Prediction of the onset of epileptic seizures from iEEG data
CS229 Final Report

Shima Alizadeh, Scott Davidson, and Ari Frankel

Seizures in epileptic patients can be anxiety-inducing, and the resulting medical and social issues can cause distress to a patient. A predictive mechanism to anticipate the onset of a seizure could help a patient prepare for a seizure and take medication. The literature has shown that the onset of a seizure is directly correlated with a distinct neurological change that can be detected using iEEG measurements. Thus the development of a classification model based on IEEG data can be created to aid epileptic patients. In this project we develop a feature set based on spectral and statistical features of the IEEG data, with and without ICA pre-processing of the data. Several classication algorithms were trained on the resulting data sets, including Naive-Bayes, SVM, logistic regression, and lasso-regularized logistic regression. It was found that the optimal results were given for regularized logistic regression on PIB features without ICA pre-processing, though generally all algorithms performed better than a random naive predictor model.

1. Introduction

1.1. Motivation

Epilepsy affects nearly 1% of the population, and is characterized by the sporadic onset of violent seizures. Due to the health-threatening nature of the seizures as well as the social embarassment induced, patients feel a constant anxiety that a seizure may happen at any moment. Medicine is available that can reduce the frequency of seizures; however, the side effects associated with long-term constant medication can also cause discomfort to the patient. The ability to predict the imminent onset of a seizure is thus attractive in that a predictive system can alert an epileptic patient and give him time to take medication as necessary and avoid the side effects.

Epilepsy is a neurological condition and is sporadic in nature, meaning that most of the time an epileptic patient exhibits no symptoms at all. This suggests that the condition may be associated with a distinct physiological change in the brain that occurs at the time of a seizure. Recent studies support this hypothesis and show that the brain activity of an epileptic patient may be classified into four stages: interictal (normal brain activity), preictal (prior to seizure), ictal (during seizure), and postictal (immediately following the seizure) [1]. It has been demonstrated that these four stages may be associated directly with voltage readings in the brain from intracranial electroencephalogram (iEEG) measurements at different locations, and the preictal stage can be detected with high sensitivity. This suggests that iEEG recordings can be used to detect preictal brain activity and serve as a predictive tool for seizure onset [3].

1.2. Problem Statement

In this project we aim to develop a reliable learning model for seizure forecasting based on long-term, continuous, iEEG records containing multiple seizure events. The data which is provided by Kaggle.com, was recorded from 5 dogs and 2 human patients with epilepsy. The iEEG record from the dogs were sampled from 16 electrodes (channels) at 400 Hz, and recorded voltages were referenced to the group average. In addition, datasets from human patients with epilepsy were obtained by using varying numbers of channels and sampled at 5000 Hz. Each training/testing example is a set of time series of voltages measured at different locations of the brain using a number of channels, and each example labeled as interictal (normal brain activity) or preictal (immediately prior to seizure). Using the provided training examples, we exploit different machine learning algorithms to build a model which can distinguish between the iEEG records of pre-ictal with iEEG clips of interictal activity with sufficient accuracy.

To gain a better insight about the efficacy of the features used for this classification problem we decided to create different sets of possible features which have been frequently used in related studies. The feature sets include multiple spectral iEEG power bands (PIB), variance, and time correlation of iEEG records. The extraction of each feature set, their types and properties and also the number of extracted features are described in section (2). Moreover, a variety of learning algorithms such as logistic regression, Naive-Bayes, and SVM have been utilized for learning different sets of features. At the end of this project, the outputs and the performance of each learning method will be evaluated by looking at the test and training errors as well as taking advantage of various error metrics such as the precision, recall, and AUCROC statistics.

2. Pre-processing and Feature Extraction

2.1. Independent Component Analysis

The data received from an iEEG measurement can be thought of as a linear combination of independent signals in the brain, each containing a unique set of frequency characteristics. While some frequency characteristics can indicate important parts of brain activity, some frequency characteristics may be associated with lower level functions (e.g. breathing). Thus independent component analysis was used in order to separate out the signals and thus potentially elicit the features that change the most between interictal and preictal brain activity.

Since the entire data set consists of multiple subjects in multiple sessions, it was important to train the unmixing matrix on individual subjects. To reduce potential bias in the training, a random mixture of data was taken from multiple time segments, half from the interictal and half from the preictal parts of the data. The unmixing matrix was derived using the ICA capabilities of the EEGLAB library for MATLAB.

