

Artificial Intelligence II - Homework 4

Ioanna Oikonomou

March 2023

Contents

1	Introduction	1
2	First part	2
2.1	Assignment	2
2.1.1	Data preprocessing	2
2.1.2	The model	2
2.1.3	Review Dataset	2
2.2	Hyperparameter tuning	2
2.2.1	Learning rate	2
2.2.2	Optimizer	2
2.2.3	Number of warmup steps	2
2.2.4	Batch size	2
2.2.5	Dropout probability	3
2.2.6	Loss function	3
2.3	Results	3
2.3.1	Learning curves	3
2.3.2	F1 score, recall and precision	4
2.4	Testing	4
2.5	Sources	5

1 Introduction

In this report I'm going to explain my solution regarding the forth project of the course. The firts part of this project requires developing a sentiment classifier for movie reviews using the pretrained BERT-base model available on Hugging Face.

2 First part

2.1 Assignment

2.1.1 Data preprocessing

The data preprocessing is exactly the same as in all the previous projects.

2.1.2 The model

My model is implemented in the class named Classifier and it consists of 3 layers. The first layer is the pretrained BERT model, then we have a linear layer and, lastly, a softmax layer.

2.1.3 Review Dataset

This class was implemented so as to provide the data to the model. I have implemented 2 additional functions (len and getitem) that I understood were necessary for the model to be provided with the data. At first my implementation didn't contain this class, but I decided to include it for cleaner code.

2.2 Hyperparameter tuning

2.2.1 Learning rate

The best learning rate based on my experiments was $1e-5$. I also tried $2e-5$ and $3e-5$ but they didn't seem to bring as good results as the value I finally picked.

2.2.2 Optimizer

I used two different optimizers in my experiments. Adam and AdamW. The latter seemed to bring better results so I included that one in my final solution.

2.2.3 Number of warmup steps

The truth is I didn't experiment with the value of the number of warmup steps because it was taking too much time. As a result, I considered the number of warmup steps to be 0.

2.2.4 Batch size

I tried two different values for the batch size. 16 and 64. Although the latter seemed to bring better results, it caused the cuda memory to run out really quickly which was not at all useful for me experimenting a lot. That's why I preferred to leave the batch size at 16.

2.2.5 Dropout probability

At first I had inserted one more dropout layer to my model. However, I noticed that it caused a slight overfitting so I removed it and the results were better in terms of fitting.

