COMP 4211 - Machine Learning Programming Assignment 2 Report

Cheng Chi Fung
cfchengac@connect.ust.hk

1 CNN Classifier

1.1 Screen Shots of the CNN Classifier

The following are the screen shot of the CNN Network and loading the pretrained encoder.

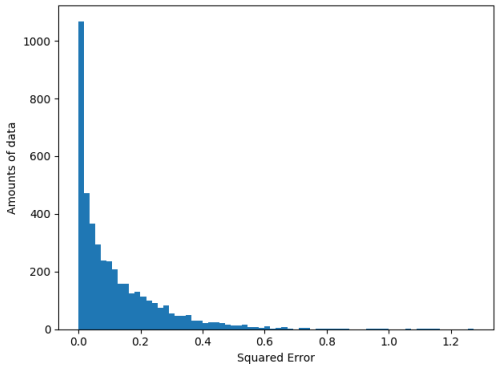

Figure 1: Screen shot of the CNN Network

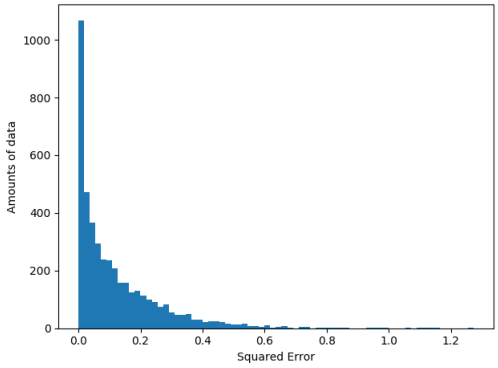

Figure 2: Screen Loading the Pretrained Encoder

1.2 Hold out validation result of CNN classifier from scratch

The following are the hold out validation results of CNN classifier from scratch. The network archiecture of using in the hold out validation was the same as the snapshot above. In the validation, we partitioned the training set into the training and validation sets with ratio 4:1. For each candidate set of hyperparameters, we trained the network for 20 epochs batch size 32 and validate it against the validation set. The cross entropy loss shown in the results table was the optimal one .

Table 1: Hold out Validation Results of scratch CNN classifer

Parameters set	Optimizer	Learning Rate	Num of Hidden	Cross Entropy Loss
1	Adam	0.001	32	awd
2	SGD	0.1	32	wd
3	SGD	0.01	32	wd
4	Adam	0.001	64	wd
5	SGD	0.1	64	wd
6	SGD	0.01	64	dw

The cross entropy loss of parameters set x had the lowest optimal cross entropy loss after conducting the hold out validation, so we would choose the parameters set [Adam, 0.001, 32] as the hyperparameters for the training in the testing phase.

1.3 Hold out validation result of the CNN classifier with Pretrained Encoder Weights

The following are the hold out validation results of the CNN classifier with pretrained encoder weights. The network archiecture of using in the hold out validation was the same as the snapshot above. In the validation, we partitioned the training set into the training and validation sets with ratio 4:1. For each candidate set of hyperparameters, we trained the network for 20 epochs batch size 32 and validate it against the validation set. The cross entropy loss shown in the results table was the optimal one .

Table 2: Hold out Validation Results of scratch CNN classifer

Parameters set	Optimizer	Learning Rate	Num of Hidden	Cross Entropy Loss
1	Adam	0.001	32	awd
2	SGD	0.1	32	wd
3	SGD	0.01	32	wd
4	Adam	0.001	64	wd
5	SGD	0.1	64	wd
6	SGD	0.01	64	dw

The cross entropy loss of parameters set x had the lowest optimal cross entropy loss after conducting the hold out validation, so we would choose the parameters set [Adam, 0.001, 32] as the hyperparameters for the training in the testing phase.

1.4 Testing Results of CNN Classifier

In the testing phase, for both CNN classifier from scratch and with Pretrained Encoder Weights, we trained using the entire training set with the best set of hyperparameters obtained from the hold out validation and the same network architecture as the hold out validation, and tested with the testing test. We used 32 as our batch size and trained with 20 epoch. The following was the results obtained from the testing phase.

