Dimension Reduction in fMRI Imaging Data

David (Wei) Jia, Kannu Mehta

August 14, 2013

1 Current Progress

So far, we have tried several different methods to reduce the dimensions of our data. We will go through each and describe them briefly. We will first describe the methods we used in reducing the dimension of fMRI data. Then we give our method of evaluating our results, which are in the same way as the original paper by Mitchell et, al to ensure that our baseline comparisons are the same. Finally, we summarize our current results and provide a brief discussion. Any feedback would be appreciated!

2 Method

In this section we describe the methods that we have tried so far and give a brief overview of each. First, we determined different reduced dimensions to use in our analysis. Then we ran this new, reduced dimension data to train classifiers.

2.1 Determining Dimensions

We first used Principal Component Analysis (PCA) to determine the best dimensions to utilize for our classifiers. The initial dimensions of the data set were equal to the number of voxels in each data sample, which was approximately in the order of 80,000, a very high number. We used PCA to reduce the dimensions to lower than 80. We also chose several different dimensions, each of which took into account successively a larger percent of the variation in the orignal data sample. Let p_i be the percentage variance of the original data retained for threshold level i. Then for each threshold level $p_i \in [0, 1]$, we chose the smallest N_i such that

$$p_i \leq \frac{\sum_{j=1}^{N_i} \lambda_j^{(k)}}{\sum_{j=1}^{M^{(k)}} \lambda_j^{(k)}} \quad \forall k \tag{1}$$

where each $\lambda_j^{(k)}$ is the j-th eigenvalue given by the SVD decomposition of our original data for the k-th human subject, ordered descendingly, i.e. $\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \cdots \geq \lambda_{M^{(k)}}^{(k)}, \forall k$. For our experiment, we chose thresholds

$$P = (0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00) \tag{2}$$

and obtained

$$N = (52, 57, 61, 65, 69, 73, 78) \tag{3}$$

as the smallest N_i that satisfied each threshold value in (2) for the inequation given in (1).

For each P, which defines a "max dimension," we tested using pure PCA, PCA with Whitening, and ZCA with whitening, which we will describe below.

2.2 Principal Component Analysis

In PCA, our goal is to reduce the number of dimensions of our data. The intuition is that the original data may have many dimenions that are highly correlated with each other. We can remove those dimensions by essentially a rotation of the axis. This change of coordinates will effectively collapse dimensions which are highly correlated. Our goal then is to choose the right axis, i.e. new basis vectors to maximize the accountability of variance in our data.

To do this, consider the dataset $\{X^{(1)}, X^{(2)}, \ldots, X^{(m)}\}$. We first perform a set of pre-processing steps to zero-out the mean and standardize the variance. Let $\mu = \frac{1}{m}\sum_{i=1}^{m} X^{(i)}$ be the mean of the original dataset. For each $1 \le i \le m$, set $X^{(i)} := X^{(i)} - \mu$. Let $\sigma_j^2 = \frac{1}{m} = \sum_{i=1}^{m}(X_j^{(i)})^2$ be the variance of the j-the dimension. Then set $X_j^{(i)} = \frac{X_j^{(i)}}{\sigma_j}$ for each $1 \le i \le m$ and $1 \le j \le d$ where d is the original number of dimensiosn in our data.

We now choose unit vectors u such that

$$u = \arg\max_u \frac{1}{m}\sum_{i=1}^{m}((X^{(i)})^T u)^2$$

Let $\Sigma = \frac{1}{m}\sum_{i=1}^{m}(X^{(i)})^T X^{(i)}$ be the covariance matrix. Then we solve the constraint optimization problem

$$\begin{aligned} \text{maximize} \quad & u^T \Sigma u \\ \text{s.t.} \quad & u^T u = 1 \end{aligned} \tag{4}$$

Solving this, we obtain that u must be the principal eigenvector of the covariance matrix Σ. Finally, to reduce our data to k dimensions, we choose $u_1, u_2, \ldots, u_k$ to be the k highest eigenvectors of Σ. We can then obtain a new dataset $X_{\text{PCA}}^{(i)} \in \mathbb{R}^k$ where

$$X_{\text{PCA}}^{(i)} = (u_1^T X^{(i)}, u_2^T X^{(i)}, \ldots, u_k^T X^{(i)})$$

While we have given the general implementation of PCA, our implementation is further complicated by the fact that the original data dimensions is in the order of 80,000, which could make it infeasible to compute using MATLAB on a standard computer. To solve this issue, we applied the following.

Let our training data be $X \in \mathbb{R}^{M \times N}$ where M is the number of dimensions and N is the number of training samples. In our case, we have that M is in the order of 80,000 and N is 78. OUr aim with PCA is to find the eigenvalues of $XX^T \in \mathbb{R}^{M \times M}$. From linear algebra, we know that it is possible to decompose any non-singular matrix Z such that

$$Z = Q_1 \Sigma Q_2^T \tag{5}$$

where $Q_1 \in \mathbb{R}^{M \times M}$, $\Sigma \in \mathbb{R}^{M \times N}$, and $Q_2 \in \mathbb{R}^{N \times N}$. We have that Q_1 and Q_2 are orthogonal matrices, Q_1 is the eigenvector matrix of XX^T and Q_2 is the eigenvector matrix of $X^T X$. This result follow directly from linear algebra theory and we ommit this proof here.

Our goal is then to find Q_1. But recall that since the decomposition in (5) applies for $Z = XX^T$, but is infeasible given the dimension of XX^T, we can simply apply the same decomposition in (5) for $Z = X^T X \in \mathbb{R}^{N \times N}$ since $N << M$.