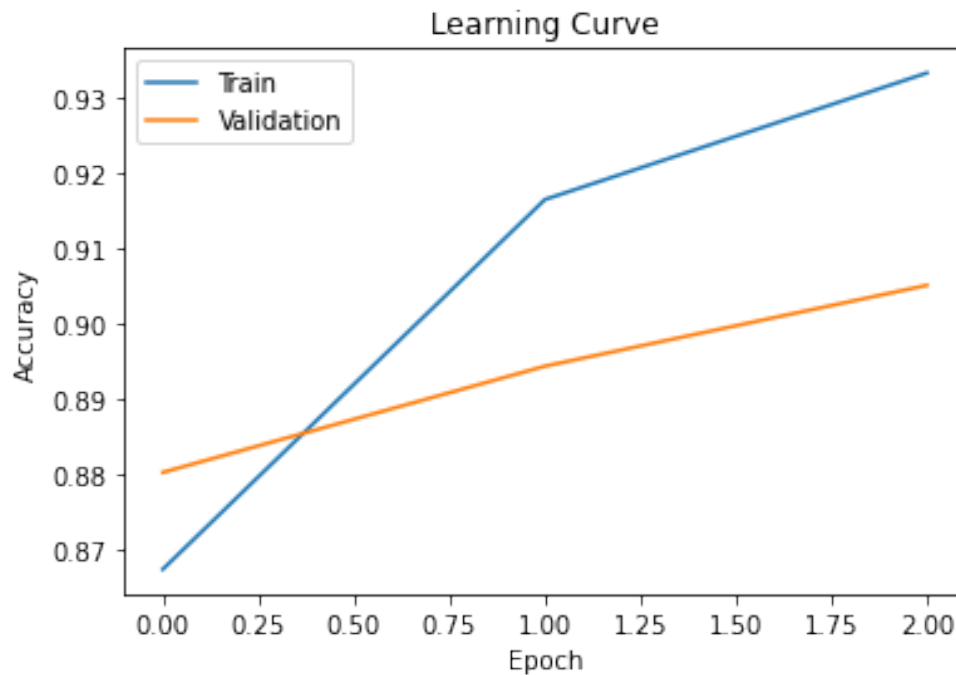
2.2.6 Loss function

I used Cross Entropy loss function because of the Softmax layer in my model.

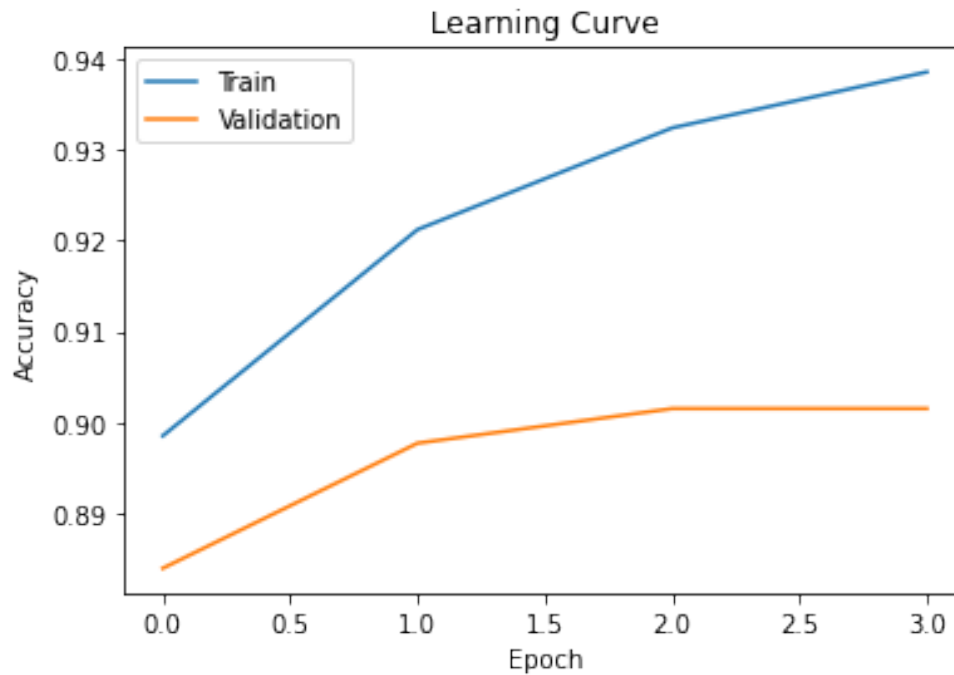
2.3 Results

2.3.1 Learning curves

Below, I provide you with a learning curve of my model over 3 epochs:



I will also provide you with a learning curve based on accuracy over 4 epochs, stating the fact that it was developed before removing the dropout layer of my model.



2.3.2 F1 score, recall and precision

Here, I'm using a screen shot of the result of the classification report of my model. In this photo we can see the values of f1 score, recall and precision which were requested in terms of the assignment.

	precision	recall	f1-score	support
0	0.91	0.90	0.90	4470
1	0.90	0.91	0.91	4463
accuracy			0.91	8933
macro avg	0.91	0.91	0.91	8933
weighted avg	0.91	0.91	0.91	8933

2.4 Testing

The last part of my code is a section you can use to test with the test set. I'm not sure this is what you needed in order to test it. I hope it helps :)

2.5 Sources

- [Hugging Face](#)
- Examples that were distributed in the lectures regarding this assignment.