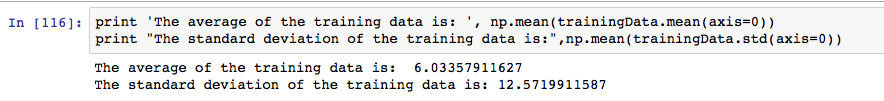
# Objective

This project is intended to classify data entries in a csv file. The given labeled dataset includes 3 parts, training data in a .csv file including 58 columns of float numbers, training label in a separate file with 0 or 1 labeled, and a testing file waiting for corresponding labels.

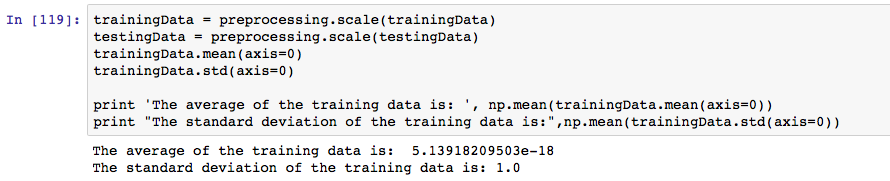
This project is a classic classification problem, so three classification models are used, SVM, linear SVC and Logistic Regression.

# Process:

1. Load csv data into pandas data frame for processing.
2. Get the mean and standard deviation of the training and testing data, as a result:



1. Normalize the data so as the standard deviation is 1.



1. Build 2 classification models using SVM, linear SVC and logistic regression, with k-fold cross-validation, where k=10 in this project.
2. Run the 3 models with training data to compare the result
3. Change the parameters of the three models and try the results, finally get the best ones.



1. Post-processing: select the model with the best accuracy, which is 0.92888 for SVM and save the result to csv format.