Homework 5

Ensemble methods very often generate a significant improvement in the classification accuracy. Your task for the next homework is to test the ensemble learning for user-independent activity recognition:

https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Preprocess the time series data using Fast Fourier Transform and concatenate the acceleration for all the axes into one feature vector. Then use machine learning to classify the signals into different activities. Use Leave-One-Out crossvalidation, where you test on one user and train on the rest of them.

In this task in particular you are required to implement a Python framework for mixture of experts, where the user can assign multiple classifiers to a different region of the feature space. The results of these classifiers are forwarded to an additional classifier (e.g. neural network) that is trained to optimally combine the output from the first stage. Instantiate this framework and use the above mentioned dataset for testing.

It is not necessary to use neural networks, you can also try to use another classifier of your choice that you consider appropriate.

Draw a graph of change of accuracy w.r.t. the number of used classifiers in the first stage. Is there any benefit of using mixture of experts on this dataset? Does the accuracy change if you use overlapping regions instead of disjunct ones?

Implement the solution with Python and numpy and put answers to the given questions in an IPython notebook.

For a detailed explanation of the algorithm, lookup the book "Combining Pattern Classifiers", by L. Kuncheva or consult one of the high number research papers or tutorials where it was used.

Have fun!