2.2. Power-in-band (PIB) features

The changes in spectral power of multichannel iEEG in certain frequency bands have been demonstrated to be good indicatiors to distinguish between pre-ictal and inter-ictal states of the brain activity. High sensitivity and specificity can be achieved by developing nonlinear classifiers which are trained based on linear features of spectral power of iEEG[1],[3]. Moreover, the rapid computation of the spectral power with low power consumption is very advantageous in manufacturing implantable predicting devices. To extract features using spectral power of iEEG records from dogs with natural epilepsy, in each of the 16 iEEG channels, the iEEG record was partitioned into non-overlapping 1-minute blocks, each block Fourier transformed, and the resulting power spectrum was divided into 6 frequency bands: (0.1–4 Hz), (4–8 Hz), (8–12 Hz), (12–30 Hz), (30–70 Hz), (70–180 Hz). Within each frequency band the power was integrated over band frequencies to produce a power-in-band (PIB) feature. Each of the 6 PIB values from a given channel and time block was summed with the corresponding PIBs from other 1-minute blocks in the same channel. Repeating this procedure on each iEEG channel eventually gave $6 * 16 = 96$ PIB values as the elements of the feature vector. The PIB features were calculated for both pre-ictal and inter-ictal data to form the training and testing data sets for the learning algorithms.

2.3. Variance and Correlation Features

Two statistical measures of a signal, variance, and correlation coefficient were other possible options to represent the features of the problem. These form a relatively simple to compute set of features which provide measures of the power in each channel of the data and the relationships between the signals in different channels. Epilepsy is known to cause increased brain activity in regions of the brain where the seizure is centered [1]. By measuring the variance of the signals of all of the channels, we anticipate that certain channel's variances will change in a consistent manner during seizures thus which our learning algorithm will be able to detect. Measuring the signal variance results in one feature per iEEG channel, thus for the 16 channel records, 16 features are formed from the variances. Another way we detect the presence of these spatially consistent brain waves is through the correlation coefficient matrix which provides a measure of the characteristic correlation of signals detected at specific electrodes in advance of seizures.

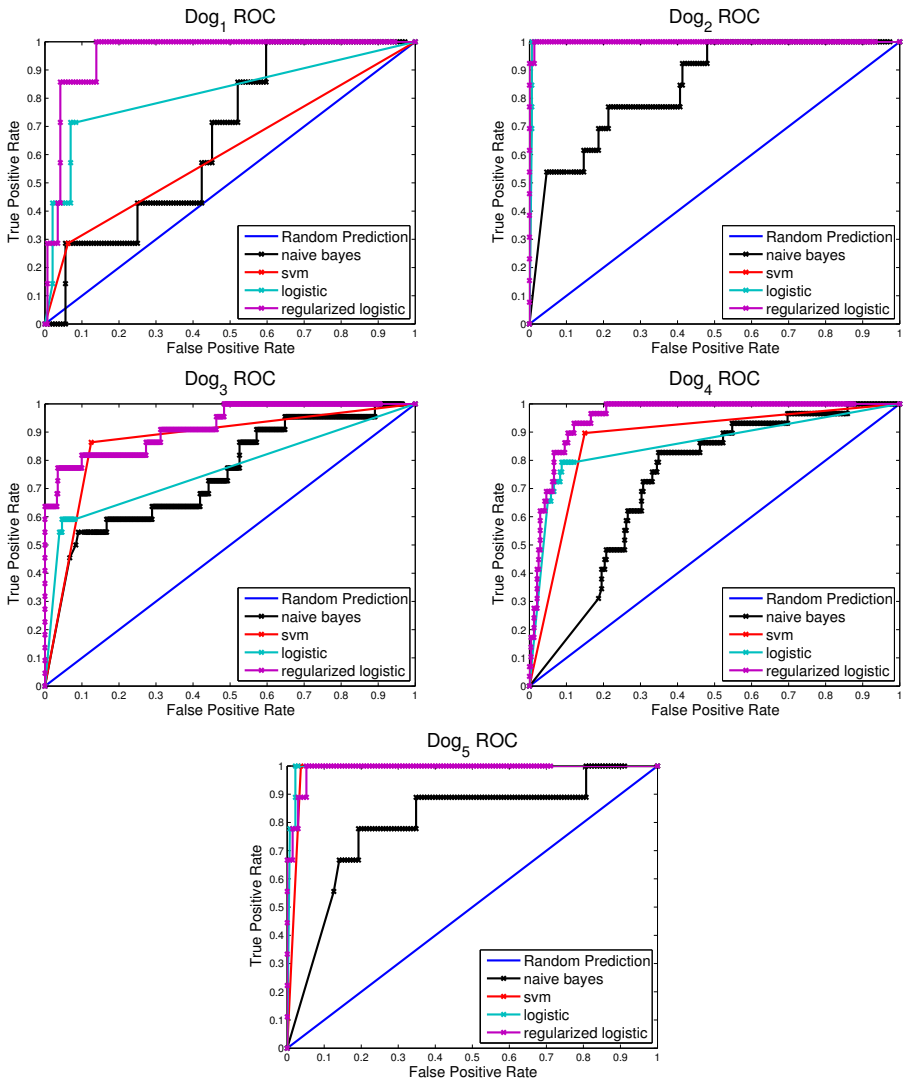

Figure 1. The ROC curves for each of the dogs using PIB features without ICA

The correlation coefficient matrix, defined as $R_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$ where C_{ij} is the covariance matrix is symmetric with ones on the diagonal, thus only the upper triangular portion of the matrix not including the diagonal is used for features. For 16 channel records, this gives 120 unique features.

