Exercise “Regression with a Multi Layer Perceptron (MLP)” Part 2/3

Prof. Dr.-Ing. Jürgen Brauer

Introduction:

“House Prices: Advanced Regression Techniques“ competition at Kaggle
https://www.kaggle.com/c/house-prices-advanced-regression-techniques

In the last exercise we learned how to …

… read in the training and test data of this competition using Pandas
… access specific rows and columns from the data table (“slicing”)
… do a fast analysis to find out which features (input columns) have a “large” linear correlation with the sale price using Pearson’s correlation coefficient
… plot the data using matplotlib

In this exercise your task is to …
… build a MLP in TensorFlow
… train it using the training data train.csv
… use the trained MLP to predict the sale prices for the 1459 test houses from test.csv
… submit your predicted sale prices and see what your ranking is in the “leaderboard” for this competition at Kaggle!

Detailed steps:

1. Implement a MLP in TensorFlow with a variable number n of input features
The MLP shall have a n-h1-h2-1 topology,
i.e. n inputs, h1 hidden neurons in layer1, h2 hidden heurons in layer2, 1 output neuron
The output neuron corresponds to the predicted sale price.

2. Train your MLP using gradient descent and use a simple loss function, where you compute the absolute difference of the predicted sale price and the actual sale price.

3. Conduct the following experiments with a different number n=1,..,6 of input features:

```
features1 = ['TotalBsmtSF']
features2 = ['TotalBsmtSF', '1stFlrSF']
features3 = ['TotalBsmtSF', '1stFlrSF', 'GrLivArea']
features4 = ['TotalBsmtSF', '1stFlrSF', 'GrLivArea', 'OverallQual']
features5 = ['TotalBsmtSF', '1stFlrSF', 'GrLivArea', 'OverallQual', 'GarageArea']
features6 = ['TotalBsmtSF', '1stFlrSF', 'GrLivArea', 'OverallQual', 'GarageArea', 'GarageCars']
```

For each experiment 1-6 with these different input features, train your MLP 100.000 steps, then compute the average error on the training data.

Question: Do more input features help to achieve a better average error on the train data?

4. Use the best of your 6 trained MLPs to predict the house sale prices for all the 1459 houses in test.csv

5. Generate a predictions.csv file of the form

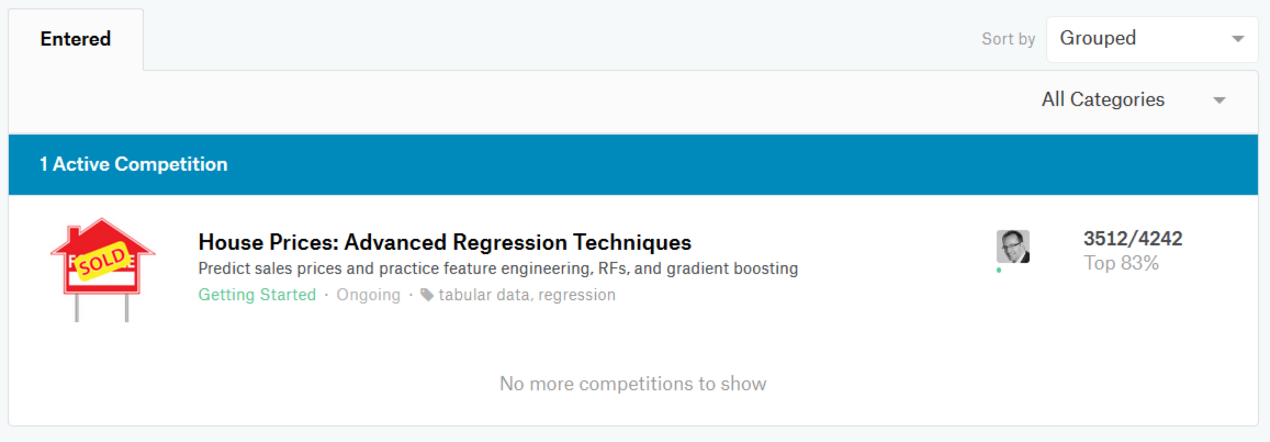
Id, SalePrice
1461, <Predicted Price for House #1461>
1462, <Predicted Price for House #1462>
1463, <Predicted Price for House #1463>
…
2919, <Predicted Price for House #2919>

Then submit your predictions.csv at Kaggle and check your position in the leaderboard.

With such a simple MLP my position was 3512 of 4242.
Great! I am better than 730 other “Kagglers” with my very first submission and there is much room above to improve! 😉

Note: Here we used just 6 of the 80 input features for prediction. Of course, we need more features to improve. This will be done in the next exercises.

Note2: Your ranking can become lower if other “Kagglers” become better.



House Prices: Advanced Regression Techniques

Predict sales prices and practice feature engineering, RFs, and gradient boosting

Getting Started · Ongoing · tabular data, regression

3512/4242

Top 83%

No more competitions to show