COMP4211 Assignment 1 Report

NG Chi Him
20420921
chngax@connect.ust.hk

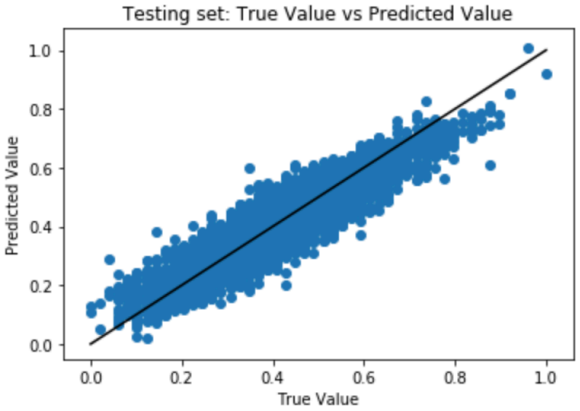
Computing Environment

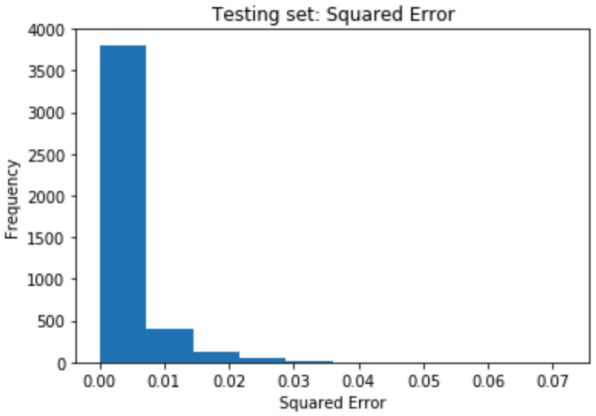
All tasks were implemented in Jupyter Notebooks in Anaconda environment with Python 3.6. Instructions for placing the dataset and opening the source files are included in README.md of the source code submission.