3. Learning Algorithms and Performance

In order to classify the data as interictal or preictal, multiple classification algorithms were trained on the data and tested using holdout cross-validation with 30% of the data. The algorithms used were Naive-Bayes, SVM, logistic regression, and lasso-regularized logistic regression.

Since different epileptic patients have different seizure characteristics [2], each classifier had to be trained on each subject individually. The ROC curves for each of the dogs and each of the classifiers is shown in figure 1. Generally, the lasso-regularized logistic regression classifier outperforms each of the other classifiers, and Naive-Bayes has the worst performance.

AUCROC	PIB, no ICA				PIB, ICA			
	NB	SVM	Logistic	Lasso logistic	NB	SVM	Logistic	Lasso logistic
Dog 1	0.664	0.612	0.818	0.955	0.780	0.615	0.615	0.798
Dog 2	0.845	0.997	0.996	0.999	0.765	1.000	1.000	1.000
Dog 3	0.751	0.869	0.766	0.921	0.735	0.921	0.868	0.971
Dog 4	0.728	0.874	0.858	0.954	0.753	0.897	0.820	0.951
Dog 5	0.800	0.981	0.992	0.989	0.776	0.993	0.939	0.994

Table 1. AUCROC statistics for the dogs using PIB features, with and without ICA pre-processing

Precision	PIB, no ICA				PIB, ICA			
	NB	SVM	Logistic	Lasso logistic	NB	SVM	Logistic	Lasso logistic
Dog 1	0.571	0.286	0.429	0.286	1.000	0.286	0.143	0.286
Dog 2	0.923	1.000	0.923	0.923	1.000	1.000	1.000	1.000
Dog 3	0.727	0.864	0.591	0.636	0.773	0.955	0.773	0.773
Dog 4	0.828	0.897	0.724	0.828	0.828	0.931	0.621	0.724
Dog 5	0.778	1.000	0.778	0.889	0.778	1.000	0.778	0.889

Table 2. Precision statistics for the dogs using PIB features, with and without ICA pre-processing

Recall	PIB, no ICA				PIB, ICA			
	NB	SVM	Logistic	Lasso logistic	NB	SVM	Logistic	Lasso logistic
Dog 1	0.060	0.182	0.333	0.333	0.101	0.200	0.143	0.286
Dog 2	0.152	0.929	0.923	0.923	0.137	1.000	1.000	1.000
Dog 3	0.072	0.260	0.394	0.636	0.072	0.300	0.515	0.773
Dog 4	0.194	0.419	0.553	0.585	0.220	0.450	0.529	0.677
Dog 5	0.184	0.643	0.778	0.667	0.132	0.818	0.778	0.667

Table 3. Recall statistics for the dogs using PIB features, with and without ICA pre-processing

AUCROC	Covariance, no ICA				Covariance, ICA			
	NB	SVM	Logistic	Lasso logistic	NB	SVM	Logistic	Lasso logistic
Dog 1	0.654	0.687	0.793	0.809	0.737	0.748	0.880	0.871
Dog 2	0.868	0.990	0.996	0.991	0.888	0.990	0.991	0.998
Dog 3	0.680	0.888	0.930	0.964	0.650	0.907	0.844	0.956
Dog 4	0.716	0.813	0.899	0.895	0.766	0.796	0.758	0.850
Dog 5	0.713	0.863	0.825	0.980	0.800	0.974	0.875	0.987

Table 4. AUCROC statistics for the dogs using covariance features, with and without ICA pre-processing

Precision	Covariance, no ICA				Covariance, ICA			
	NB	SVM	Logistic	Lasso logistic	NB	SVM	Logistic	Lasso logistic
Dog 1	0.571	0.429	0.286	0.143	0.571	0.571	0.286	0.286
Dog 2	0.923	1.000	1.000	0.846	0.923	1.000	1.000	1.000
Dog 3	0.909	0.864	0.682	0.727	0.818	0.909	0.682	0.636
Dog 4	0.483	0.759	0.690	0.690	0.483	0.724	0.552	0.586
Dog 5	0.778	0.778	0.667	0.667	0.778	1.000	0.667	0.667

Table 5. Precision statistics for the dogs using covariance features, with and without ICA pre-processing

