Matlab code

Data directory

Original	Should contain train.csv and test.csv, the given Kaggle data
ConvertedData	Will be created after running PreprocessingKaggleData.m It will contain Train.csv and Test.csv, corresponding to the given Kaggle data and without header row, index column, labels in numeric
SignificanceTesting	Will be created after running PreprocessingKaggleData.m It will contain 6 pairs of ITest, ITrain set, used to train and evaluate models

The given Kaggle data are converted and saved to files in different directories. Doing this way, I could speed up the training time each time I train a model.

Code directory

Utils	Contains functions that are called many times
MLP1	Contains the code to train simple MLP, it does not use weight decay, momentum, etc. The code here are based on my solution for an assignment of the course Neural Networks

Other files

PreprocessingKaggleData.m

Converts the given Kaggle data to the format suitable for processing

Splits the given labelled data into 6 pairs of train and test set

Otto_MLP1.m

Trains and tests MLP using 3-fold cross-validation

Otto_MLP.m

Trains and tests MLP 6 times

Otto_MLP_Balance.m

Trains and tests MLP 6 times, applies under- or oversampling

Otto_MLP_Outliers.m

Trains and tests MLP 6 times, applies removing noises

Otto_SVM.m

Trains and tests SVM 6 times

Otto_TreeBagger.m

Trains and tests TreeBagger 6 times

Otto_TreeBagger_Balance.m

Trains and tests TreeBagger 6 times, applies under- or oversampling

Otto_TreeBagger_Outliers.m

Trains and tests TreeBagger 6 times, applies removing noises

Otto_TreeBagger_Cost.m

Trains and tests TreeBagger 6 times, applies cost function

Otto_TreeBagger_PCA.m

Trains and tests TreeBagger 6 times, applies PCA

Otto_TreeBagger_Pairwise.m

Trains and tests TreeBagger, based on pairwise classification

Otto_TreeBagger_Features.m

Trains and tests TreeBagger, using subset of features

Otto_TreeBagger_Pairwise23.m

Trains and tests TreeBagger on class 2 and 3, using cost function

Otto_TreeBagger_Features23.m

Trains and tests TreeBagger on class 2 and 3, using subset of features

Otto_TreeBagger_Final.m

Trains TreeBagger on full labelled set, and create submission

selectFeaturesLDA23.m

Selects subset of features to classify class 2 and class 3 better

features.csv and featuresSVM.csv

Contains result of feature selection using LDA and SVM, respectively.