National University of Singapore
School of Computing
Tutorial 2:
**SUPPORT VECTOR MACHINE**

Introduction to Support Vector Machine

Supervised Machine Learning algorithm used to split data. It takes as input data and output a plane (or line) called SVM which separates the data. Its separation is based on some data points (not all data points). These data points are called the support vectors.

It uses geometry to solve the problem. Plots each data item in n-dimensional space, with values of each feature being coordinate. Then finds plane (called SVM) which differentiates different classes.

For multi-class classification problem, it forms n * (n-1)/2 classifiers i.e. separate based on "In class" or "Not in class".

1. Which line can be considered as SVM? (A or B or C)?
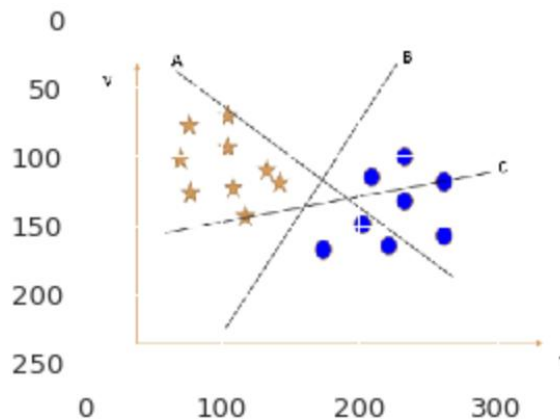
2. All the lines below will separate the data. Which of these lines is considered as SVM?
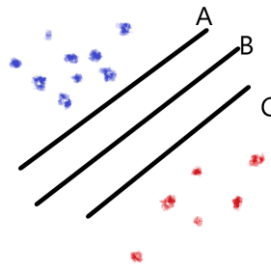
3. Use sklearn SVM.svc classifier for the following classification:

   Classify whether a recipe is of Muffin or Cupcakes:

a. You are provided with CSV file "recipes_muffins_cupcakes.csv". Read the CSV file into Pandas dataframe.
b. Divide the data into X_train, X_test, y_train, y_test keeping train to test data ratio to be 80:20.
c. Define an SVM.svc classifier and fit the train data.
d. Predict the test_data and print the accuracy of classification.
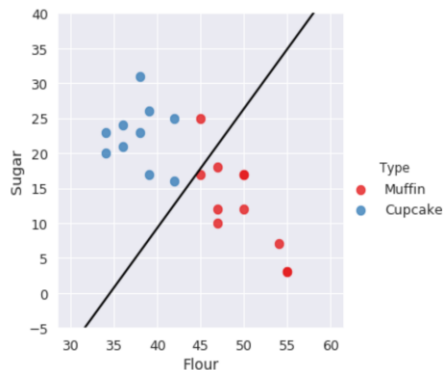
There are 3 important parameters of SVM.
- kernel
- C parameter
- gamma parameter

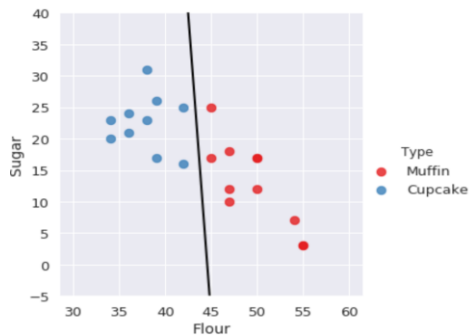4. Which of these is a valid value of kernel parameter?
A. Linear
B. RBF
C. Poly
D. All the above

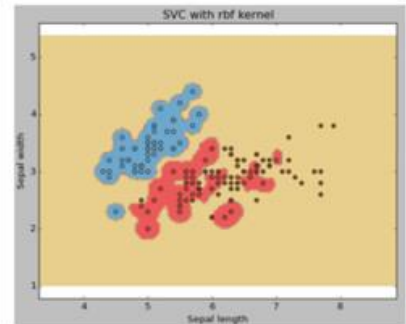5. Which of the 2 SVMs below do you think have higher C parameter value?

A.

B.
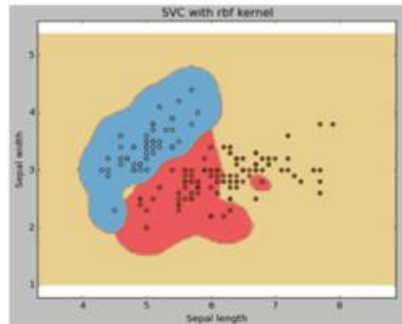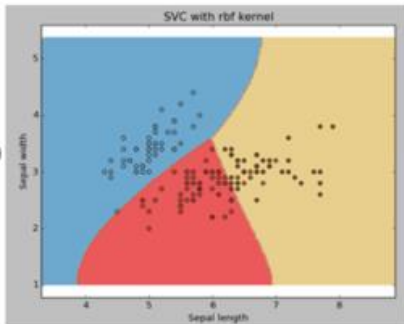
6. Which of the 3 pictures below have highest gamma values?
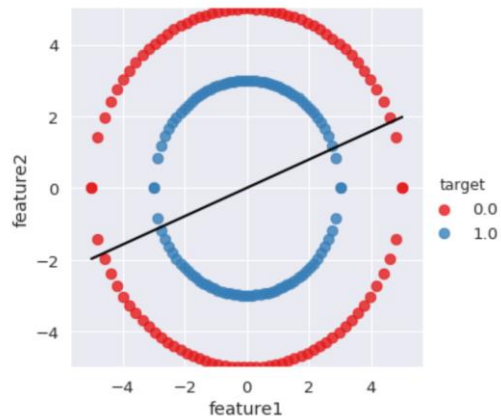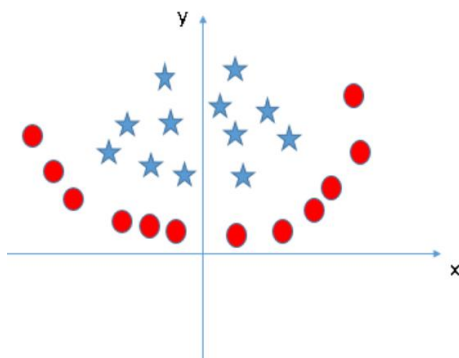
A.   B.   C.

7. What additional feature would be required to create a linear hyper-plane for segregating the classes in the two diagrams below?

8. Use sklearn svm.SVR classifier for time series data
    a. Use pandas 'read_csv' to read 'price.csv'
    b. Create X_train, X_test, y_train, y_test. Put 1st 15 points in training data and 5 points in testing data.
    c. Fit your data into svm.SVR using different kernels.
    d. Plot your predicted fitted data vs actual data points.
    e. Predict for test data and print Mean Square Error in data points predictions.