Results: Linear Regression Model

FIFA Dataset

R^2 Scores: Training Set 0.838, Testing Set 0.842





Testing Set Mean Squared Error: 0.003

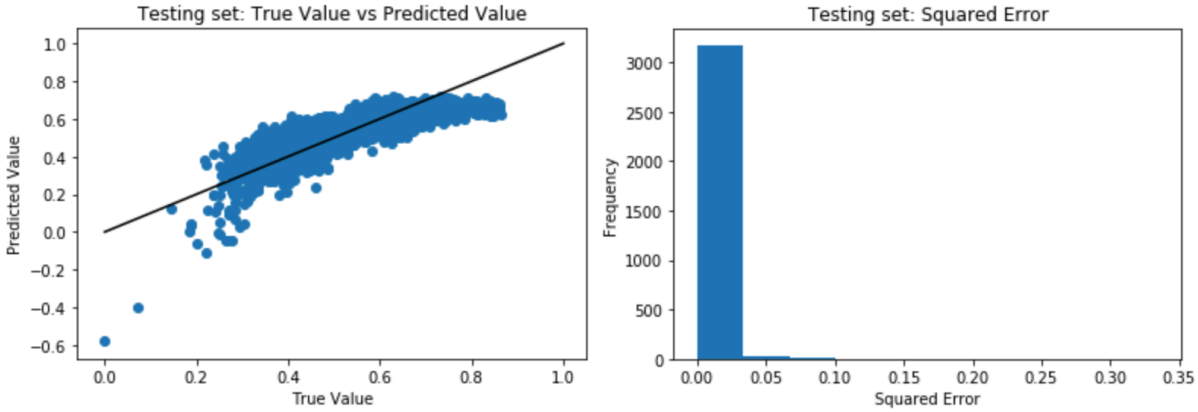
Finance Dataset

R^2 Scores: Training 1.000, Testing 1.000

Testing Set Mean Squared Error: 0.000

Orbits Dataset

R^2 Scores: Training 0.686, Testing 0.695



Testing Set Mean Squared Error: 0.005

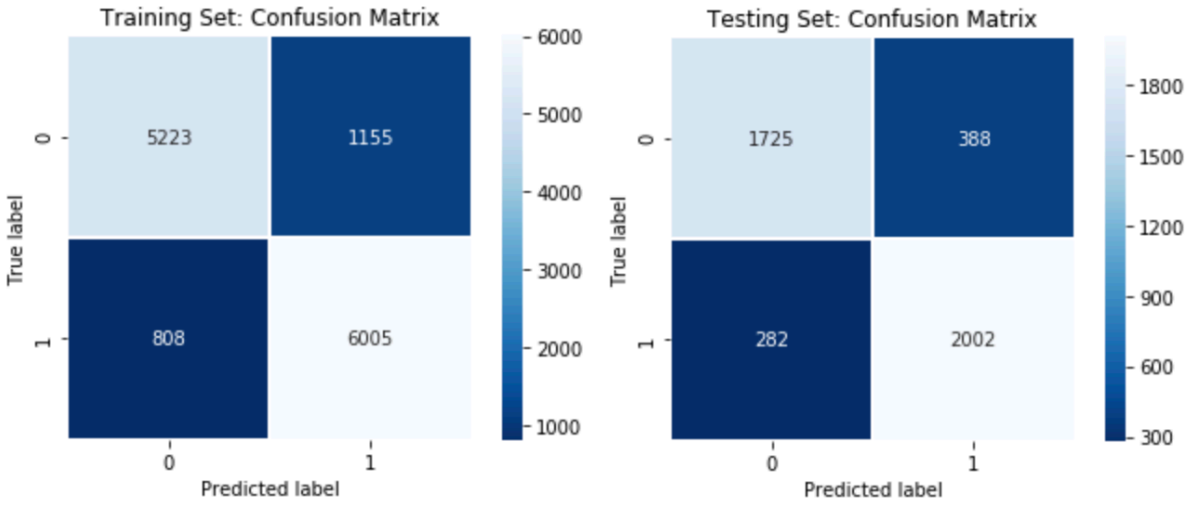
<u>Results: Logistic Regression</u>

Model Settings

- Learning Rate: 1e-2 for FIFA, 1e-1 for Finance and Orbits
- Early Stopping: if validation accuracy not improved by 1e03 in 5 epochs

