Anshul Shivhare MTech AI 17828

MLSP assignment 3

Problem 5

Supervised Sentiment Analysis using SVM:
Dataset Details : (./data/movieReviews1000.txt)
Movie Review data (each line is a individual review),Each Review has a label associated with it either 1 (Positive Review) or 0 (Negative Review). Label is mentioned at the end of each review.

Features :

We extract TF-IDF feature for each Review
Reduce the TF-IDF feature dimmension to 10 using PCA
Split Reviews for Training (700 Reviews) and Testing (300 Reviews)
Tfidvectorizer was used from sklearn library for TF-IDF model.

TF-IDF

Here in TF-IDF, rather than focusing on regular words, the words with more importance are taken
into consideration. In TF-IDF model TF stands for term frequency and IDF stands for inverse document. In TF-IDF the words which occurs most in every document is given less Weightage and the words which occurs more in a single document (not in other documents),then it is given more weightage.We make a vocabulary which measures the informativeness of word(i.e log(N/ni)).

N is total number of documents, ni is the word appearing in ith document.IDF reduces the weight given to common words, and highlights the uncommon words in

Library Used for Implementing svm: Sklearn
Library Used for Implementing pca: from sklearn.decomposition

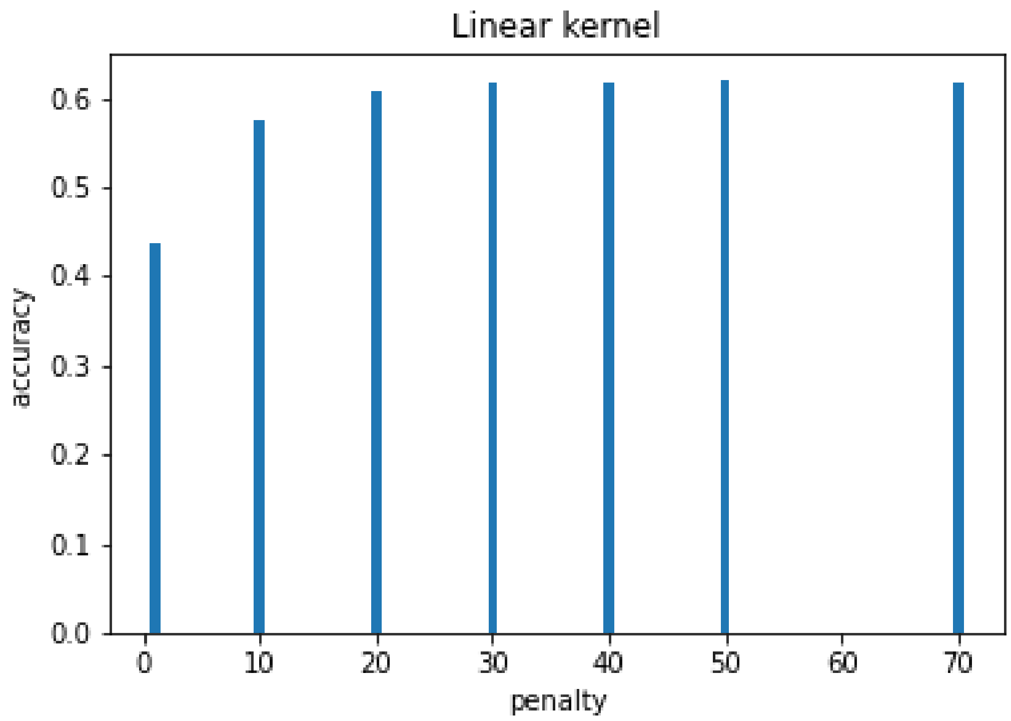
Applied PCA to whole data, for k=10.
The data is split into 2 , train and test. First 750 samples are taken as train data and the rest 250 are taken as test data.
then labels are extracted from those train and test data and stored in y_train and y_test.

