Exercise “Regression with a Multi Layer Perceptron (MLP)” Part 4/4

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 …

… work into Jupyter Notebooks
… build a MLP in Keras
… use scikit-learn’s MinMaxScaler to normalize the data
… to use all 38 numerical input features for predicting the house sale prices
… think about how you could encode non-numerical features as input for a MLP
… improve your Kaggle leaderboard position!

In this exercise the task is to …

… think about how to use the categorial (non-numerical) features that we did not use so far
… preprocess the categorial column data such that you can use it as input for a MLP
… use both numerical and categorial features in order to predict house prices

Detailed steps:

1. Understand the categorial data: Have a look at the file data_description.txt that comes with the dataset and get a feeling what the categorial data looks like. E.g., which values can appear in the “LandSlope” column?

2. How can you prepare a Pandas data frame that only consists of the categorial data columns?

3. How can we count how often each catergorial value (e.g. “Gtl”= Gentle slope) appears in a categorial data column (e.g. “LandSlope” data column)?

4. How can we encode categorial data such that we can use it as input for a MLP?

5. Prepare a Jupyter notebook where you show your solution to 2-4

6. Train a MLP with ALL numerical feature data columns and the categorial data columns in order to improve your prediction accuracy regarding the house sale prices of your test data set