

Data Visualization - Semester B Final Project

Author: Elinor Buskila

Part 1: Improvement of model from semester A - Bank Churn Model

The purpose of this notebook is to take a classification model I made last semester and improve it. This is made possible through the models and methods learned during the current semester: dimensionality reduction by PCA, clustering by K-Means, ensemble learning, etc.

In this notebook I have used the following methods to improve the performance:

- Voting (soft,hard)
- XGBoost
- AdaBoostClassifier
- StackingClassifier

But I did not find place for improvement.

Part 2: Part 2: Fashion Mnist Dataset

This is a multivariate classification problem with ten different types of fashion items that we need to classify. The success rates in this issue are quite high because the images are beautifully arranged. Accordingly, we would like to create a compact model that requires as few pixels as possible in order to successfully classify our labels.

The tools that were used:

1. Voting
2. XGBoost
3. KNN
4. RandomForestClassifier
5. PCA
6. Stacking

Part 3: Dogs VS Cats Dataset

In this part, we will get a huge amount of pictures of dogs and cats. Some are simple and nice pictures and some are pictures that can be very confusing to a computer. We will try to find for us a model that will do its best to classify between pictures of dogs and cats. Recall that even in this part, the use of deep learning methods is prohibited, although they are the leading methods today in the image classification process. According to the knowledge we have earned, we will try to maximize the capabilities of machine learning in order to build the best mode we can offer.

The tools that were used:

1. KNN
2. XGBoost
3. RandomForest
4. Stacking

5. K Means (pre-processing)
6. PCA
7. openCV

Pre-processing: PCA to 90% of the variance and image standartizing.

Part 4: Hands sync

I tried to classify all kinds of hand movements according to the situation they were in. That is, we would like to classify in which situation the person was when he moved his hands.

The data conducted of 3 states:

1. Spontaneous - 2 people moving their hands spontaneously in front of each other
2. Sync - 2 people trying to sync their hand movement
3. Alone - a person that moves his hand randomly

We have had to determine whether we can use machine learning to see if we can understand whether the hands are moving spontaneously, in sync, or just alone.

Pre processing: Even though I made a great progress in this part, I also got stuck on it and could not finish it on time due to miserable combination of work and then sickness that came on the days I took off work to work on the project.