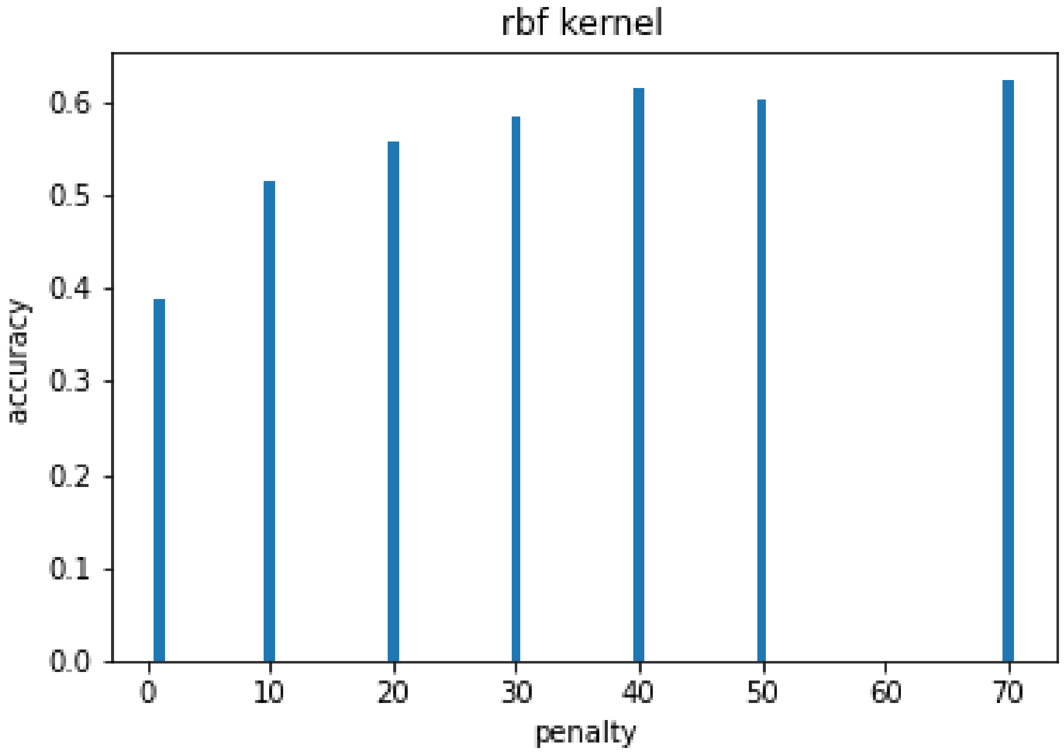
the dimension of whole data before applying pca and before dividing it to test and train : (1000, 3109)

the dimension of X_test after applying pca is : (250, 10)
the dimension of X_train after applying pca is : (750, 10)

the dimension of y_test that is labels of test dataset : (250,)
the dimension of y_train that is labels of train dataset : (750,)

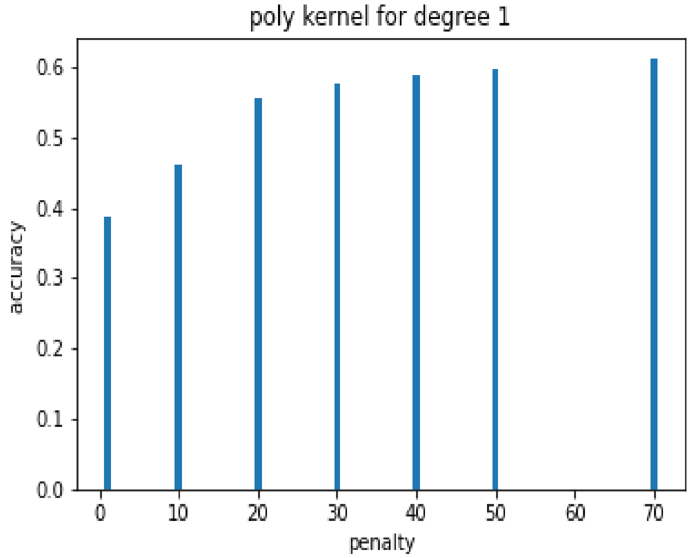
Then for different kernels and using different penalties the accuracy is compared.the graphs regarding those are plotted below:

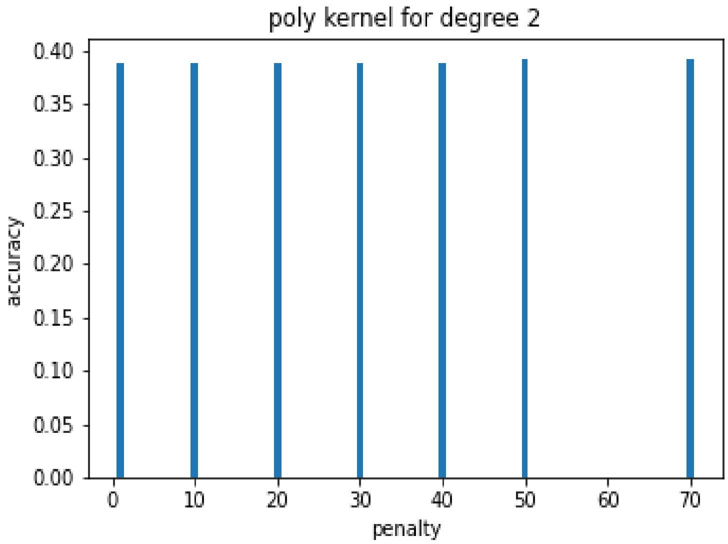


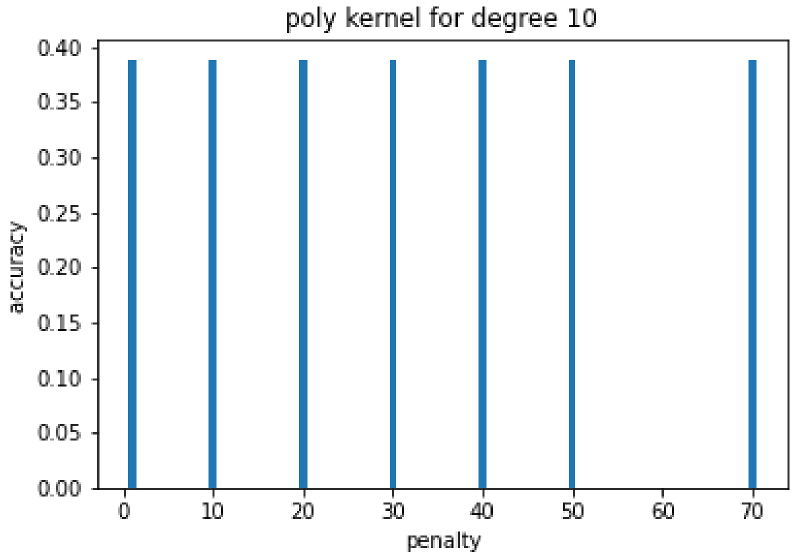


for the above data set, Linear and rbf kernel gives almost same accuracy for different penalties. Though rbf kernel is allways a little less accurate for the current data. when there is 0 penalty then rbf is less accurate compared to linear, but as penalty increases both of them performs same.

For polynomials with different degrees the accuracy bar graphs are below:

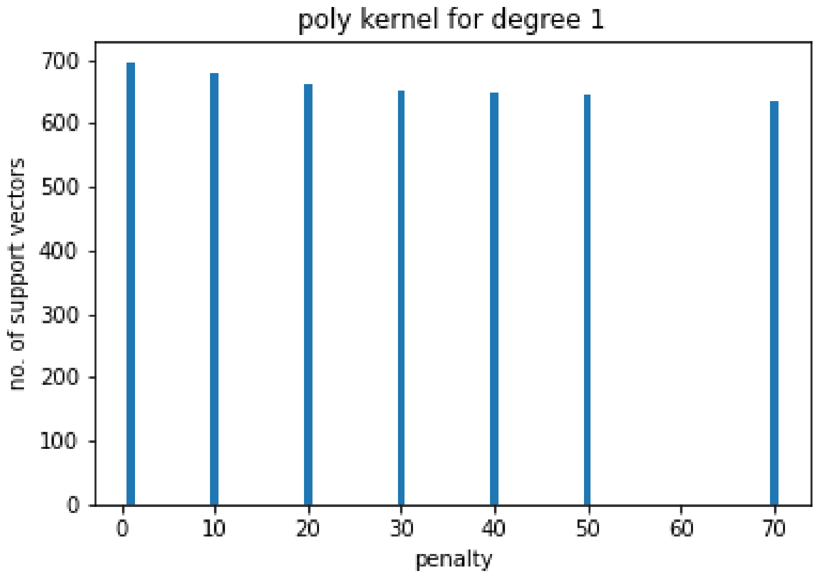






polynomial kernel, we know that as we increase the degree of polynomial our training accuracy increases.on comparing linear kernel with degree 1 polynomial , we are getting the same accuracies. but as degree increases our accuracies with penalty do not vary much in polynomial kernel, this may be because the data is already overfitted.
hence as we increase the degree of polynomial ,the traning data fits more properly ,but on test data the accuracy decreases . for the above data linear (degree 1) polynomial is best fit , because after degree 2 we are getting less test accuracy (overfiiting on train data).
so here the order of polynomials play a very important role for generalization purpose. also as the order of polynomial increases both kind of error decreases ,later is becomes stable.

the below plot is for degree 1 polynomial with penalty vs no. of support vectors



for all types of kernel, if we increase the penalty for misclassification then accuracy increases. but here mostly accuracy becomes stable after few penalty as we have less data .

Note:Estimated parameters converge in probability to true parameters when the number of samples is sufficiently large.

For penalty 10 ,i.e C=10 we have following observation for all 3 kernels:

Kernel	Number of Support vectors	Classification Accuracy
Linear	626	0.62
RBF	664	0.57
Polynomial(degree 1)	678	0.46

Problem 4:

Supervised Sentiment Analysis using SVM:

Dataset Details : leap.ee.iisc.ac.in/sriram/teaching/M LSP 21/assignments/data/Data.tar.gz

15 subject faces with happy/sad emotion are provided in the data. Each image is of 100 × 100 matrix .

Performed PCA to reduce the dimension from 10000 to K. Implemented a classifier on the training images with linear kernel based support vector machine.

Library Used for Implementing svm: Sklearn

Library Used for Implementing pca: from sklearn.decomposition

First the data is fetched (both train and test), and there labels are separated. labels are stored in y_train and y_test.

The dimension of X_test before applying pca is :(10,10201) and of X_train is (20,10201).

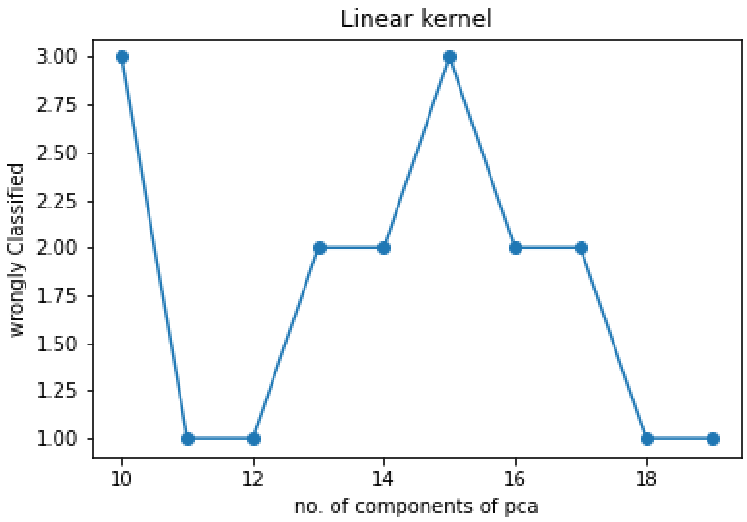
the dimension of y_test (i.e labels of test dataset) is (10,) and dimension of y_train (i.e labels of train dataset) is (20,) .

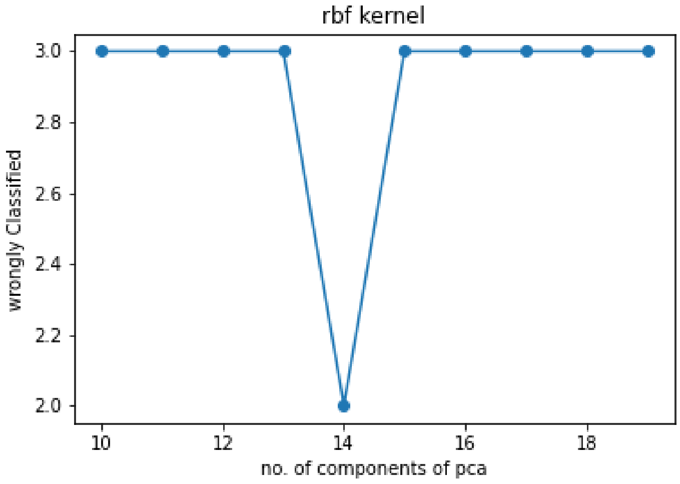
4 kernels are compared (linear , rbf , polynomial, sigmoid). parameters C and epsilon is varied for all. With different K the graphs are plotted .