We can now find Q_2 easily. Also Σ is a diagnoal matrix where the diagonals are the square roots of the eigenvalues obtained by the decomposition of $X^T X$. Then we can find Q_1 as follows

$$Q_1 = (XQ_2)\Sigma^{-1}$$

It's easy to check that this give us the desired result. Note also that Σ is initially a $M \times N$ matrix, but we can strip it to a $N \times N$ matrix since the other elements are zero.

2.3 PCA Whitening

The goal of whitening is to make the input less redundant. That is, we want the input training samples such that (a) the features are less correlated with each other (b) all the features have unit variance. After the application of PCA, dimensions are already uncorrelated, and therefore, we simply normalize each dimension to have the same variance.

2.4 ZCA Whitening

ZCA is a technique to map X_{PCA} back to original number of dimensions while still obtaining the identity covariance. The goal of ZCA, usually, is not to reduce the dimensions, but rather simply to decorrelate them. Let X_{ZCA} be the newly obtained data after ZCA whitening on X, then X_{ZCA} is obtained by:

$$X_{\text{ZCA}} = Q_1 X_{\text{PCA}}$$

where Q_1 is the eigenvector matrix obtained during the application of PCA. It can be shown that the equation produces the data points which are very close to original data points but with an identity covariance.

2.5 Evaluation

To test the efficacy of our dimensions reduction approach, we trained Gaussian Naive Bayes (GNB) classifiers. Trained classifiers were evaluated by their cross-validated classification error or leave-one-trial-out in this case. Note that the same scheme was employed in the original publication (Mitchell et al.), which ensures that we will have the same baseline results.

Following the evaulation methods of Mitchell et al, for each test (which we will describe in the following section), we trained our classifier on 78 of the 80 reduced-dimension data, alternting the 2 unused sets of data for each iteration. We then evaluated our results by testing the classifier on the 2 unused sets of data. These incremental accuracy values are saved and finally averaged for each training-testing iteration. Finally, we report the average accuracies across all iterations.

3 Experiments and Results

For each of the threshold values given in (2), we applied our technique and obtained classifiers with three possible options: (i) with PCA, (ii) with PCA and whitening (iii) with ZCA and whitening. The results are summarized in the table below. Note that the 6 subjects are named A, C, D, E, K, and M corresponding to the naming scheme from the original publication.

3.1 PCA

For no whitening, we obtained the following accuracy results.

Dimensions	Subject A	Subject C	Subject D	Subject E	Subject K	Subject M
$N_1 = 52$	0.5500	0.5000	0.5875	0.6375	0.5250	0.7000
$N_2 = 57$	0.5125	0.5000	0.5250	0.6625	0.5125	0.7875
$N_3 = 61$	0.5375	0.5125	0.5125	0.7375	0.5250	0.7375
$N_4 = 65$	0.5375	0.5000	0.5250	0.7500	0.5375	0.6375
$N_5 = 69$	0.6125	0.6125	0.6125	0.7250	0.5375	0.6125
$N_6 = 73$	0.6375	0.6750	0.7125	0.6750	0.6875	0.5875
$N_7 = 78$	0.6125	0.6875	0.9125	0.6875	0.7750	0.8000

3.2 PCA and Whitening

For PCA with whitening, we obtain the following accuracy results.

Dimensions	Subject A	Subject C	Subject D	Subject E	Subject K	Subject M
$N_1 = 52$	0.5125	0.5000	0.5625	0.6250	0.5000	0.7125
$N_2 = 57$	0.5125	0.5000	0.5000	0.6875	0.5000	0.7875
$N_3 = 61$	0.5250	0.5000	0.5125	0.7250	0.5125	0.7250
$N_4 = 65$	0.5375	0.5000	0.5125	0.7375	0.5125	0.6125
$N_5 = 69$	0.6125	0.6250	0.5375	0.6875	0.5125	0.5875
$N_6 = 73$	0.6375	0.6500	0.6375	0.5875	0.6125	0.5750
$N_7 = 78$	0.6000	0.7000	0.9375	0.7375	0.8375	0.8250

3.3 ZCA and Whitening

For ZCA with whitening, we obtain the following accuracy results.

Dimensions	Subject A	Subject C	Subject D	Subject E	Subject K	Subject M
$N_1 = 52$	0.5500	0.5000	0.5875	0.6375	0.5250	0.7000
$N_2 = 57$	0.5125	0.5000	0.5250	0.6625	0.5125	0.7875
$N_3 = 61$	0.5375	0.5125	0.5125	0.7375	0.5250	0.7375
$N_4 = 65$	0.5375	0.5000	0.5250	0.7500	0.5375	0.6375
$N_5 = 69$	0.6125	0.6125	0.6125	0.7250	0.5375	0.6125
$N_6 = 73$	0.6375	0.6750	0.7125	0.6750	0.6875	0.5875
$N_7 = 78$	0.6125	0.6875	0.9125	0.6875	0.7750	0.8000

4 Discussion

As we can see from the tables above, our reduced dimensions retain high accuracy with some subjects, and as high as in the 90% and lower with others, falling to the 50-60%. Not surprisingly, as we moved our threshold up to take into account more variability, our results became better as a whole. Using whitening did not give us any benefits. It may also be important to note that we have reduced the dimension of our data from the order of 80,000 to lower than 78 dimensions, which is a substaintial reduction. Although we were unable so far to produce *better* results, our ability to retain accuracy for some subjects while reducing dimension by magnitudes of order 10^3 shows promise in our dimension reduction techniques.

In our final paper, we will provide more detailed analysis of our results and give more visual representations of our data. We will also likely try other methods to help improve our results. In the mean time, any feedback or suggestions about how we can futher enhance our work would be greatly appreciated!