Analysis of EEG signals using classification methods

Aayushi Malik
2015003

Shreya Sharma
2015096

I. Problem Statement and Motivation

Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activity of the brain. It is typically noninvasive, with the electrodes placed along the scalp. It plays an important role in the domain of brain computer interface (BCI). Compared with other biomedical signals, the EEG is extremely difficult for an untrained observer to understand because of the spatial mapping of functions onto different regions of the brain and electrode placement. Besides, feature extraction and pattern recognition are crucial for data processing. So, we aim to classify the EEG patters by applying dimensionality transformation techniques like PCA and LDA.

II. Brief Literature Review

There have been many studies on EEG pattern analysis and classification. We combined the approaches used by these studies and compared the performance of various approaches. In [4] , the authors used two approaches: PCA+HMM and PCA+HMM+SVM to classify EEG samples.The authors in [1] also used PCA with HMM for classification. In [3], the authors compared the classification done by LDA, QDA and MD (Mahalanobis Distance) based classifiers. In [2], the authors applied the joint PCA-LDA method and the stepwise discriminant method.

III. Dataset Used

THERE are two datasets. One is EEG dataset with classes: Alcoholic, Non-alcoholic and the other is EGG dataset with classes: Seizure, Non-seizure.

A. Seizure vs Non-seizure

A csv file containing 11500(rows) samples which contain 178 elements which represent the brain activity over 1 second. Each of the 178 element is the value of the EEG recording and is treated as a variable or feature. The last column of a row is the class label. There are 5 labels = 1,2,3,4,5. All subjects falling in classes 2, 3, 4, and 5 are subjects who did not have epileptic seizure. Only subjects in class 1 have epileptic seizure.

B. Alcoholic vs Non-alcoholic

It contains EEG recordings for 122 subjects, each having 120 trials. The fourth letter of line 1 of each trial is the label ('a': alcoholic and 'c': control). In each trial, recordings have been taken from 64 electrodes placed at different positions. The four columns of data are: the trial number, sensor position, sample number (0-255), and sensor value (in micro volts).

IV. Proposed Architecture

The best architecture that can be used for this binary classification is using PCA for dimensionality reduction and classifying using Hidden Markov Model. The Hidden Markov Model had The following structure:

- Possible value of states: 0 and 1 (representing non-alcoholic/non-seizure and alcoholic/seizure respectively)
- Initial probabilities: These were calculated by simple counting. For both values of states, the total number of samples were counted.
- Transition probabilities: The matrix used for the transition probabilities was: [[1,0],[0,1]]
- Observation probabilities: The observations were discretized and then through counting, the transition probability for each value of the observation corresponding to each state value was calculated.

This is a good architecture as it provides good accuracy for both datasets (0.9282 for Seizure dataset and 0.5895 for Alcoholic dataset). The reason why this architecture works well with the data we have used is because EEG patterns are a time series data and is a perfect fit for HMM as HMM works well for sequential observations.. Using PCA for dimensionality reduction and Naive Bayes for classification provides better accuracy in the Seizure dataset than using HMM (0.9507). However, it performs worse than random for the Alcoholic dataset (0.4716). PCA is used for dimensionality reduction as LDA is preferrable to PCA when the number of samples and the number of classes is large. In our case, despite the number of samples being large, the number of classes was only two so LDA did not exhibit optimal results. The training time on using LDA was very large (on using our implemented version as well as using an existing implementation) and the accuracy was only between 0.5 and 0.6.

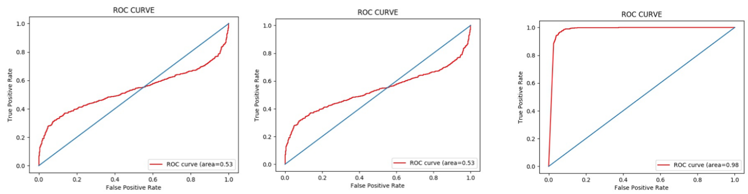
V. Results

The results can be seen from the graphs and tables.

