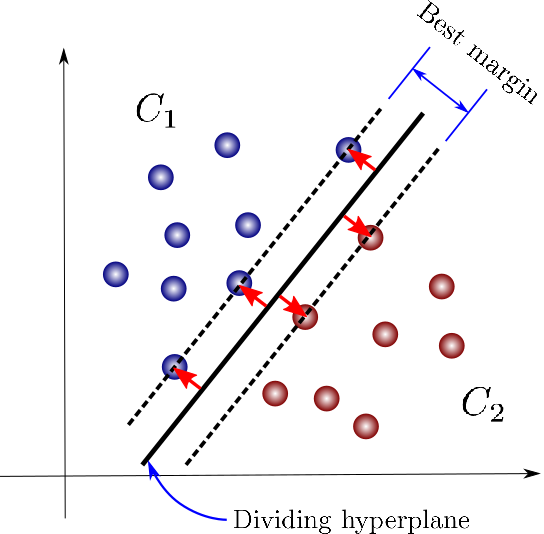
**SUPPORT VECTOR MACHINE**

Support vector machine (SVM) is a useful technique for data classification, regression and prediction. Previously there has been a lot of study using artificial neural network (ANN) in these areas, especially in the field of prediction. However, in the stock market, because the data often has enormous noises and structure dimensionality, the ANN method has some limitations. Support vector machine can treat higher dimensional data better even with a relative low amount of training set. It can present a good ability of generalization for complex model.

* The goal of SVM is to create a flat boundary called Hyperplane, which divides the space to create fairly homogeneous partition on either side.
* SVM can be used both classification and regression.
* Maximum margin should be their for two data points.
* Using **kernel trick** to classify non-linear classification.
* In the field of prediction for stock market and account, the most important thing is to improve the prediction accuracy rate. However, little study has validate the suitability of stock market prediction by SVM.
* Mathematical logics are kept secret in **Black Box** technique.



In this article, we propose an original and universal method by using SVM with financial statement analysis for prediction of stocks. Commonly there are so many technical analyses for prediction in stock market. But these technical indices such as RSI (Relative strength index), BIAS, etc. appear to fluctuate with the quantity of stock exchanges. Compared with the technical indices, the financial indices from the financial statement are much more reliable, nonvolatile and accurate. The goal of this article is to improve the accuracy rate of prediction and to meet different kinds of stockholders’ expectations.

Requirement:

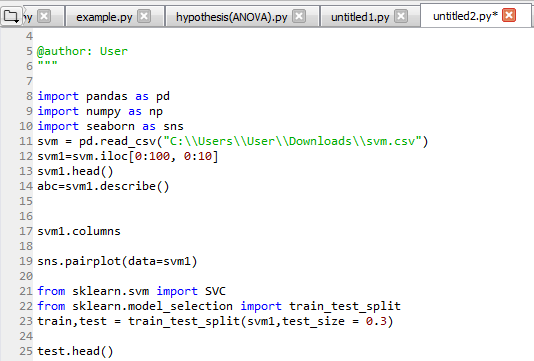
The research data used in this study is selected from the financial distress released by the stock companies. By the ending of this study, the numbers of the financial distresses released by the stock companies is 422.

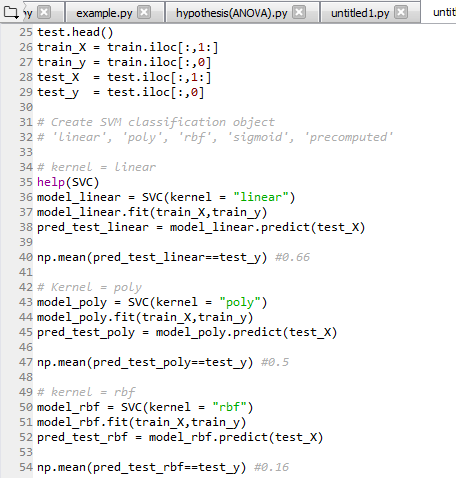
Data set deals with the financial distress prediction for a sample of companies.

* In the below company 2 is financially distressed at time 4 but company 2 is still healthy at time 14.

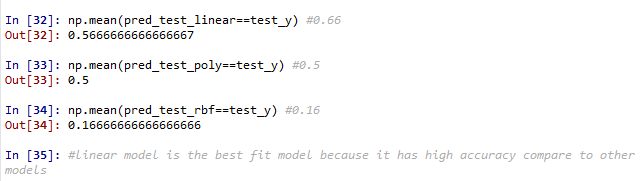
Python code:

* Package:
* import pandas as pd is used for data management
* Import numpy as np is used mathematical operations
* import seaborn as sns is used to visualize the pairplot
* from sklearn.svm import SVC is used to fit the data you provide, , returning a best fit “hyperplane” that divides .
* from sklearn.model\_selection import train\_test\_split is used to split the data in train and test





**Output:**



**R code:**