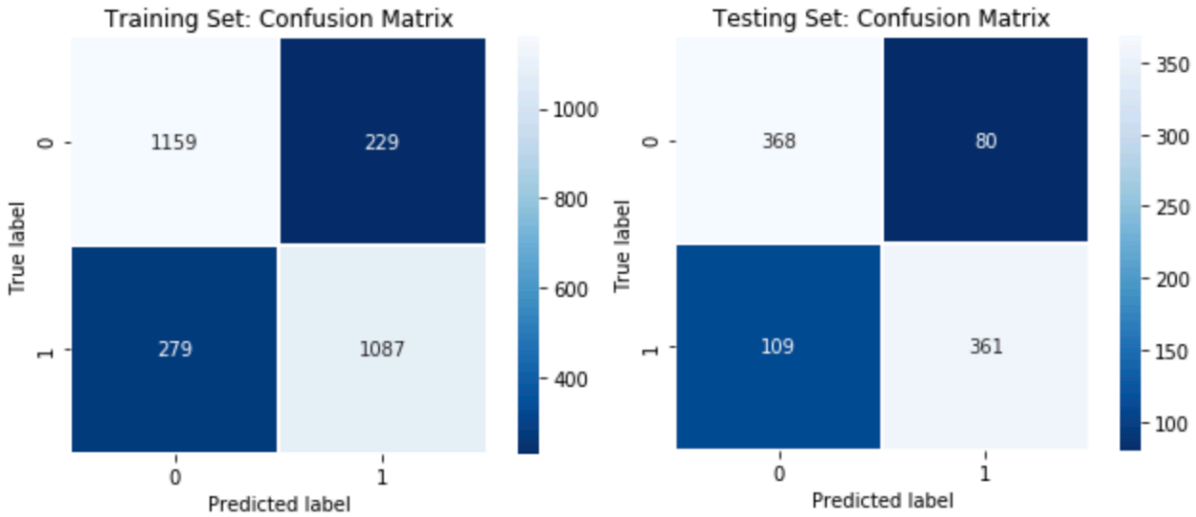
FIFA Dataset

Accuracy: Training 0.851, Testing 0.848



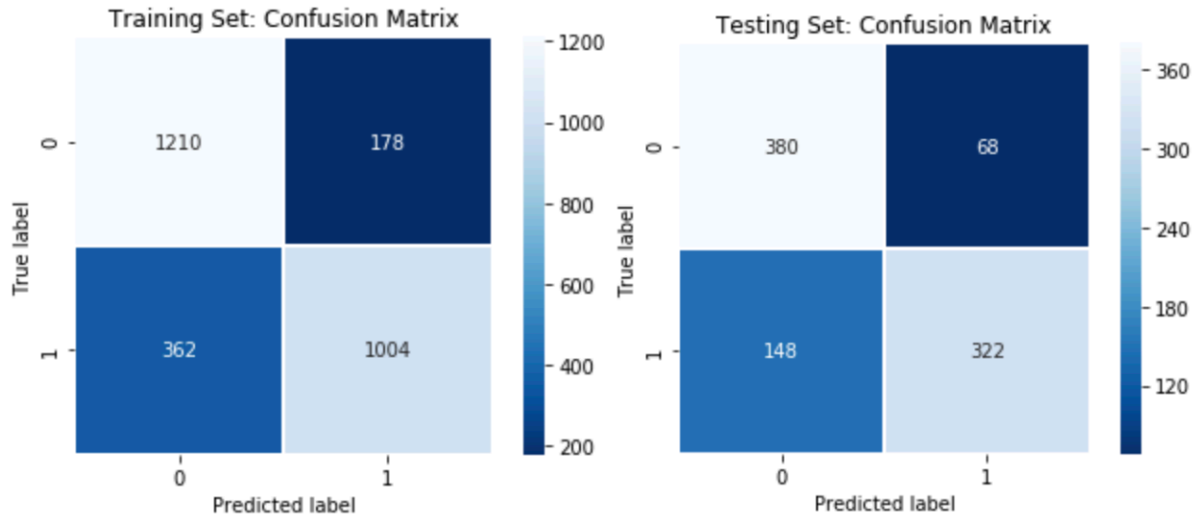
Finance Dataset

Accuracy: Training 0.816, Testing 0.794



Orbits Dataset

Accuracy: Training 0.804, Testing 0.765



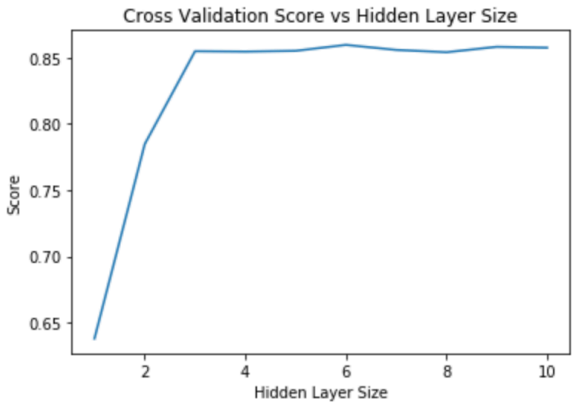
<u>Results: Neural Network</u>

Model Settings

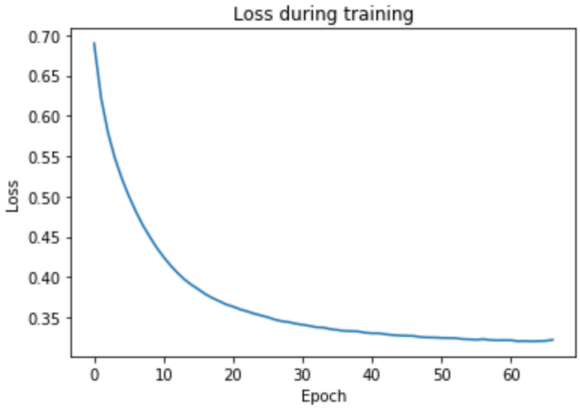
- Hidden Layer Size: Chosen from {1, 2, …, 10} after cross validation
- Activation Layer: ReLU
- Optimizer: Adam (Learning Rate 1e-3, beta1 0.9, beta2 0.999, epsilon 1e-8)
- L2 Regularization Strength: 1e-4
- Early Stopping: If validation accuracy not improved by 1e-4 over 10 epochs