Table 3: Testing Result Metric of CNN classifier from scratch

	Cross Entropy Loss	Top-1 Accuracy	Top-3 Accuracy
Mean	SGD	0.1	32
Std	SGD	0.1	32

Table 4: Testing Result Metric of CNN classifier with Pretrained Encoder Weights

	Cross Entropy Loss	Top-1 Accuracy	Top-3 Accuracy
Mean	SGD	0.1	32
Std	SGD	0.1	32

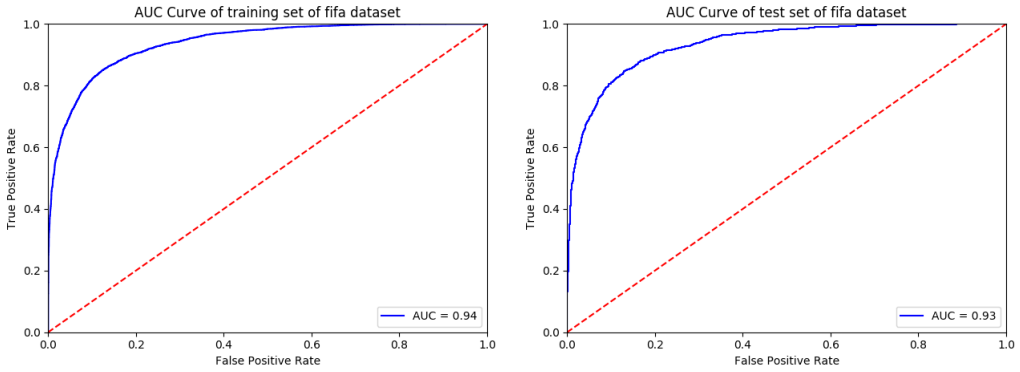

Figure 3: Learning Curve of Cross Entropy Loss

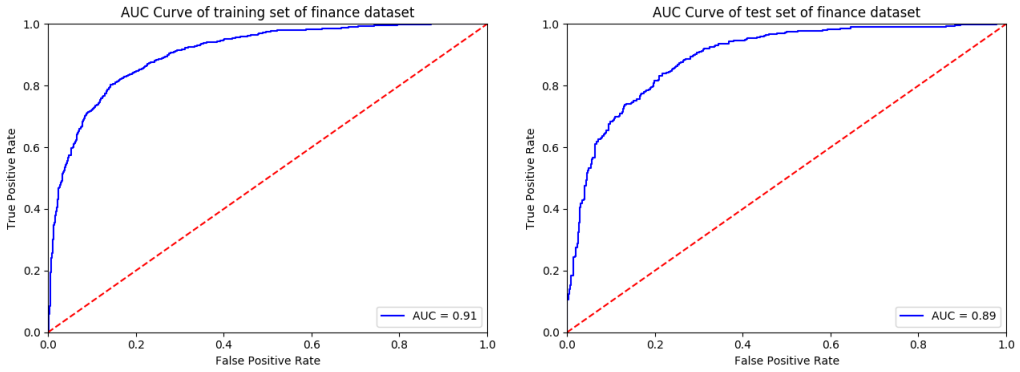

Figure 4: Learning Curve of Top-1 Accuracy

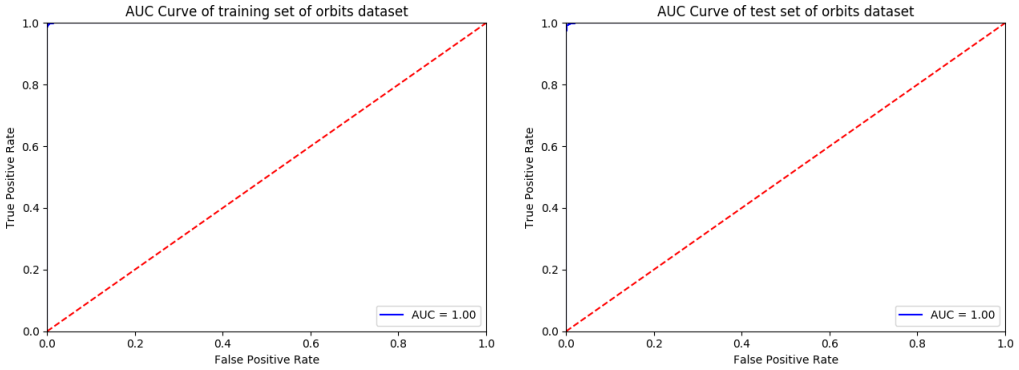

Figure 5: Learning Curve of Top-3 Accuracy

From the square loss histogram of test sets, we found out that the mean squared error of the most of the data points are around zero. It means that this model perform quite well for most of the data points. However, for the finance datasets, there are some data points that have the sqaured error far away from zero (23.77581) which means that there might be some **outliers** that far away from the curve.

