CSc 84020 Neural Networks and Deep Learning
Homework 2: Classification and Prediction with DL Multi-Layer Perceptron
Andrea Ceres
Shao Liu

3c. Analysis

The model we use in this section is similar as the one we use in *3b*. The main architecture is a five-layer perceptron neural network, with batch normalization after the first and fourth dense layer and with dropout after the second dense layer. The first four layers have 32 neurons per layer and the last layer has 16 neurons. We replaced the optimizer to *Adam*, setting the learning rate to 0.001. The activation function we chose for each layer is *LeakyReLU*, which is a modification of *ReLU*. The difference is that for negative input *x*, instead of outputting 0 as *ReLU* does, *LeakyReLU* will output *k*x*, where *k* is a small constant value less than one (0.1). We conducted several experiments with different parameters of dropout and regularization. We tested the model without dropout and regularization, dropout rates of 0.1, 0.2 and 0.4, and L2 regularization rates of 0.0005 and 0.01. The results are shown in *Table.1*.

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Dropout	0.0	0.1	0.2	0.4	0.0	0.1	0.2	0.4	0.0	0.1	0.2	0.4
L2 Regularization	0.0	0.0	0.0	0.0	0.0005	0.0005	0.0005	0.0005	0.01	0.01	0.01	0.01
Accuracy	**0.7641**	0.7571	0.7538	0.7375	**0.7613**	0.7588	0.7559	0.7245	**0.7371**	0.7078	0.7220	0.7068
Loss	0.6984	**0.6967**	0.7123	0.7637	**0.7440**	0.7538	0.7577	0.8783	**0.8738**	0.9320	0.9059	0.9366

Table.1

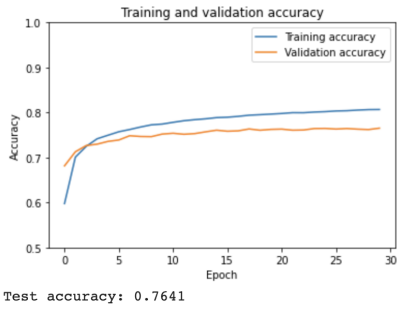

Fig.1

From *Table.1*, we can see that no dropout and regularization have achieved the highest accuracy. With the same regularization rate, the accuracy decreases as the dropout rate increases. Moreover, the loss increases as well when the dropout rate increases. The dropout may have had a negative effect on the testing accuracy because the parameters in our model are relatively few. We only have five layers and the total parameter count is 29,174. Moveover, we only trained 30 epochs. We can surmise our model is still underfitting. Our model in its current form is not complex enough to accurately capture relationships between the training dataset's features and the target labels.

From *Fig.1* and *Fig.2*, we can see that during the training process of the no dropout and no regularization model, the accuracy keeps increasing and loss keeps decreasing. We can conclude that if we train the model for more epochs, the loss will be improved and the model will become more accurate. This means the parameters still need more training to capture the relationship between training data and labels. So 30 epoch is not enough for the model to be overfitting to the training data.

The other reason that dropout does not work for our model is that our model is not complex enough to capture the relationship between the training dataset and labels. The training process of dropout (0.4) and L2 regularization (0.01) show that the accuracy fluctuates around the training accuracy, as shown in *Fig.3*.

The differences are caused by the dropout. Because our model is not complex enough, dropouts will strengthen the weakness of our model. Then, it will be hard to predict on testing data. This is why the accuracy of the testing set cannot have a stable improvement.

Fig.2

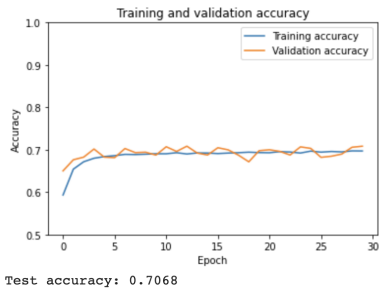

Fig.3

To improve the performance of our model, we think the main methods are: (1) to increase the complexity of our model such as adding more layers and (2) to train it for more epochs.