Seizure Prediction With Maximum Mutual Information Based Feature Selection

William Boxx, Johns Hopkins University
EN 525.776 Information Theory

Motivation

Seizure forecasting systems have the potential to help patients with epilepsy lead more normal lives. In order for EEG-based seizure forecasting systems to work effectively, computational algorithms must reliably identify periods of increased probability of seizure occurrence. If these seizure-permissive brain states can be identified, devices designed to warn patients of impeding seizures would be possible. Patients could avoid potentially dangerous activities like driving or swimming, and medications could be administered only when needed to prevent impending seizures, reducing overall side effects.

There is emerging evidence that the temporal dynamics of brain activity can be classified into 4 states: **Interictal** (between seizures, or baseline), **Preictal** (prior to seizure), Ictal (seizure), and Post-ictal (after seizures). Seizure forecasting requires the ability to reliably identify a preictal state that can be differentiated from the interictal, ictal, and postictal state. The primary challenge in seizure forecasting is differentiating between the preictal and interictal states. The goal of this project is to demonstrate the existence and accurate classification of the preictal brain state in dogs with naturally occurring epilepsy [1].

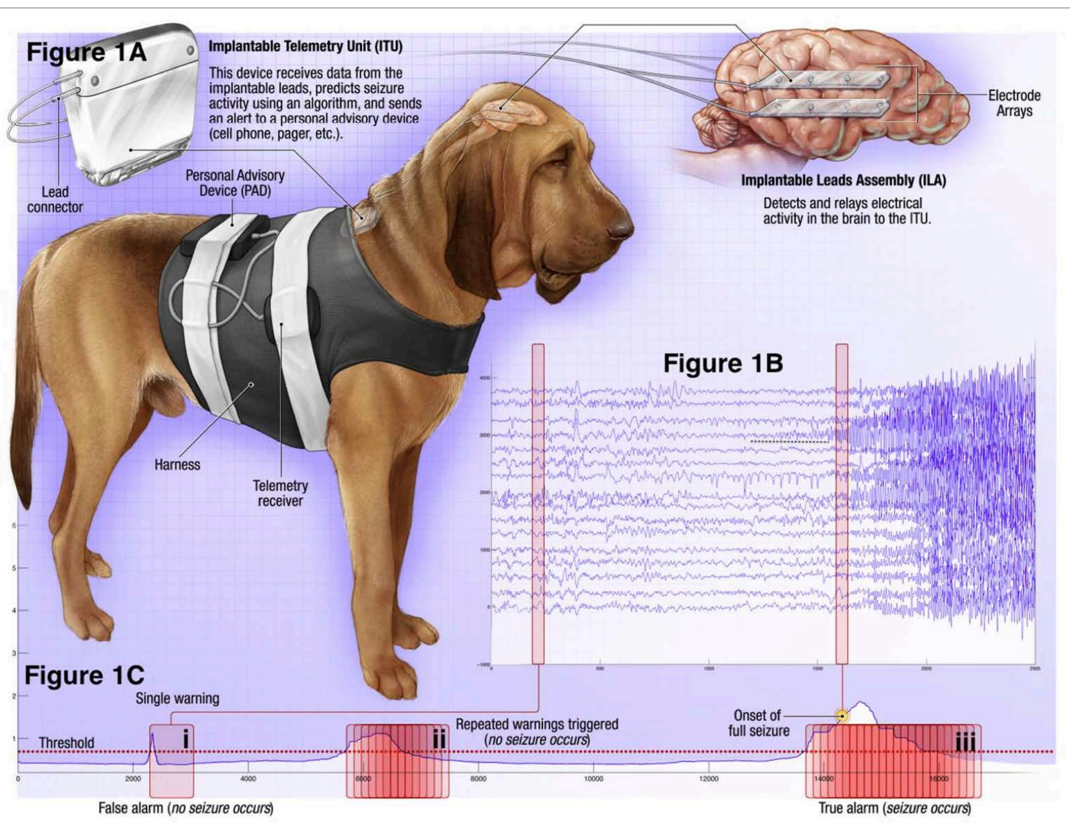


Fig. 1: Dog Recording System

Analysis

Datasets

Intracranial EEG was recorded from dogs with naturally occurring epilepsy using an ambulatory monitoring system (see Fig. 1). EEG was sampled from 13 electrodes at 400 Hz, and recorded voltages were referenced to the group average. Three dogs were included in the test set for this project. The test sets for each dog included a number of interictal segments and a number of preictal segments. Each segment contained 10 minutes of data for the 13 electrodes. The data was separated into 90% training data and 10% testing data for each dog.

