Dimension Reduction in fMRI Imaging Data

Kannu Mehta, David (Wei) Jia

August 8, 2013

1 Introduction

The development of functional Magnetic Resonance Imaging (fMRI) technology has been critical in allowing computer scientists and neuroscience researchers to analyze brain activity and inference about cognitive states. However, for many reasons, such as limitations to artificial intelligence and computer vision, this has proven to be difficult. In this project, we hope to improve upon existing techniques and shed light on the growing field of fMRI-based brain research.

1.1 Motivation

Existing studies have been able to use machine learning techniques with some success to classify fMRI imaging data and distinguish different cognitive states such as whether a human subject is viewing a sentence or picture [1]. However, their efforts were focused on high dimensional data with a very small set of training samples. Although their research proved to have decent initial results, a large set of training data is required in order to effectively train classifiers on high dimensional data. Because higher dimensional data usually results smaller number of samples, generating a large amount of sample data is difficult. Because the dimensionality of features are usually directly related to the dimensionality of sample data, having a small number of sample data also leads to problems such as over-fitting. One way to solve this problem is instead to reduce the dimensionality of the sample data to make up for the lower number of samples.

1.2 Problem Definition

Conventional classifiers like Gaussian Naïve Bayes (GNB), Support Vector Machines (SVM), Logistic Regression (LR) perform well when the data has low dimensional inputs and large number of training examples. However, as the dimensions of the input features increase these classifiers tend to suffer from the "curse of dimensionality." In other words, as dimensionality increases, the amount of available data tends to decrease. However, to produce any statistically significant results, higher dimension data require an even high number of samples, which only exacerbates the original problem. This is particularly true for brain scan images as the number of voxels per image is large, typically in the range of 1,000 to 5,000.

Thus, a way to reduce the dimensionality of data while retaining the effectiveness of results would be highly beneficial. In this paper, we will apply various methods, such as Principal Component Analysis (PCA) and Zero Component Analysis (ZCA) whitening on classification errors in an effort to achieve this goal.

2 Proposed Research

Our goal is to efficiently and effectively train a classifier on high dimensional data. In order to do so, a very large training set is needed to improve generalization and reduce over-fitting on the training data. In some studies, researchers have adopted the feature selection techniques in order to deal with such high dimensional data from brain scans [1]. In this project, we plan to explore the dimensionality reduction techniques to reduce the dimensions of the input features.

2.1 Hypothesis

In our project, we hypothesize that a reduction of dimension in analysis of fMRI data can lead to similar or better results than that of higher dimensions.

2.2 Method

Conventional implementations of dimension reduction techniques such as PCA require memory that grows polynomially with respect to the number of input dimensions. Hence, we implement PCA which has $O(N^2)$ space complexity, where N is the number of training examples, and can efficiently find the eigenvectors.

In addition, we also plan to observe the effects of PCA and ZCA whitening on the classification error. Further, that GNBs assume that the dimensions are conditionally independent or uncorrelated is well known. But this assumption may not hold true for the fMRI images because adjacent pixels or patches may be highly correlated. Hence, we will also try the Discrete Cosine Transform (DCT) in order to de-correlate the dimensions and comply with the initial conditional independence assumption.Finally we intend to train a GNB and SVM classifier in order to study the effect of dimension reduction and de-correlation of dimensinons on the classification error.

2.3 Evaluation

To evaluate our methods, we plan to run our methods on test data similar to the ones presented in previosu research that used higher dimension data, and we hope to demonstrate similar or better results.

References

[1] Tom M Mitchell, Rebecca Hutchinson, Radu S Niculescu, Francisco Pereira, Xuerui Wang, Marcel Just, and Sharlene Newman. Learning to decode cognitive states from brain images. *Machine Learning*, 57(1-2):145–175, 2004.