

```

from sklearn import svm
from sklearn.naive_bayes import GaussianNB
import csv
import numpy as np
import matplotlib.pyplot as plt

f = 'data/kaggle_train_tf_idf.csv'

with open(f, 'r') as fin:
    file_reader = csv.reader(fin)
    next(file_reader)
    data = np.array(list(file_reader), dtype=np.float)

NUM_TRAININGS = 3000
X_train = data[1:NUM_TRAININGS, 1:-1]
Y_train = data[1:NUM_TRAININGS, -1]
X_test = data[NUM_TRAININGS:, 1:-1]
Y_test = data[NUM_TRAININGS:, -1]
X = data[1:, 1:-1]
Y = data[1:, -1]

def get_error(G,Y):
    error = 0
    for i in xrange(len(G)):
        if G[i] != Y[i]:
            error += 1

    return 1.0 * error / len(G)

clf_NB = GaussianNB()
clf_NB = clf_NB.fit(X_train,Y_train)
G_test = clf_NB.predict(X_test)
test_error = get_error(G_test,Y_test)

K = 5
from sklearn import cross_validation
clf_NB = GaussianNB()
scores = cross_validation.cross_val_score(clf_NB, X, Y, cv=K, scoring='accuracy')
avg_score = sum(scores) / len(scores)
print('Scores = {}'.format(scores))
print('avg_score = {}'.format(avg_score))
print "test_error of naive_bayes classifier:%.6f" % test_error

# svm
clf_SVM = svm.SVC(kernel='rbf',C=1)
# cross validation
print "Performance of SVM:"
K = 5

scores = cross_validation.cross_val_score(clf_SVM, X, Y, cv=K, scoring='accuracy')
avg_score = sum(scores) / len(scores)
print('Scores = {}'.format(scores))
print('avg_score = {}'.format(avg_score))

```