FIFA Dataset

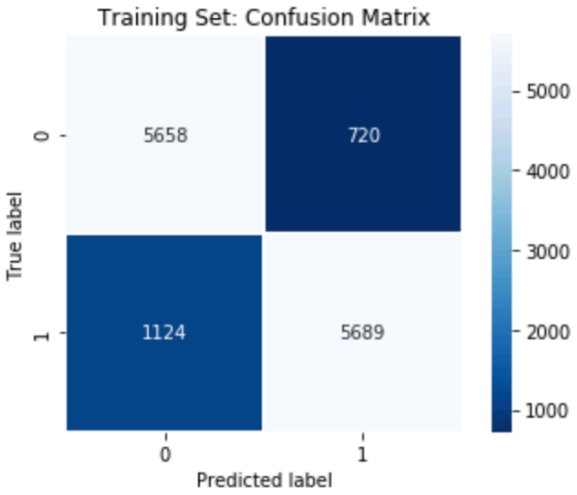
Best Hidden Layer Size: 6

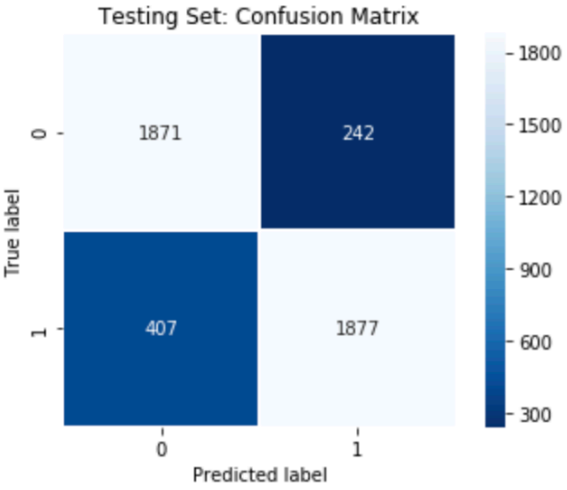


Model Best Loss: 0.322067



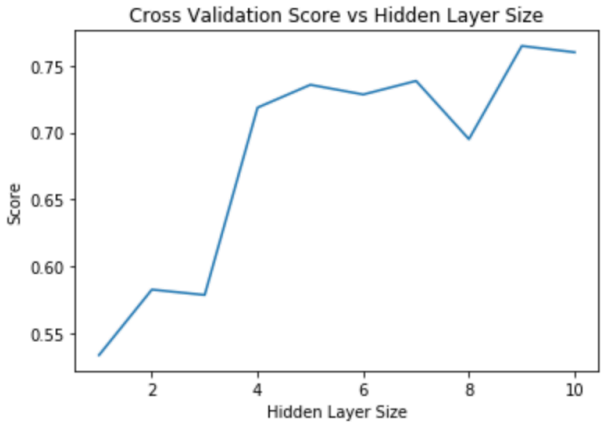
Accuracy: Training 0.860, Testing 0.852



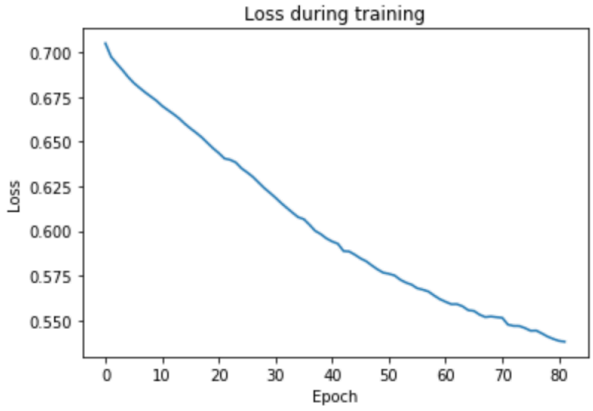


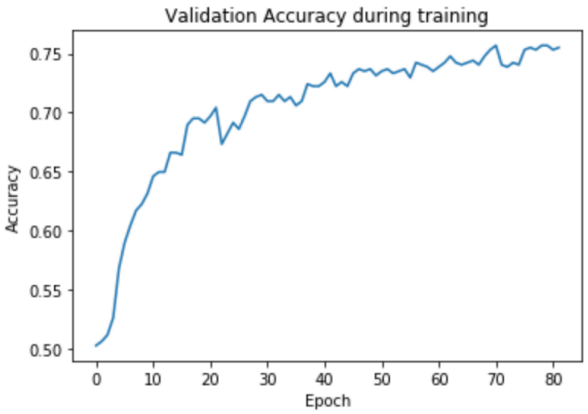
Finance Dataset

Best Hidden Layer Size: 9

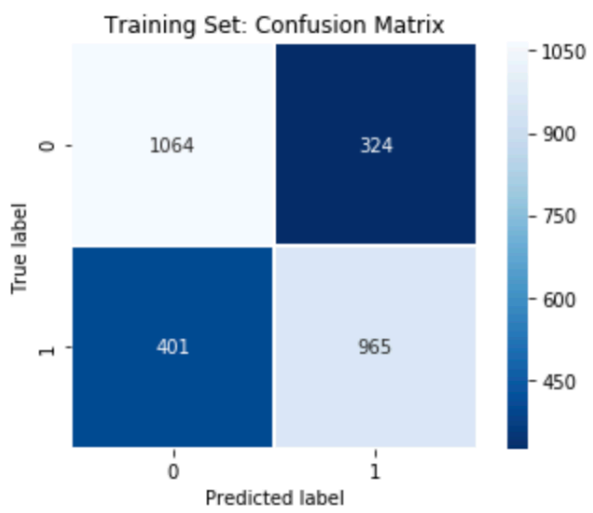


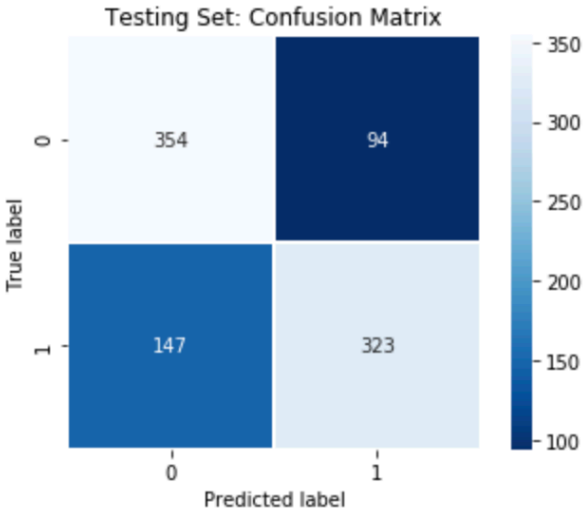
Model Best Loss: 0.537994





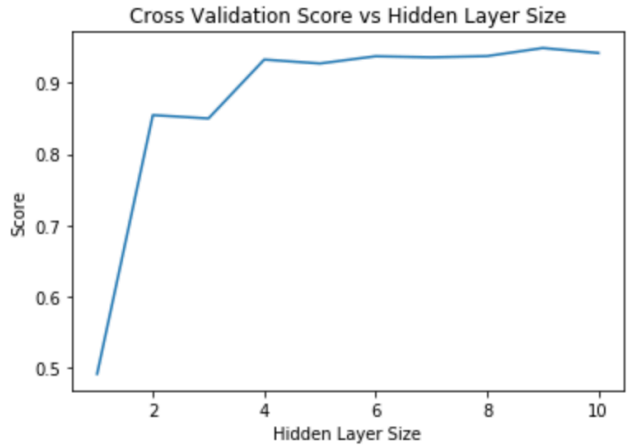
Accuracy: Training 0.737, Testing 0.737



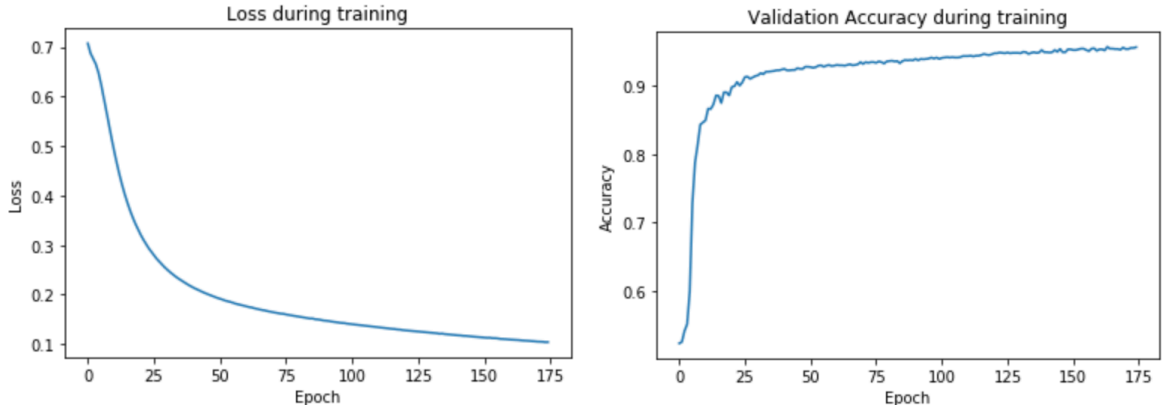


Orbits Dataset

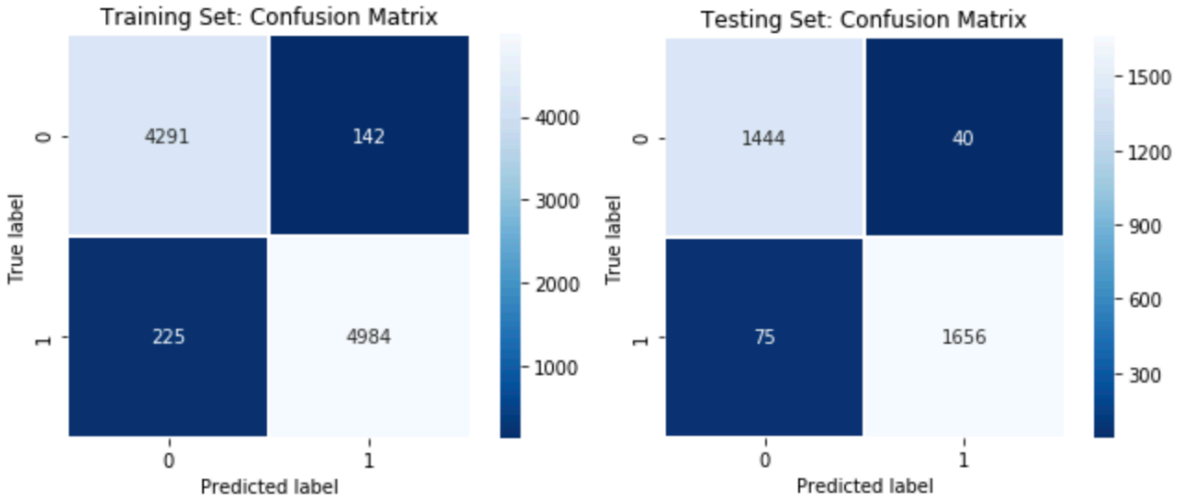
Best Hidden Layer Size: 9



Model Best Loss: 0.104032



Accuracy: Training 0.962, Testing 0.964



Comparing Performance

Time Taken

Below recorded durations includes training, testing and hyper-parameter tuning processes.

