Deep Learning FSS22
Assignment 2: CNNs and RNNs

Timur Michael Carstensen - 1722194

08.05.2022

Contents

1 Convolutional Neural Networks

1.1 a)

(i) *3x3 convolutional layer with bias, 1 input channel, 32 output channels, stride 1, padding 1 on each side*

Input: 28x28x1

Output: 28x28x32

No. of Parameters: 32 3x3 kernel matrices and 32 bias terms

(ii) *Logistic activation function*

Input: 28x28x32

Output: 28x28x32

No. of Parameters: none

(iii) *2x2 max-pooling layer with stride 2*

Input: 28x28x32

Output: 14x14x32

No. of Parameters: none

(iv) Linear layer with 10 hidden units

Input: 14x14x32

Output: 10

No. of Parameters: 10x6270=62720

(v) Log-softmax function

Input: 10

Output: 10

No. of Parameters: none

Interpretation of the outputs: log-probabilities for the 10 different classes of garments.

1.2 b)

Cf. Code.

1.3 c)

Cf. code.

1.4 d)

Cf. Code.

1.5 e)

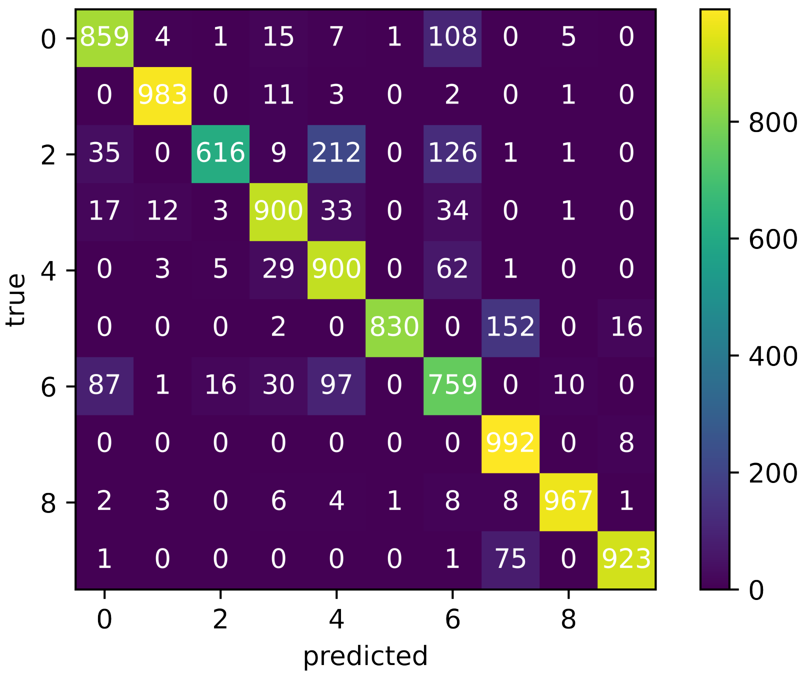

Figure 1: Confusion matrix for the fashion MNIST dataset

The model achieved an accuracy of 87.29%. When inspecting the individual classes, we can see that the accuracy varies from 99.2% to as low as 61.6%. The model performs the worst on pullovers and t-shirts with accuracies of 61.6% and 75.9% respectively. It performs best on sneakers and ankle boots with accuracies of 99.2% and 98.3% respectively. The rest of the classes are in the 80% to 90% range, with bags also performing particularly well (accuracy: 96.7%). I would expect the model to perform well on unique shapes and make mistakes on garments that are quite similar in shape. That is, shirts and pullovers are easy to confuse since they have similar shapes and sneakers, ankle boots and bags have (very) distinct shapes as compared with the rest of the garments. This is confirmed in Figure 1 where we can see that pullovers are often confused for shirts and coats. Our model can discriminate footwear from non-footwear quite well, though it does have issues distinguishing sandals from sneakers occasionally. These mistakes are reasonable for our model to make since it does not learn anything about the actual meaning of what makes a shirt a shirt or a pullover a pullover.

1.6 f)

Not attempted.

2 Recurrent Neural Networks and Pretraining

2.1 a)

Cf. Code.

2.2 b)

We can see that the error on our training set decreases with each subsequent epoch. That is, for our first batch the average training loss is 0.6865 and for our fifth epoch it is 0.3577. However, at the same time, the average loss on our validation set reaches a minimum at around epoch 4 and starts to slightly increase with epoch 5 (0.6657 vs. 0.6667). This indicates that we are overfitting our model on our training data and not converging to a *global* minimum. I suspect this to be because of our embedding layer which we train on our training data. For tokens not encountered in the training set, the corresponding embeddings retain their random initialisations and thus reduce our model's performance in the forward pass on the validation set. This effect may have been lessened if our vocabulary in the training set were larger or if we had more training examples in general.

2.3 c)

The GloVe [PSM14] embedding file contains word embeddings for 29841 of the 32362 unique word ids in our dataset. In contrast to before, the embeddings are 100-dimensional vectors containing the representations of our words. The ”word-embeddings.txt” file is organised in such a way that each row is a key, value pair with the key being the word id and the value being the embedding $\in \mathbb{R}^{100}$. This way, we can match the embeddings to our corpus with the word ids.

2.4 d)

The first observation that we can make is that the average validation losses of both the RNN with and without finetuning stay constantly below that of the simple RNN. The average training set loss stays above the simple RNN when do not finetune our embeddings and falls below it if we do. The average training loss for the RNN with GloVe embeddings and fine-tuning also decreases much faster than that of the other models. When inspecting the validation loss in the last epoch of the GloVe model with fine-tuning we can see that we are again on an upward trajectory toward the end of the epoch. Hence, we are starting to overfit. The model with GloVe embeddings and without fine-tuning seems to stabilise and converges at an average validation loss of 0.6. We can conjecture that this is the best loss that we can achieve given our embeddings and model design. Better generalisation performance could possibly be achieved by using a larger training set, initialising more (or all) of the word vectors with GloVe embeddings and fine-tuning the model up until we observe an up-tick in validation loss and stop training at that point.

2.5 e)

References

[PSM14] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.