Data Science

Final project - Second term

By Noa Nussbaum

In this term's assignment, we were instructed to create three Jupyter notebooks, as well as edit one that we've previously made.

In my first project, the classification notebook, I worked to improve my prediction accuracy results. Since this semester we learnt about several ensemble models, I used the Voting classifier, Stacking classifier, AdaBoost, Gradient Boosting and Bagging classifiers.
My highest prediction accuracy result was given by the Stacking classifier, although it was just as high as last term's results - 94.3%.

In my second project, the Fashion Mnist notebook, I uploaded the data, used PCA to reduce it's dimensions, and proceeded to use the following models in order to reach results: Random Forest, Pipeline, Logistic Regression , KNN, AdaBoost, XGB , Gradient Boosting, Voting classifier. I then checked their accuracy using the accuracy score function and the confusion matrix.

My highest result was 86% accuracy, using the Pipeline model.

The third project is on the cats vs dogs dataset. We were given 25,000 photos of cats and dogs for analysis. I used PCA to reduce it's dimensions. I then used these models for prediction: Decision Tree, Logistic Regression and KNN, as well as these emsemble learning models - Random Forest, Pipeline, , AdaBoost, XGB, Gradient Boosting classifier, Voting classifier. I then checked their accuracy using the accuracy score function and the confusion matrix. The best result achieved was 65.88% accuracy, using the Pipeline model, once again.

The purpose of my final project was to classify between three different situations in the way people communicate with each other, given many datasets describing human hand movement.
I had to upload a lot of data, manipulate it and clean it as instructed (a task I found to be not so simple). I then split the data into training and testing sets, and used several models to make predictions. Amongst others, the models I used were: Random Forest, KNN and Pipeline.

I used the accuracy score function and the confusion matrix to check their accuracy levels.