

Assignment 4 Neural Network With TensorFlow Analysis
Ashley Brinker

A. Predicting Loan Defaults

- a. Beginning with the max nodes (twice the number of input variables), we saw some overfitting in our model. By removing nodes until the model no longer improved, we found the best value was a third of our original number of nodes.
- b. Softmax was the activation function for the output layer in all results below. The Tanh activation function took much longer to run than the other functions, with Relu performing the best in terms of duration.
- c. The table below represents the accuracies with all variables used. The best model used the Relu activation function with one hidden layer and a dropout layer, as highlighted below.

Classification Accuracy

Activation Function	Relu				Sigmoid				Tanh			
# Hidden Layers	1		2		1		2		1		2	
Dropout Layer?	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Training Data Accuracy	93%	93%	95%	95%	90%	90%	90%	90%	90%	91%	92%	93%
Testing Data Accuracy	91%	90%	91%	91%	88%	88%	88%	88%	89%	90%	90%	91%

- d. After the best model was found with all variables, we selected only those variables that were selected by the previous Random Forest model, as it has performed the best out of all other models. The results of this model did not improve from the previous best with the following results:
 - i. Training Data Accuracy: 87%
 - ii. Testing Data Accuracy: 86%

B. Predicting Loan Amounts

- With this model we began with the max number of nodes. When we began to remove nodes, we found that the model was not improving by removing a large number of nodes. The best RMSE values were found by using 95% of the max number of nodes.
- Linear was the activation function for the output layer in all results below. The duration for all models was pretty comparable.
- The table below represents the accuracies with all variables used. The best model used the Relu activation function with two hidden layers and no dropout layer, as highlighted below.

RMSE Accuracy

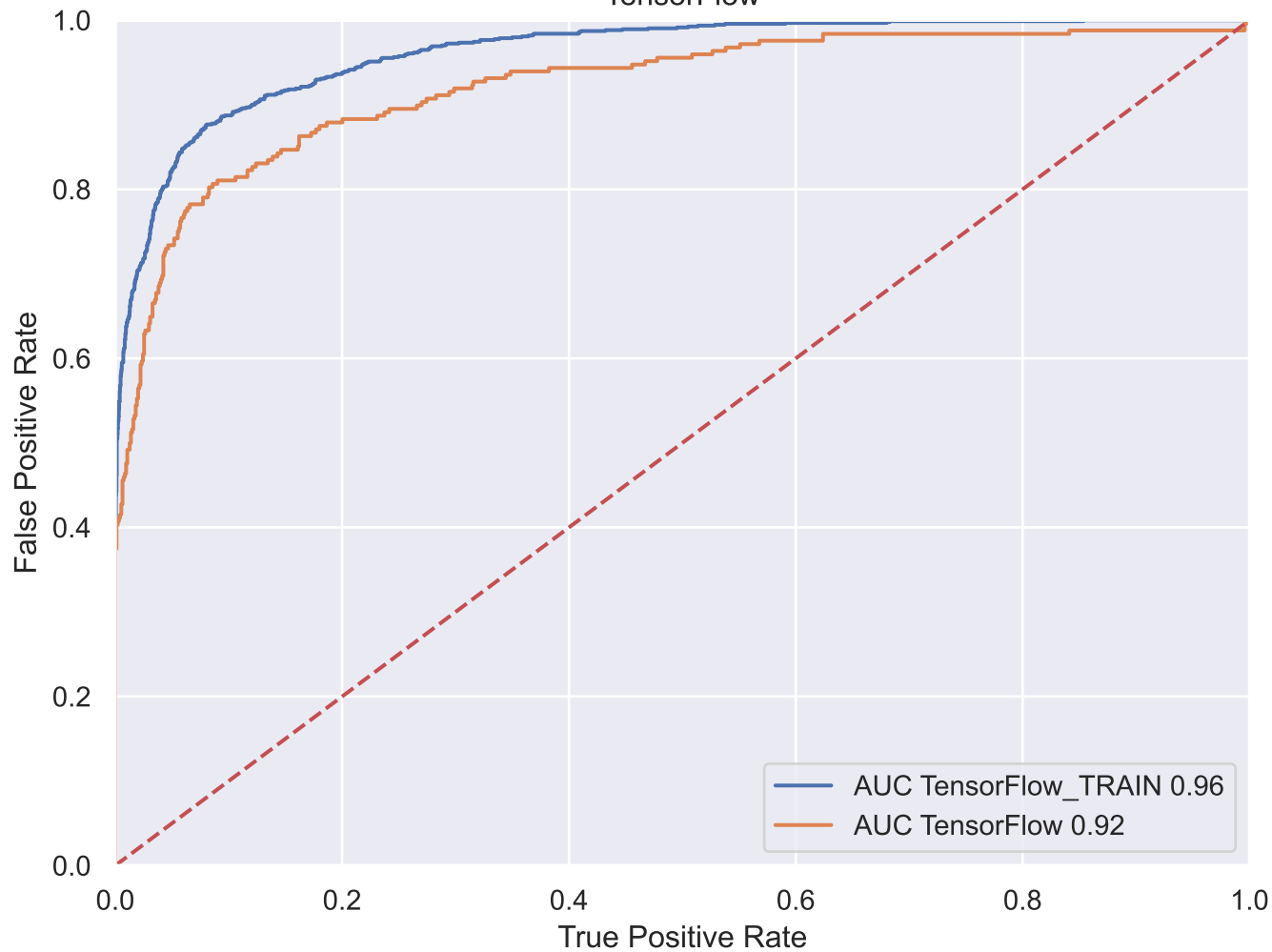
Activation Function	Relu				Sigmoid				Tanh			
# Hidden Layers	1		2		1		2		1		2	
Dropout Layer?	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Training Data RMSE	13338	12790	6061	5698	16793	16792	16836	16848	16788	16787	16795	16793
Testing Data RMSE	12617	12097	6877	6643	15943	15941	15984	15995	15938	15936	15944	15942

- After the best model was found with all variables, we selected only those variables that were selected by the previous Gradient Boosting model, as it has performed the best out of all other models. The results of this model did not improve from the previous best with the following results:
 - Training Data RMSE: 10424
 - Testing Data RMSE: 10071

C. Overall Results:

- a. Classification Model: **The Random Forest model has outperformed all other models (tree based, regression and neural networks) and should be selected.** These results can be found in the ROC curves of all models.
- b. Amount Model: **The Gradient Boosting model has outperformed all other models (tree based, regression and neural networks) and should be selected.** These results can be seen in the RMSE of all models:
 - i. TREE : 6175.72
 - ii. RANDOM_FOREST : 3739.26
 - iii. GRADIENT_BOOSTING : 3147.54
 - iv. REG_ALL : 3926.04
 - v. REG_TREE : 5679.84
 - vi. REG_RF : 4730.41
 - vii. REG_GB : 4441.78
 - viii. REG_STEP : 4441.78
 - ix. TensorFlow : 6869.42

TensorFlow



ALL MODELS

