"[CMSC828C] Project-1"

Statistical Pattern Recognition

PROJECT REPORT

NEVIL PATEL-116897068

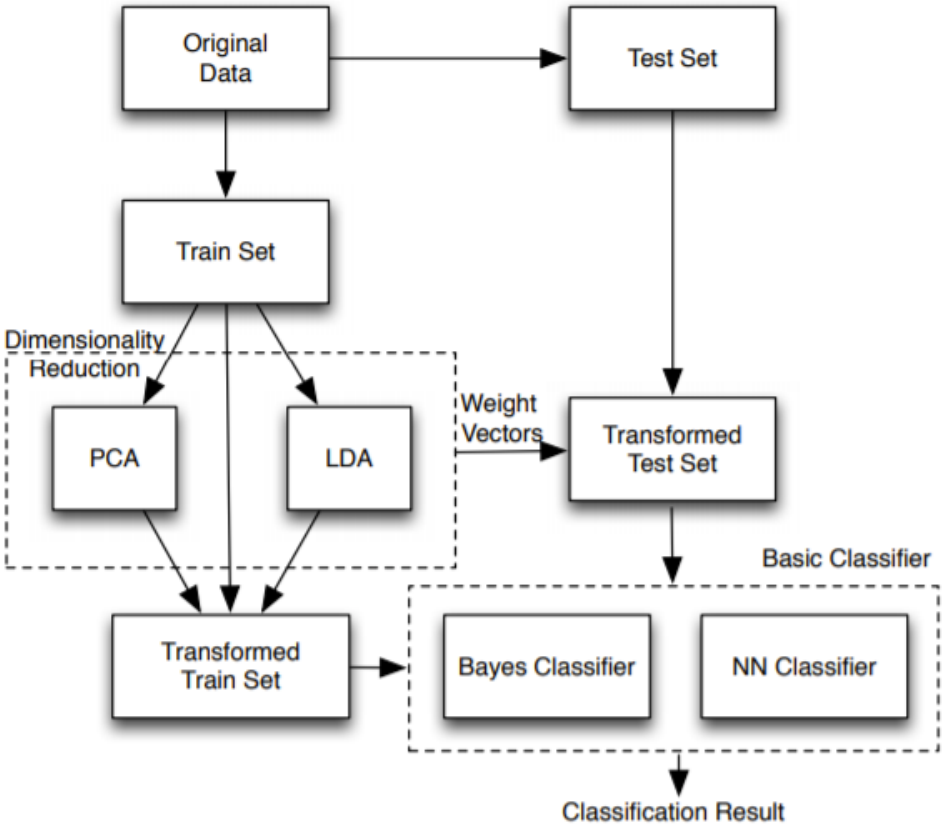
PROJECT INTRODUCTION:

This project serves as a research and evaluation for different popular techniques such as:

- ML estimation with Gaussian assumption followed by Bayes' classification.
- k-Nearest Neighbors (k-NN) rule.
- Kernel SVM and Boosted SVM.
- PCA followed by Bayes classifier and K-NN rule.
- MDA followed by Bayes classifier and K-NN rule.

Here different data sets were used to perform the different tasks such as binary classification and multiclass subject classification.

Code Flow:



DATA VISULIZATION FROM DATA.MAT

C1: NEUTRAL C2: SMILING FACE C3:ILLUMINATION

BINARY CLASSIFICATION (NEUTRAL VS. FACIAL EXPRESSION):

We have used the data from DATA.mat , The images were 21 by 24 pixels gray scale images and were converted into a column vector of length 504. Here we have ignored the illumination expression from data for this binary classification.

Bayesian Classification:

The data was partitioned in two classes namely, neutral class for neutral expression and expression class for smiling facial expression. the covariance matrix formed from a single class was found to be near singular. Hence, a simple regularization technique was used to remove the singularity of the matrix by adding small amount of noise to the diagonal elements of the matrix.

- After running Bayes classifier several times, we got the maximum accuracy of 87%.
- PCA captures the maximum variances in the data and selects good features for classification along the components which are orthogonal to each other, it is necessary to include only the components corresponding to the maximum Eigen values, so it does not collect bad features

and contaminates the data unnecessarily. At 75 components the variance captures was maximum and accuracy was 86%

- MDA is a very good transformation to achieve separation between classes on a dimension which minimizes the in-class variance and maximizes the class separation. The accuracy achieved by this method was 93.75% since now the data was more separable.

The results for the following Bayes classification along with applying PCA and MDA are:

NEUTRAL VS. FACIAL EXPRESSION CLASSIFICATION ACCURACY(%ac)

CASE	TRAINING(%)	TESTING(%)	BAYES(%ac)	BAYES+PCA(%ac)	BAYES+MDA(%ac)
1	50	50	87	89.5	90
2	20	80	84.0625	71.875	93.75
3	40	60	88.3333	71.6667	91.6667
4	60	40	86.8750	88.1250	87.5000
5	80	20	83.7500	85	81.2500
6	90	10	82.5000	87.5000	80

COMPARISION OF PCA COMPONENTS AND ACCURACY

Principal Components	**Accuracy**
55	84%
65	85.5%
75	86%
85	77%
95	66%

K – Nearest Neighbors:

K nearest neighbor algorithm was performed again on the above given dataset. This was performed at first on the original data and then using PCA and MDA methods. The accuracy was tested for different values of K. The results are as follows:

NEUTRAL VS. FACIAL EXPRESSION CLASSIFICATION ACCURACY(%ac)

K	KNN(%ac)	KNN+PCA(%ac)	KNN+MDA(%ac)
1	71	81	81
3	79	80	81
7	77	83	80
10	78	89	80
14	76	82	80

COMPARISION OF PCA COMPONENTS AND ACCURACY (K = 6)

Principal Components	Accuracy
5	83%
10	84%
15	88%
20	92%
25	90%

Here in DATA + PCA there is a growing trend in accuracy and then it minimizes the effect of adding more principal components. This captures approximately 97% of the variance from the data.

Support Vector Machines:

The SVM algorithm met all KKT conditions and was modelled very easily. Here the slack parameter was taken into consideration to generalize the linear inseparability of the data. From the comparison table given below we know that if the value of the slack is too low, the training error will be high and vice versa. The table below shows the comparison of accuracies for different values of slack parameters

NEUTRAL VS. FACIAL EXPRESSION CLASSIFICATION ACCURACY(%ac)

SLACK PARAMETER	SVM(%ac)
0.10	70.5
0.15	82
0.20	83
0.25	85.5
0.30	86
0.35	86

We can approximate slack value for our classification task as 0.30. We get the maximum accuracy and less training error at that range.

Kernel SVM Classifier:

The Kernel SVM is implemented by solving the Dual optimization using lagrangian form.
Radial Basis Function(RBF)-

σ^2	Accuracy(%ac)
1	75.5
0.01	72
0.02	76.5
0.06	75.5
0.001	72

MULTICLASS CLASSIFICATION:

Here the task of multiclass classification using data.mat was taken into account. The data was now partitioned in such a way that the training consisted on 2 image characteristic and testing was performed on 1 remaining image for each subject. Hence, the anticipated accuracy was already very low.

Bayesian Classification:

Here one point was evaluated with 200 Bayesian posterior probabilities and it was performed for 200 testing points. Hence it required a lot of computing power and also a lot of time because the evaluation was done 200x200 times. The table below shows the Results:

NO.	TRAINING SET	TESTING SET	Bayes(%ac)
1	NEUTRAL+EXPRESSION	ILLUMINATION	63.5%
2	NEUTRAL+ILLUMINATION	EXPRESSION	67%
3	EXPRESSION+ILLUMINATION	NEUTRAL	71.5%

Multiclass Support Vector Machines:

Here In order to create a classifier between many classes we can perform SVM with one and rest others classification task. We used the best option available as illumination.mat dataset. It contained 68 subjects with 21 different illumination variations in them. For each class, first 20 images were selected as the training data and only one image was selected as the testing image to make the experiment computationally fast.

The data would be considered only as classified if it is classified in X class (one) rather than any other class (all). To check X class is correct than other classes , only then it is correctly classified otherwise it its misclassified or simply ignored. Here the classification task was split into 2 sides one X class and all other vague data on other, it is highly likely that any random point to get classified correctly.

Here the number of points in our class is 20, which are perfect points. All other vague points and testing is done using only one point at a time, this method proves that the accuracy will be extremely high. Here the final accuracy was 100%.

Illumination and Pose Dataset Classification using KNN:

The illumination.mat data set is really a robust data set with good features for each subject which makes it easy for the classifiers to classifies test-points correctly. As evident from above the illumination data set performed accurately for SVM which suggests that there is good separability between the classes. On the other hand, the pose.mat data set performed poor for major classification models. This evaluation was performed for both data sets using the KNN classification techniques and this time for fair evaluation and increase the randomness the data set was not split into training and testing sets but rather every point was evaluated against every other point in the data set keeping the labels. This was easily possible in KNN.

Illumination data set:

We see from the data as the number of nearest neighbor increases the KNN performs bad because there are many tie brake and the KNN algorithm performs poorly under such conditions.

K	ACCURACY(%ac)
1	100
3	86.33
5	70.56
7	60.4765
9	54.59
11	51.787

Pose data set:

The same trend can be observed for the pose data set. It already started with a poor accuracy and it went on getting worse as the number of K increased.

K	ACCURACY(%ac)
1	58.3236
3	38.8448
5	36.9
7	35.7871
9	37.3726
11	37.1461

CONCLUSION:

Various techniques were finally evaluated and experimented with different transformations for different data sets with varying parameters and results are as follows. It could be said:

- Transformations such as MDA, PCA and kernel trick try to exploit linear separability in higher and lower dimensions and thus are more likely to improve the performance
- Classification problem performance is proportional to the dataset given e.g: Illumination set performed best for most classifiers
- To avoid overfitting of the data, training accuracy and error should be optimised.
- The Kernel SVM takes full advantages of the dimensionalities of the data.
- Bayesian, KNN and SVM classification followed by kernelizations and transformation techniques produces different and extraordinary results as per the dataset.

REFERENCES:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.721.9408&rep=rep1&type=pdf

https://pdfs.semanticscholar.org/50ad/3df1230607a1d39ba5c78d84b65c8b2e7edf.pdf