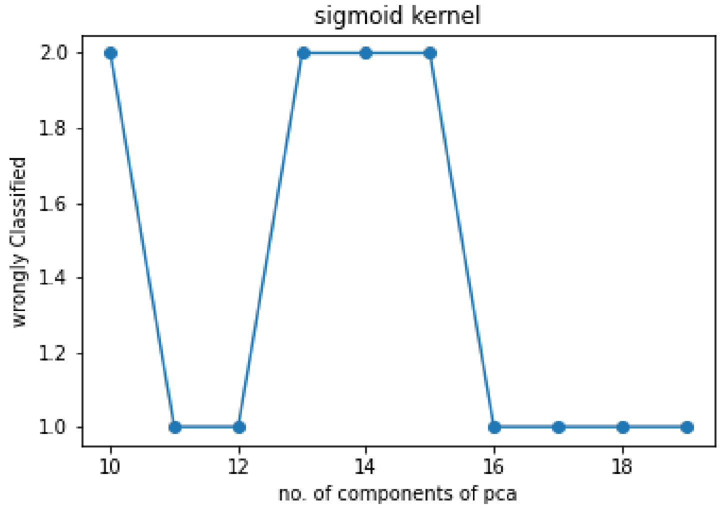
C=[0.2,1,10,100,500] and epsi= [0.1,0.001,0.0001,0.00003] are the values which are varied for every kernel. and for polynomial kernel extra parameter degree is varied. Degrees =[1,2,5,8,10] .

few of the graphs and observations are plotted below:

Note : The value of K is taken from 10 to 21 instead of 1 to 21 , as for 1 to 10 the stooping criteria(epsilon) is satisfied very very late, hence it was taking a lot of time (especially for linear kernel).