2 CAE with Pretrained Encoder

2.1 Screen Shots of the CAE with Pretrained Encoder

The following are the screen shot of the Network of CAE Decoder and loading the pretrained encoder.

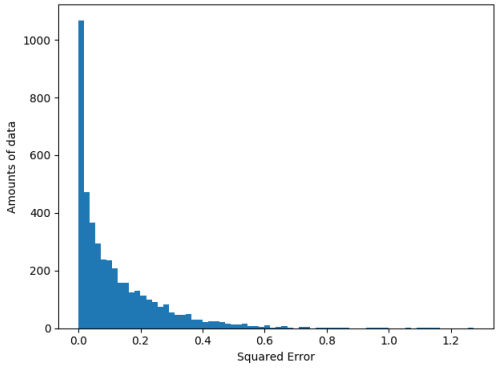

Figure 6: Screen shot of the Network of CAE Decoder

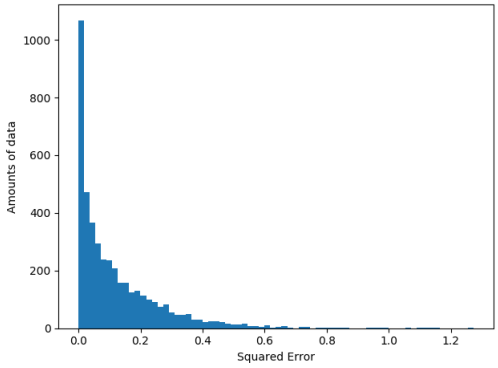

Figure 7: Screen Loading the Pretrained Encoder

2.2 Hold out validation result of CAE with Pretrained Encoder

The following are the hold out validation results of the CAE with Pretrained Encoder. The network archiecture of using in the hold out validation was the same as the snapshot above. In the validation, we partitioned the training set into the training and validation sets with ratio 4:1. For each candidate set of hyperparameters, we trained the network for 20 epochs batch size 32 and validate it against the validation set. The cross entropy loss shown in the results table was the optimal one .

Table 5: Hold out Validation Results of scratch CNN classifer

Parameters set	Optimizer	Learning Rate	Num of Hidden	Cross Entropy Loss
1	Adam	0.001	32	awd
2	SGD	0.1	32	wd
3	SGD	0.01	32	wd
4	Adam	0.001	64	wd
5	SGD	0.1	64	wd
6	SGD	0.01	64	dw