VI. Analysis of Results

A. Without PCA

Without applying PCA on the datasets we can see the Naive Bayes performs the best. Even though it's accuracy is less



(a) LDA (b) Neural Net (c) Naive Baye's

Fig. 1. ROC curves of multiple classifiers applied on Seizure dataset without PCA

TABLE I
SEIZURE : CONFUSION MATRICES OF CLASSIFIERS WITHOUT PCA

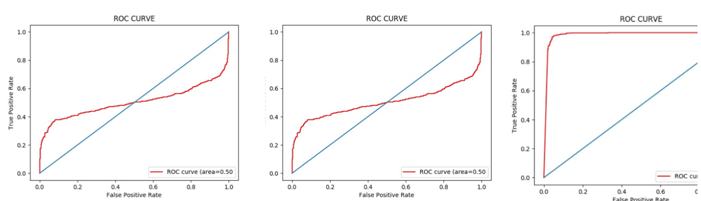
	NAIVE BAYES	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.500	fp = 0.017
Actual: 0	fn = 0.043	tp = 0.439
Accuracy	0.9391	
	LDA CLASSIFIER	
	Predicted: 1	Predicted : 0
Actual: 1	tn = 0.356	fp = 0.162
Actual: 0	fn = 0.263	tp = 0.220
Accuracy	0.5754	
	NEURAL NETWORKS	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.499	fp = 0.019
Actual: 0	fn = 0.054	tp = 0.429
Accuracy	0.9275	
	HIDDEN MARKOV MODEL	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.800	fp = 0.200
Actual: 0	fn = 0.018	tp = 0.982
Accuracy	0.8942	

TABLE II
SEIZURE : CONFUSION MATRICES OF CLASSIFIERS WITH PCA

	NAIVE BAYES	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.501	fp = 0.017
Actual: 0	fn = 0.033	tp = 0.450
Accuracy	0.9507	
	LDA CLASSIFIER	
	Predicted: 1	Predicted : 0
Actual: 1	tn = 0.195	fp = 0.322
Actual: 0	fn = 0.227	tp = 0.256
Accuracy	0.4507	
	NEURAL NETWORKS	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.491	fp = 0.027
Actual: 0	fn = 0.075	tp = 0.408
Accuracy	0.8985	
	HIDDEN MARKOV MODEL	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.881	fp = 0.119
Actual: 0	fn = 0.028	tp = 0.972
Accuracy	0.9282	

TABLE III
SEIZURE : CONFUSION MATRICES OF LDA CLASSIFIER (IMPLEMENTED)

	LDA CLASSIFIER (implemented)	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.423	fp = 0.094
Actual: 0	fn = 0.191	tp = 0.292
Accuracy	0.7152	
	LDA CLASSIFIER (implemented) on PCA	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.260	fp = 0.257
Actual: 0	fn = 0.207	tp = 0.292
Accuracy	0.5362	



(a) LDA (b) Neural Net (c) Naive Baye's

Fig. 2. ROC curves of multiple classifiers applied on Seizure dataset with PCA

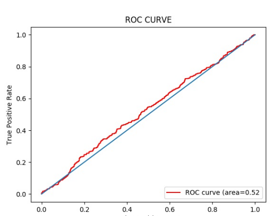
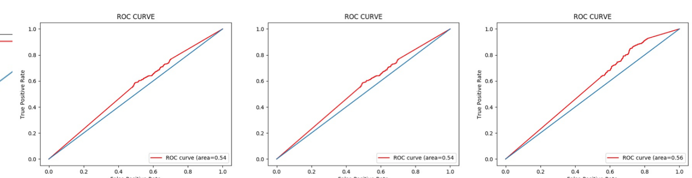
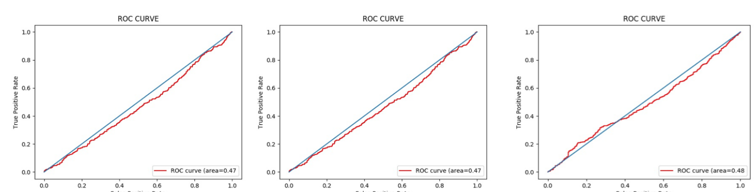


Fig. 3. ROC curves of implemented LDA classifiers applied on Seizure dataset with PCA



(a) LDA (b) Neural Net (c) Naive Baye's

Fig. 4. ROC curves of multiple classifiers applied on Alcoholic dataset without PCA



(a) LDA (b) Neural Net (c) Naive Baye's

Fig. 5. ROC curves of multiple classifiers applied on Alcoholic dataset with PCA

than that of LDA classifier, the difference is insignificant. It is not surprising as Naive Bayes gives the least error out of all classifiers in general. This is because the Bayes error is irreducible error.