	FIFA Dataset	Finance Dataset	Orbits Dataset
Linear Regression	0.016 s	0.004 s	0.003 s
Logistic Regression	0.053 s	0.013 s	0.011 s
Neural Network	50.987 s	10.106 s	45.626 s

Performance

Below figures represents the performance of each model during testing.

	FIFA Dataset	Finance Dataset	Orbits Dataset
Linear Regression (Mean Squared Error)	0.003	0.000	0.005
Logistic Regression (Accuracy Score)	0.848	0.794	0.765
Neural Network (Accuracy Score)	0.852	0.737	0.964

Discussion

From the results listed above, we can see that Linear Regression and Logistic Regression tasks took significantly less time to finish versus Neural Network. One of the major factors is that Neural Network is computationally more expensive to train, while another factor is that hyperparameter tuning was included in the Neural Network task, which scaled up the time taken by the number of parameter candidates.

Linear Regression performed very well across all three datasets, with testing results for Finance Dataset having almost no error at all. Unfortunately, under the scope of this study, no other regression methods were implemented for further comparison.

Comparing the two classification methods, Logistic Regression and Neural Network, both performed similarly in FIFA dataset, while Logistic Regression gives slightly better result in Finance Dataset and Neural Network performed significantly better in Orbits dataset. It is interesting that Neural Network, being the more sophisticated and expensive method of the two, does not necessarily yield better results. Granted that this is not a fair comparison as hyperparameters could be further optimized for Neural Network, we could conclude that Logistic Regression can be seen as a viable option for classification and that Neural Network may not be the best “default option” among the two.