

Colab


Problem Definition: This project was a continuation of last weeks Kaggle Digit Recognition competition . RandomForestClassifier was previously used to generate a classification model. In contrast, the goal this week was to create a model using an Artificial Neural Network. The model was implemented using Keras, which is a high level library for deep learning that sits on top of TensorFlow.

Research Design/Programming:

To begin, a Neural network without hyperparameter tuning was used to generate a baseline model. The baseline model outperformed last week's RandomForestClassifier with an accuracy score of 96.5 %.

1955

Dan Smith




0.96514

3

36m

Your Best Entry ↑

Your submission scored 0.96514, which is an improvement of your previous score of 0.92714. Great job!

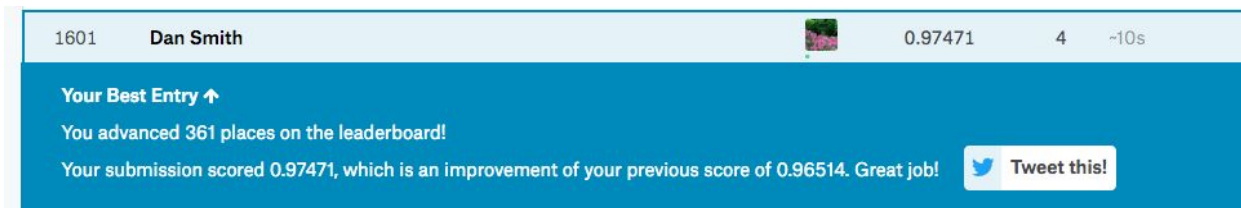
 Tweet this!

The focus moving forward was placed on achieving a higher accuracy score.

RandomizedSearchCV was used for hyperparameter tuning. The number of neurons per layer and the number of hidden layers were manipulated to improve model accuracy. The results in the table below show that the optimal hyperparameter settings are 861 Neurons/Layer and 5 Hidden Layers.

	Mean Test Accuracy	#Neuron/Layer	# Hidden Layers	Mean Fit Time(s)
rank_test_score				
1	0.971405	861	5	63.371054
2	0.969619	431	9	65.512504
3	0.969381	776	8	75.682360
4	0.968048	96	8	43.516399
5	0.967762	467	5	50.172492

The model was retrained using the abovementioned hyperparameter settings and predictions were made on the Kaggle out of sample data.



A screenshot of a Kaggle leaderboard entry for user 'Dan Smith'. The entry shows a score of 0.97471, which is an improvement over a previous score of 0.96514. The user has advanced 361 places on the leaderboard. The snippet includes a 'Your Best Entry' header, a congratulatory message, and a 'Tweet this!' button.

Rank	Username	Score	Placements	Time
1601	Dan Smith	0.97471	4	~10s

Your Best Entry ↑
You advanced 361 places on the leaderboard!
Your submission scored 0.97471, which is an improvement of your previous score of 0.96514. Great job! [Tweet this!](#)

The accuracy improved significantly to approximately 97.5 %.

Recommendations for Management:

The results show that a neural network using 861 neurons/layer with 5 hidden layers is optimal. A neural network without hyperparameter tuning outperformed RandomForestClassifier with tuning by approximately 4 % accuracy. Based on this I recommend a neural network is used going forward.

This was merely an exploratory effort to evaluate neural networks as a viable alternative to RandomForestClassifier. We gained insight from this experiment however, there is room for model improvement. Due to processing power limitations, only 5 iterations were performed for RandomizedSearchCV. 5 iterations of RandomizedSearchCV with 3 fold cross validation and 10 epochs results in 150 model fits. This alone took 1355.63 (s). Exploring more hyperparameter combinations at this point in time with our current computational resources is not viable. Moving forward, I recommend that we invest in a cloud computing platform such as AWS to improve processing capacity. This will allow us to perform a larger and more comprehensive randomized search thus improving model accuracy.