to do so?

Ans: The data was normalized. This was necessary for the test set.

(2) Why is this problem a case of regression (rather than classification)?

Ans: Because the main objective of regression is to predict a continuous value instead of a discrete label like in the case of classification.

(3) Does your model suffer from overfitting? If so, what would you suggest doing about it?

Ans: No, looking at the graphs we can see that overfitting is almost nonexistent.

(4) Is mean absolute error (MAE) a good metric of success in this case? Why (not)?

Ans: Yes, because MAE is the absolute value of the difference between the predictions and the targets. The approximate MAE score in this case of 2.6 would mean this prediction is off by \$2.600 on average.

### References

**François, C. (2017). *Deep Learning with Python*. Manning.**