

Exercise “Regression with a Multi Layer Perceptron (MLP)”

Part 3/4

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 ...

- ... 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!

In this exercise the task is to ...

- ... work into Jupyter Notebooks
- ... build a MLP in Keras
- ... use scikit-learn’s MinMaxScaler to normalize the data
- ... use all 38 numerical input features for predicting the house sale prices, not just six as in the exercise before
- ... improve your Kaggle leaderboard position!

Detailed steps:

1. Implement a Jupyter Notebook that shows how to implement a simple MLP in Keras

In the exercise present your Jupyter Notebook and go through it step by in order to teach your fellow students how to use Keras to implement and train a MLP.

The example MLP shall learn a simple mapping: for two inputs x_1 , x_2 , it shall learn to compute x_1+x_2 .

2. Work into scikit-learn’s MinMaxScaler

Prepare another Jupyter Notebook that shows how to use the MinMaxScaler from scikit-learn in order to normalize a matrix column-wise. Use your notebook to teach the usage of the MinMaxScaler to your fellow students in the exercise.

3. Use all numerical features for house price prediction

Now use all 38 numerical features that are available for the 1460 training houses, put them into a matrix of shape (1460,38) and normalize them using MinMaxScaler. Then use Keras to implement and train a model that predicts the house prices on basis of a 38-dimensional input vector.

Prepare your solution as a Jupyter Notebook such that you can teach your fellow students how to solve this task step-by-step.

Question: Is a small error on the training dataset a good indicator whether your model works well for predicting house prices?

4. Submit your prediction results

When using all 38 numerical features instead of just six, I could improve my Kaggle position for this competition from 3512 to position 2384.

1 Entered Competition



House Prices: Advanced Regression Techniques

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

Getting Started · Ongoing · 📁 tabular data, regression



Knowledge

2384/4454

Submit your prediction results as well and check whether using more features helped to improve your predictions!