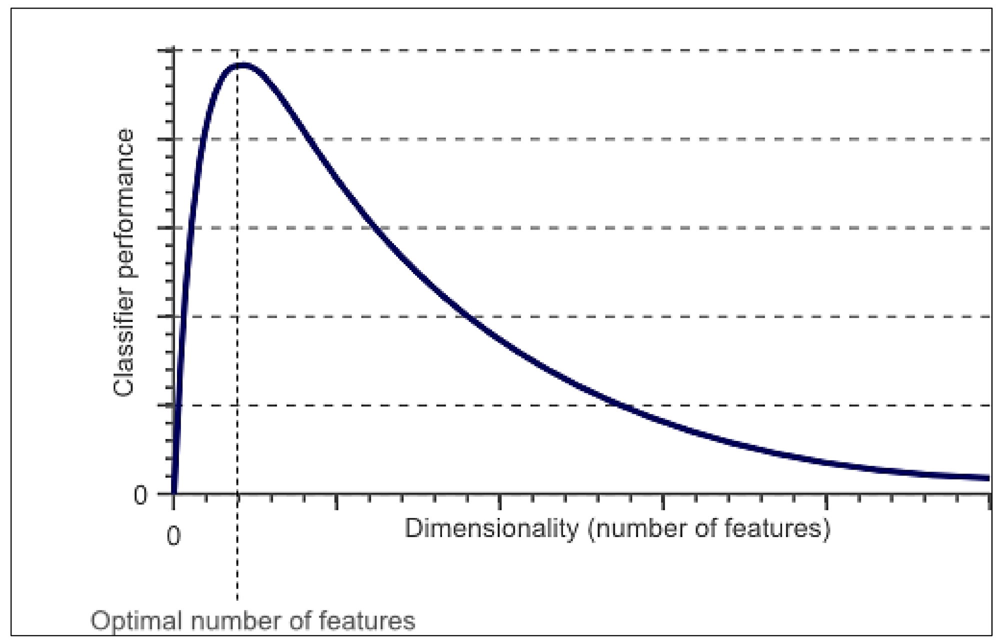


Fig. 2: Dimensionality Vs. Performance

Curse of Dimensionality

The performance of our classifier peaks at a certain dimensionality (see Fig. 2). As the dimensions in our feature set increase past this peak, the curse of dimensionality takes over. In 1 dimension, we can model a data set using a histogram with a certain number of bins each with a constant number of points. The total number of points represented by N^1. If we wish to maintain this precision as we increase the dimensionality, D, we will need N^D points [2].

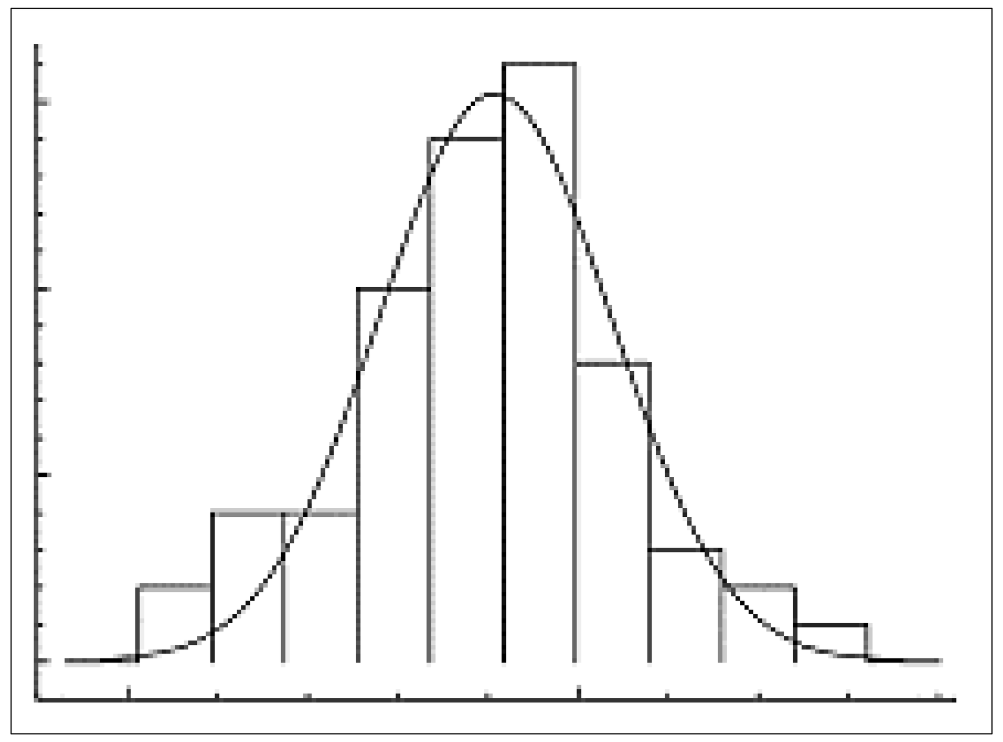
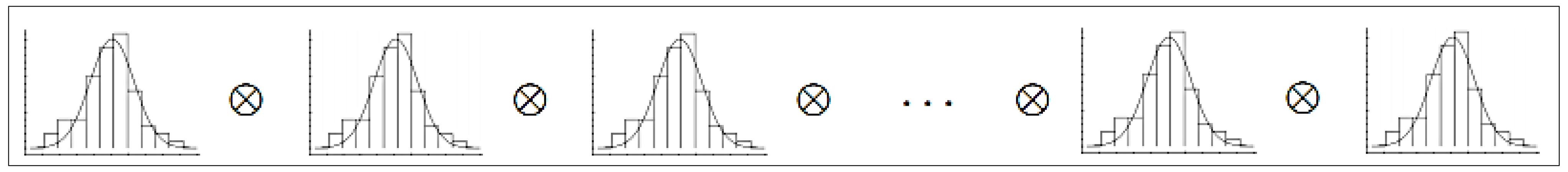