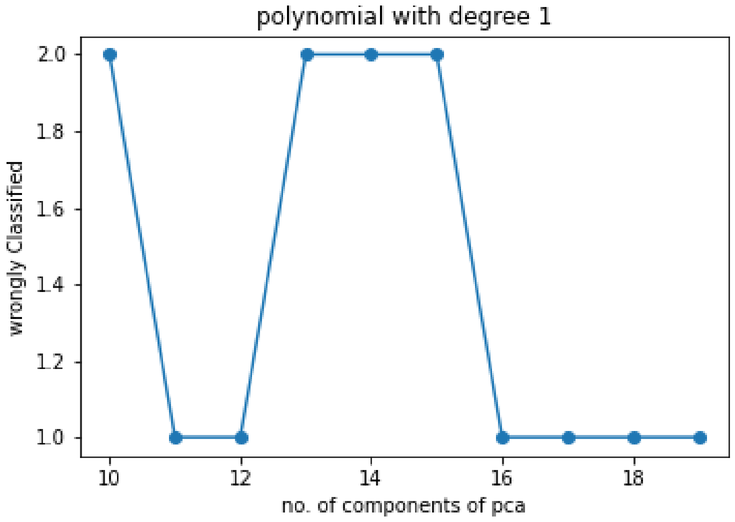


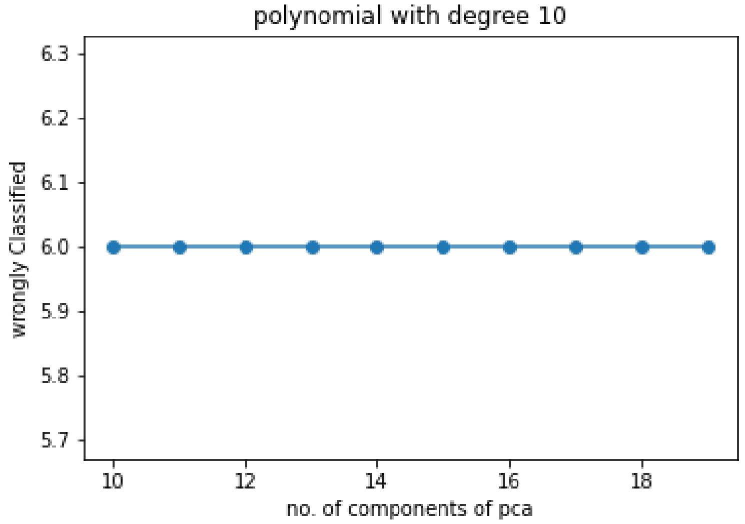


for linear classifier ,on varying the value of components in linear kernel we encounter least missclassification as 1 and maximum misclassification as 3.

for rbf kernel ,on varying the value of components in rbf kernel we encounter least missclassification as 2 and maximum misclassification as 3.

for sigmoid kernel ,on varying the value of components in sigmoid kernel we encounter least missclassification as 1 and maximum misclassification as 2.
on comparing rbf and linear ,linear performs far better than rbf and sigmoid . wheareas sigmoid is better than rbf.



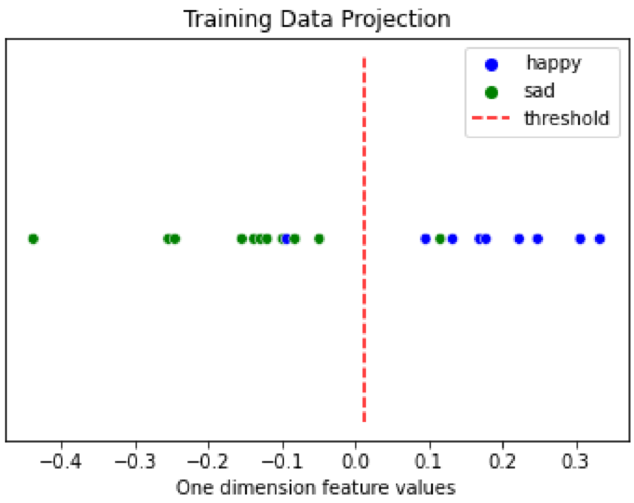


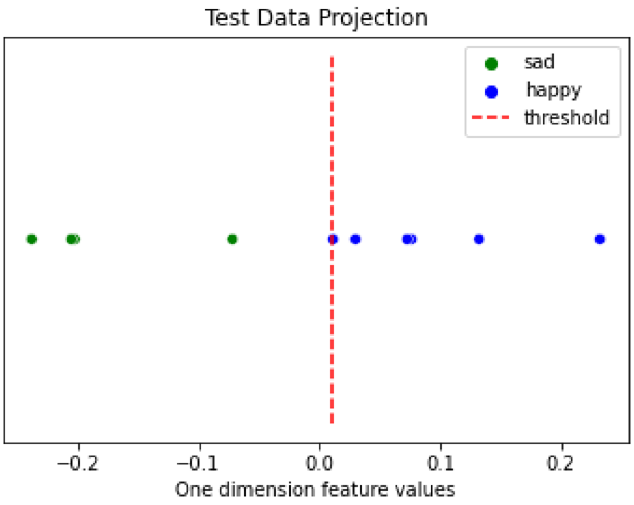
or polynomial kernel with different degrees , polynomial with degree 10 is missclassifying the same number for all K's , this is because the data is already overfitted ,hence not varying much with change in value of k.
and for 1 degree polynomial,on varying the value of components in polynomial kernel we encounter least missclassification as 1 and maximum misclassification as 2.

LDA vs SVM

FOR This Particular data set ,LDA is better performing than svm for K=10(Pca component), rest for other K's it is approximately same as more or less missclassification is 1 ,2, or 3 .
both of them are kind of simliar in performace in all the case ,because data is almost linearly separable, for 1 degree polynomial.

plot for test and train accuracy of lda is plotted below : For K=10


Training Data Projection
● happy
● sad
threshold
One dimension feature values


Test Data Projection
● sad
● happy
threshold
One dimension feature values