We can also see that Neural Network performs nearly as good as Naive Bayes on the smaller dataset (seizure) but not so much with the larger dataset (alcoholic). This may be

TABLE IV
ALCOHOLIC : CONFUSION MATRICES OF CLASSIFIERS WITHOUT PCA

	NAIVE BAYES	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.224	fp = 0.294
Actual: 0	fn = 0.172	tp = 0.309
Accuracy	0.5333	
	LDA CLASSIFIER	
	Predicted: 1	Predicted : 0
Actual: 1	tn = 0.265	fp = 0.253
Actual: 0	fn = 0.208	tp = 0.273
Accuracy	0.5383	
	NEURAL NETWORKS	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.245	fp = 0.273
Actual: 0	fn = 0.203	tp = 0.278
Accuracy	0.5233	
	HIDDEN MARKOV MODEL	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.149	fp = 0.851
Actual: 0	fn = 0.173	tp = 0.827
Accuracy	0.3984	

TABLE V
ALCOHOLIC : CONFUSION MATRICES OF CLASSIFIERS WITH PCA

	NAIVE BAYES	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.215	fp = 0.303
Actual: 0	fn = 0.225	tp = 0.257
Accuracy	0.4716	
	LDA CLASSIFIER	
	Predicted: 1	Predicted : 0
Actual: 1	tn = 0.232	fp = 0.287
Actual: 0	fn = 0.242	tp = 0.240
Accuracy	0.4716	
	NEURAL NETWORKS	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.241	fp = 0.278
Actual: 0	fn = 0.241	tp = 0.241
Accuracy	0.4816	
	HIDDEN MARKOV MODEL	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.843	fp = 0.156
Actual: 0	fn = 0.858	tp = 0.163
Accuracy	0.5895	

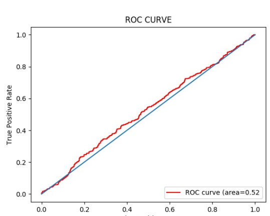


Fig. 6. ROC curves of implemented LDA classifiers applied on Alcoholic dataset with PCA

because the number of features of alcoholic (1684) is larger than seizure's (178) and this is true for all the classifiers. The training and testing time is also larger when the feature size increases. LDA (both library and implemented version) performs very poorly compared to other classifiers. This may be due to the fact that LDA reduces the the dimensions of both the classes to 1 in this case as it is a binary classification

TABLE VI
ALCOHOLIC : CONFUSION MATRICES OF LDA CLASSIFIER (IMPLEMENTED)

	LDA CLASSIFIER (implemented)	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.423	fp = 0.094
Actual: 0	fn = 0.191	tp = 0.292
Accuracy	0.7152	
	LDA CLASSIFIER (implemented) on PCA	
	Predicted: 1	Predicted: 0
Actual: 1	tn = 0.260	fp = 0.257
Actual: 0	fn = 0.207	tp = 0.292
Accuracy	0.5362	

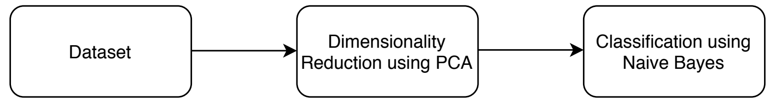


Fig. 7. The two best proposed architectures

problem and the dataset is such that it is not able to increase the separation between the classes. HMM performs reasonably good on the smaller dataset but gives the least accuracy with the larger one.

B. With PCA

Applying PCA on both the datasets resulted into a decrease of the accuracy of all the classifiers except HMM classifier. This is justified above. PCA in a way reduces the feature space by transforming it and then selecting the top principle components. This does lead to information loss. This loss may have been harmful for all the classifiers and may not have affected HMM as the data and HMM both are series based.

C. Implemented LDA

The implemented version of LDA classifier gives more accuracy than the library one. The implemented version classifies a point based on the difference between the mean of the transformed classes and the point. The library version fits a Gaussian distribution to each class and assumes that both the classes have same covariance matrix. Implemented version calculates the covariance matrices for both the classes and transforms the test poi

VII. INDIVIDUAL CONTRIBUTIONS

REFERENCES

[1] A.O. Argunsah and M. Cetin. AR-PCA-HMM Approach for Sensorimotor Task Classification in EEG-based Brain-Computer Interfaces. 2010 International Conference on Pattern Recognition.

[2] N.S. Dias, M. Kamrunnahar, P.M. Mendes, S.J. Schiff and J.H. Correia. Comparison of EEG pattern classification methods for brain-computer interfaces. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:2540-3.

[3] O.D. Eva and A.M. Lazar. Comparison of Classifiers and Statistical Analysis for EEG Signals Used in Brain Computer Interface Motor Task Paradigm. (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.1, 2015.

[4] H. Lee and S. Choi. PCA+HMM+SVM for EEG Pattern Classification. Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.