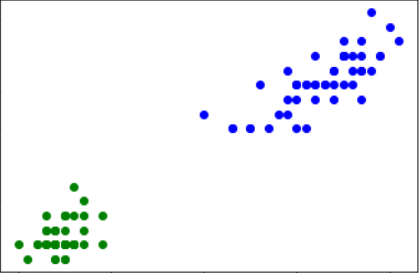
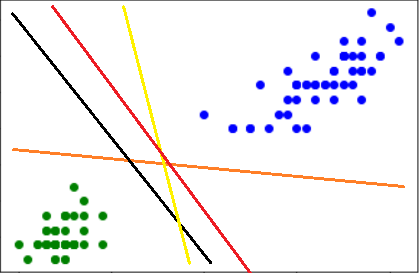
**Support Vector Machine**

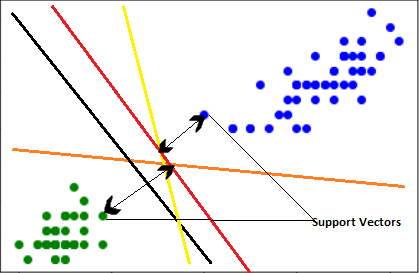
Support vector machines was initially designed for prediction of numeric and categorical outcomes. Today SVM is mostly used as a classification technique for predicting categorical outcomes.



Above mentioned graph represents two categories/groups, one with blue color and other with green. In SVM we need to draw a line also called as Decision boundary to predict the test data correctly, where the test data belongs to blue or green category.



But the problem here is that many such decision boundaries are possible, but we need to choose one best/optimal line.



Red is the best decision boundary because we need to have the distance between the line and data points as maximum as possible to get better prediction. With this condition Red line is the clear winner. That distance between the decision boundary and data point(also called support vectors) is called margin. Margin is considered from both side of the decision boundary.

Now Lets perform SVM with python

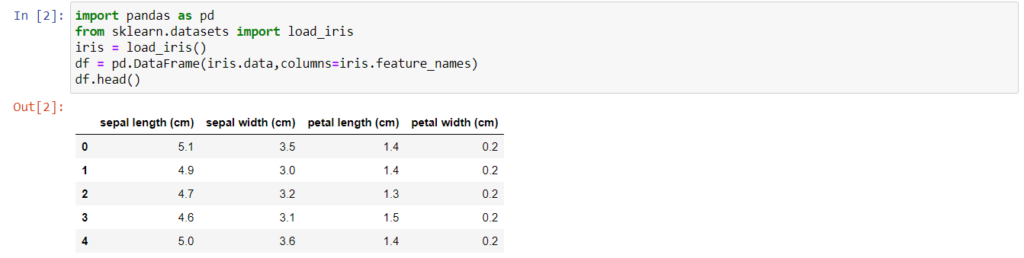
import pandas as pd

from sklearn.datasets import load\_iris

iris = load\_iris()

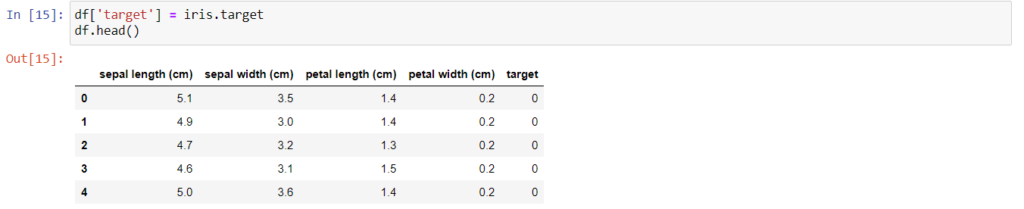
df = pd.DataFrame(iris.data,columns=iris.feature\_names)

df.head()



df['target'] = iris.target

df.head()



from sklearn.model\_selection import train\_test\_split

X = df.drop(['target'], axis='columns')

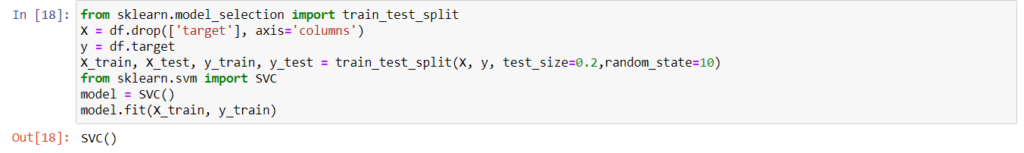
y = df.target

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2,random\_state=10)

from sklearn.svm import SVC

model = SVC()

model.fit(X\_train, y\_train)



model.score(X\_test, y\_test)



model.predict([[5.1,3.7,2.5,0.9]])