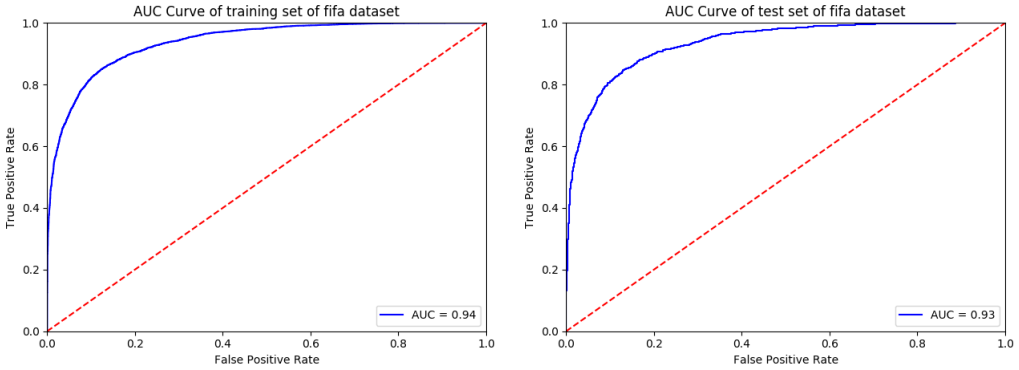

Figure 8: Learning Curve of Cross Entropy Loss

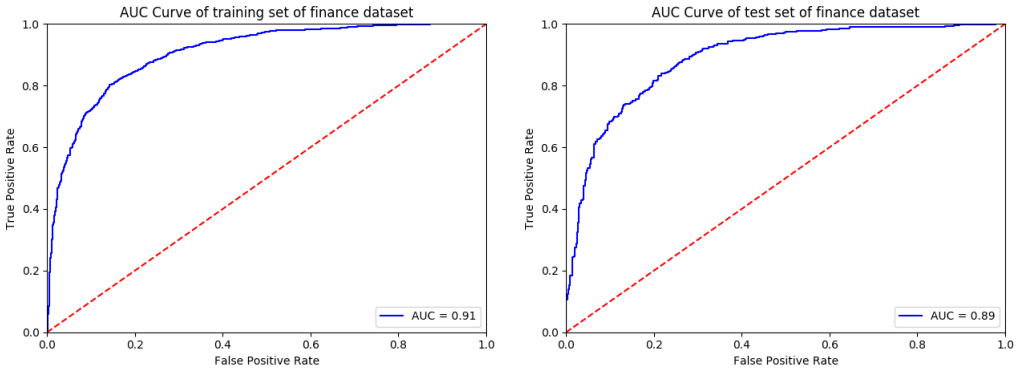

Figure 9: Learning Curve of Top-1 Accuracy

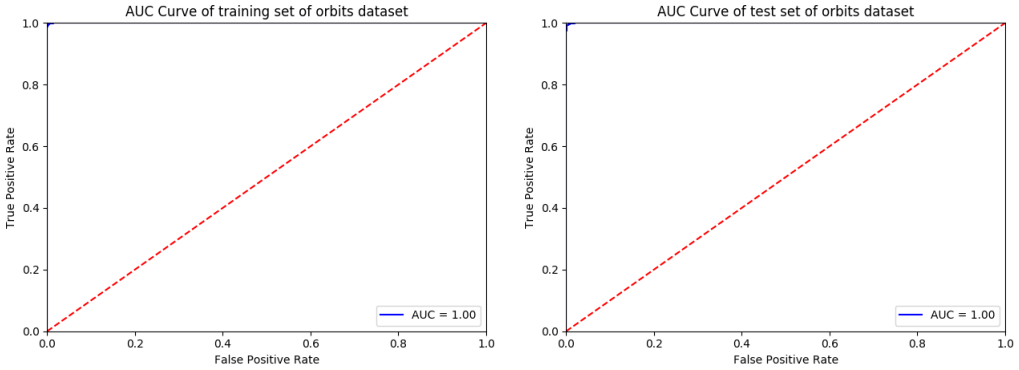

Figure 10: Learning Curve of Top-3 Accuracy

The cross entropy loss of parameters set x had the lowest optimal cross entropy loss after conducting the hold out validation, so we would choose the parameters set [Adam, 0.001, 32] as the hyperparameters for the training in the testing phase.

2.3 Testing Results of CAE with Pretrained Encoder

In the testing phase, for both CNN classifier from scratch and with Pretrained Encoder Weights, we trained using the entire training set with the best set of hyperparameters obtained from the hold out validation and the same network architecture as the hold out validation, and tested with the testing test. We used 32 as our batch size and trained with 20 epoch. The following was the results obtained from the testing phase.

Table 6: Testing Result Metric of CNN classifier from scratch

	Cross Entropy Loss	Top-1 Accuracy	Top-3 Accuracy
Mean	SGD	0.1	32
Std	SGD	0.1	32

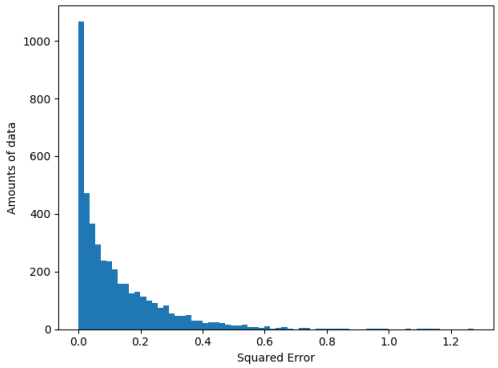

Figure 11: Screen shot of the CNN Network

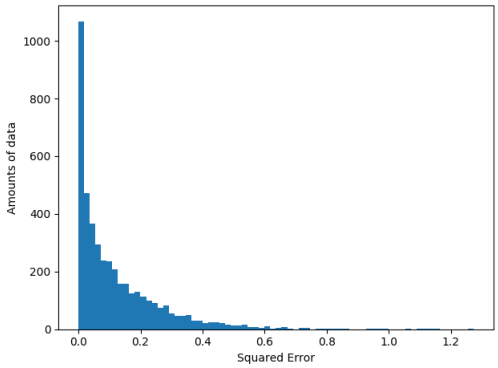

Figure 12: Screen shot of the CNN Network

From the square loss histogram of test sets, we found out that the mean squared error of the most of the data points are around zero. It means that this model perform quite well for most of the data points. However, for the finance datasets, there are some data points that have the sqaured error far away from zero (23.77581) which means that there might be some **outliers** that far away from the curve.