Notebook no.1- Improvements in Classification Capabilities:

I took a classification problem for whether it will rain or not tomorrow in Australia from Kaggle(Rain in Australia | Kaggle)

The performances from previous Semester were the following:

1. With KNN: the accuracy score was 0.8251.
2. With Logistic Reg: the accuracy score was 0.8281.
3. There have been manipulations on the data- some features were removed manually.
4. The prediction was only on one city in Australia- Sydney.

The improvements of this classification were the following:

1. I've prepared the data, using dummies.
2. I managed to reduce the features to only 1 dimension. As we've learned with less features, the classifier is considered better.
3. Now the classification is being predicted not only about Sydney, but also for Australia.
4. The prediction have been improved to 0.85 – by using KNN. (improvement of 3% in KNN)

Notebook no.2- Fashion-MNIST:

I took the Fashion-MNIST problem from Kaggle(Fashion MNIST | Kaggle)

I used in my work two different classifiers, KNN and RandomForest, performing on fashion-Mnist dataset. First we trained the raw data on both classifiers and have received accuracy of 0.89 with RandomForest and 0.86 with KNN. Also, we've found the best params for our models by GridSearchCV.

Then I tried to reduce the amount of features in our data, to 187 with PCA. Instead of 784 features, and I noticed that it didn't effect our accuracy much- 0.87 with RandomForest and 0.86 with KNN.

Conclusions:

1. The PCA method makes our data more comfortable, with less dimensions, without harming our prediction much.
2. Most of the miss- match predictions were made in between prediction of shorts- that were predicted as: T-shirts, Pullovers and Coats. De facto, it's expected that it'd be classified similarly as they look the same.

Notebook no.3- CatsVsDogs:

I took the CatsVsDogs data from Kaggle(catsVsDogs | Kaggle).

In this notebook we were asked to predict if a given image is cat or dog. From the start I expected that the traditional process of training a large data, as image that consists of pixels and each pixel is represented in 3 dimensions, Red Green Blue, will take a lot of time and resources. So some image pre-processing has to be applied.

The classifier I've used in this notebook is RandomForest classifier.

I tried two image pre-processing:

- Cany Edge Detection – this algorithm trys to recognize the edges in the image. With such pre-processing we achieved accuracy of around 60%.

- Histogram of Oriented Gradients (HOG) – this algorithm represents detection edges in small sectors in the image and translates it to vectors of oriented gradients. With the pre-processing we achieved accuracy of around 75%.

After the HOG we also apply an PCA to reduce the number of features from 40,000 to 606. And we see that it doesn't affect much our accuracy (73%).

Notebook no.4-Hands:

I received raw data from actual experiment. Its purpose to classify if the hand movement between two people is Sync, Spontaneous or just moving alone.

We were obligated to clean the data before we use it.

I wanted to make each row to contain movement of right and left hand together.

Also for alone, we have to add dummy data of movements for right hand.

The classifier I used was decision tree classifier. From this we managed to classify alone with 100 precents accuracy, sync was around 80 precents. But in spontaneous classification there was the worst accuracy of about 96 precents. In addition something about the Sync classification about Oded Medina isn't clear.

I'd expect such accuracy in alone classification, since the data of movements in right hand contained always the same data, it's indeed logical that it could classify this as alone in a good level.