Recall	Covariance, no ICA				Covariance, ICA			
	NB	SVM	Logistic	Lasso logistic	NB	SVM	Logistic	Lasso logistic
Dog 1	0.070	0.273	0.182	0.200	0.069	0.267	0.333	0.400
Dog 2	0.188	0.813	0.867	0.917	0.211	0.813	0.722	0.867
Dog 3	0.068	0.333	0.469	0.571	0.062	0.328	0.500	0.500
Dog 4	0.187	0.407	0.571	0.606	0.192	0.396	0.533	0.425
Dog 5	0.109	0.500	0.857	0.857	0.146	0.563	0.667	0.750

Table 6. Recall statistics for the dogs using covariance features, with and without ICA pre-processing

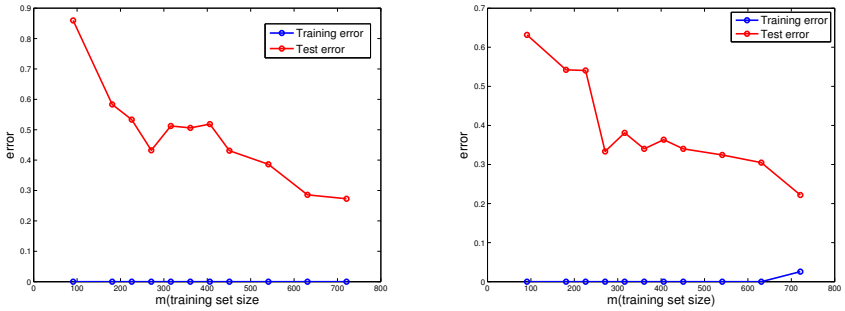

Figure 2. Learning curves for one of the dogs for logistic regression (left) and lasso-regularized logistic regression (right).

In order to estimate the effects of bias and variance, a learning curve was generated for one of the dog subjects using logistic regression and lasso-regularized logistic regression. The resulting plots are shown below in figure 2. For each of the cases the training error is nearly zero for all set sizes, while the testing error is quite large and decreases steadily as a function of training set size. In addition, the error decreases in the lasso-regularized case. Since lasso-regularization tends to set feature parameters to zero, this suggests that the large feature set is overfitting the data, and that either a larger data set or a smaller feature set would help improve the accuracy of a classifier.

4. Conclusions and Future Work

We have demonstrated the use of machine learning algorithms for the classification of iEEG data as interictal or preictal with a relatively high success rate. Each of the learning algorithms tested performed better than a random naive predictor, where lasso-regularized logistic regression performed the best and Naive-Bayes performed the worst. The feature sets tested included PIB features and covariance features, both with and without ICA pre-processing for signal separation. Generally the results showed no improvement with the use of the current ICA pre-processing. Furthermore, the PIB features tended to give lower testing error than covariance features. This suggests that the power spectral density of the iEEG data is more sensitive to differences in preictal and interictal brain activity than the correlation features. The learning curves also suggest that there is room for improvement, however, in that the current feature set is overfitting the data and either more data or fewer features are needed to improve the accuracy.

A number of other research directions could be included in future work. The large imbalance in the number of preictal data segments versus interictal data segments suggests that an approach to weight the preictal data more heavily than the interictal data could help improve the accuracy of a classifier. In addition, other measures to filter the data and reduce noise could help (e.g. using a Kalman filter on the raw iEEG data).

Using Discrete Wavelet Transform (DWT) is another option for extracting features from the frequency bands which carry more information about the preictal or interictal nature of the iEEG signals than the other bands. As a

future work, we plan to utilize the features obtained by DWT to generate new predictive models and compare their accuracies and properties with the developed ones.

While we explored the effects of using lasso-regularization (L1-regularization) of the logistic regression classifier, a ridge-regularization (L2-regularization) could also have been used to reduce the magnitude of the feature parameters. And finally, the overfitting problem should be handled by using a model searching method to reduce the number of extraneous features, for example by using a forward or backward search on the information content of a particular model.

References

[1] Howbert JJ, Patterson EE, Stead SM, Brinkmann B, Vasoli V, Crepeau D, Vite CH, Sturges B, Ruedebusch V, Mavoori J, Leyde K, Sheffield WD, Litt B, Worrell GA (2014) Forecasting seizures in dogs with naturally occurring epilepsy. PLoS One 9(1):e81920

[2] Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, D'Souza W, Yerra R, Archer J, Litewka L, Hosking S, Lightfoot P, Ruedebusch V, Sheffield WD, Snyder D, Leyde K, Himes D (2013) Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. LANCET NEUROL 12:563-571.

[3] Park Y, Luo L, Parhi KK, Netoff T (2011) Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. Epilepsia 52:1761-1770