Fig. 3: 1 Dimensional Histogram



Feature Selection

In order to reduce overfitting due to the curse of dimensionality, we need to develop a method to remove features, or in other terms, pick the best features. One of way of doing this is by picking the features that maximize the mutual information between predictor and response. In general, the conditional entropy of our response and predictor will be less than the predictor itself according to Eq. 1 [3]. Therefore, in order to minimize the loss due to conditional entropy, we must maximize the mutual information.

$$I(P;R) = H(P) - H(P|R) \qquad \text{Eq. 1}$$

Results

Classification

The classification method used for this problem was linear regression due to its speed and resiliency to overfitting. Using a fast classification method allowed for a test and retest method to determine the optimal amount of features. It was found that 20 features produced the best results. The full feature set consisting of 167 features was fed to a selection algorithm that ordered them based on the maximization of mutual information technique. The linear regression classifier was then trained against the training data using these feature sets. Finally a prediction was made using the trained classifier on new data. The new data was extracted in the same way as the training data and the feature sets used were the same. The results were plotted using a Receiver Operating Characteristic curve. The results for dog 1 are shown in Fig. 4 achieving an Area Under the Curve (AUC) score of 0.7441.

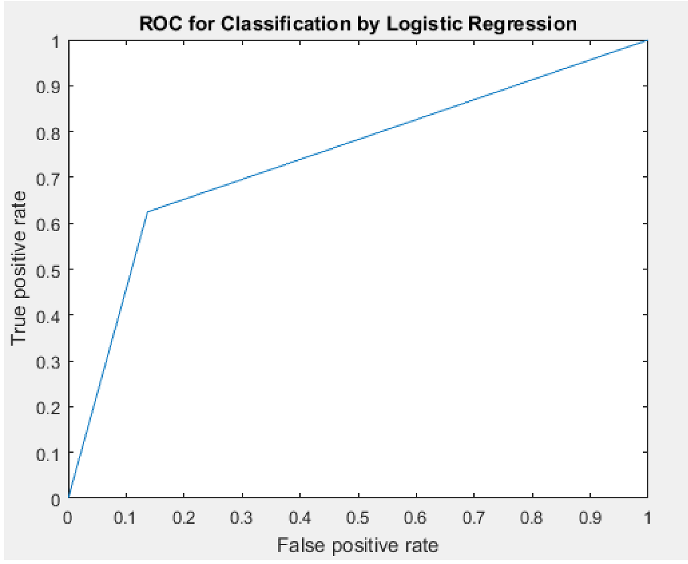


Fig. 4: ROC Curve Dog 1

Conclusion

Seizure forecasting systems have the ability to help patients of epilepsy live more normal lives. It is with this motivation that machine learning techniques are being developed to detect the onset of a seizure thus allowing preventative care. One of the challenges in machine learning problem such as this, is the ease of overfitting the data due to the curse of dimensionality. Thus, proper techniques such as maximization of mutual information must be used to optimize the dataset.

Acknowledgements and Citations

[1]"American Epilepsy Society Seizure Prediction Challenge | Kaggle", *Kaggle.com*, 2016. [Online]. Available: https://www.kaggle.com/c/seizure-prediction. [Accessed: 22- Apr- 2016].

[2]M. Maggioni, "Approximation of High Dimensional Dynamical Systems", B.I.R.S., 2015.

[3]R.Battiti, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning," *IEEE Transactions on Neural Networks*, vol. 5, No. 4, pp. 537-550, July 1994

American Epilepsy Society · Epilepsy Foundation · NIH National Institute of Neurological Disorders and Stroke