Feature Reduction and Feature Selection

One of the key elements that one needs to focus on while processing videos is the element of feature reduction or selection. Increased dimensionality results in over fitting and increased computation time. However, feature reduction methods help in minimizing dimensionality keeping the accuracy at an acceptable level at the same time. Two Methods of dimensionality reduction are to be carried out namely, Principle component Analysis (PCA) and Feature Selection using Spline Regression.

Principle Component Analysis :

The following section gives a detailed understanding of what PCA is and how its been implemented in python.

PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, ie. by reducing the number of dimensions, without much loss of information.

Steps followed:

The following section describes the steps taken to carry out the PCA. Each of these steps have been implemented in Python:

1. Obtain the data : The input data that is being fed into the PCA is our video matrix with dimensions n\*d = 34\*9. The goal of PCA is to find a transformation matrix W of a reduced dimension which is then multiplied with input Matrix X to get the input points projected on the reduced subspace.
2. Calculate the Scatter/Covariance Matrix : Firstly, we need to find a vector called the mean vector whose entities consist of the mean of each dimension across the 34 entities. It is therefore a 9\*1 Vector.

A scatter matrix is then calculate with respect to this mean vector. It results in a 9\*9 vector.

It represents the information between the influence of one dimension on another dimension.

The numbers obtained give us an idea about the magnitude of the influence and degree of correlation

The formula to compute this matrix is as follows:

S = \sum\limits\_{k=1}^n (\mathbf{x}\_k - \mathbf{m})\;(\mathbf{x}\_k - \mathbf{m})^T

Where m is

\mathbf{m} = \frac{1}{n} \sum\limits\_{k=1}^n \; \mathbf{x}\_k

1. The next step involves calculating two very important parameters for PCA. Eigen Values and Eigen Vectors. In this step, we need to first look at the data set and measure the magnitude of variances in different directions, which are perpendicular to each other. The direction vectors thus obtained are the Eigen vectors. The scalar value corresponding to each of these vectors is called the Eigen Value and it measures the strength of the variance in the data set along the direction of the corresponding Eigen vector. These Eigen value,Eigen vector pairs are combined.
2. PCA’s goal is to reduce these dimensions. We are only concerned with those directions where there is a large degree of variance and we discard those directions where the variance isn’t much. So we first sort the Eigen Vector Eigen Value pairs in descending order and take k largest Eigen Vectors. These form the k principle components and now our data can be projected on these principle components.
3. The k largest Eigen Vectors are taken together as a transformation matrix W and is multiplied with input matrix X to get n\*k matrix at the end which is our goal.

We can therefore use this matrix in our calculations from now on since they represent the original matrix with a great degree of accuracy in spite of the reduced dimensionality. Our computations become significantly faster and it is also easier to visualize our data.

Spline Regression :

//Place details of the work done in the spline regression section in our report

To select the most discriminative video features for video semantic recognition,

we assume there is a transformation matrix W n\*c (c < d)

which maps the high-dimensional video samples onto a lower dimensional

subspace, and x’ = W(T)X is the new representation for each video sample xi in such subspace.

As each row of W is used to weight each feature, if some rows of W shrink

to zero, W can be used for feature selection. A better transformation matrix W can be learned by

the minimization of T r(WTMW) , where matrix M encodes

certain structures of the training data.

To calculate the value of M, we need to calculate two matrices A and D. A is the scatter matrix obtained on the labeled data. D is spline scatter matrix obtained on applying spline regression on the input matrix X.

M = A + uD where u is the control parameter with 0<=u<=1

M is a semi-supervised scatter matrix which encodes both data distribution and label

information.

The following are the steps used to carry out spline regression:

1. Estimation of within class scatter matrix : The within class scatter matrix is computed in the same manner as the scatter matrix used for PCA. The only difference here is that we only find it for the features of the labeled videos. The formula to compute the within class scatter matrix is

S = \sum\limits\_{k=1}^n (\mathbf{x}\_k - \mathbf{m})\;(\mathbf{x}\_k - \mathbf{m})^T

Where m is

\mathbf{m} = \frac{1}{n} \sum\limits\_{k=1}^n \; \mathbf{x}\_k

2) Estimation of spline scatter matrix : Suppose matrix G (n\*n)

encodes the local similarity relationship of each pair of samples

in X, then the local structure of training videos can be

preserved in XGXT . If the local geometry of training data (both labeled and un-labeled)

are represented in G, then the unsupervised local distribution

of training data can be utilized. We define the spline scatter

matrix D to be:

D = XGXT ,

where matrix G is obtained by a local spline regression

It has been shown that splines developed in Sobolev space

can be used to interpolate the scattered distribution and

preserve the local geometry structure of training data. A

Sobolev space is a space of functions with sufficiently many

derivatives for some applications domain . One important

property of the Sobolev space is that this space provides

conditions under which a function can be approximated by

smooth functions. Splines developed in Sobolev space

are a combination of polynomials and Green’s function which

is popularly used to interpolate scattered data in geometrical

design. This spline is smooth, nonlinear, and able to

interpolate the scattered data points with high accuracy.

3) Finding Cliques : Given each datum xi in X , to exploit its local similarity

structure, we add its k−1 nearest neighbors as well as xi itself

into a local clique denoted as Ni = {xi, xi1 , xi2 , . . . , xik−1} .

The goal of local spline regression is to find a function gi :

such that it can directly associate each data point

xij in Rd to a class label yij = gi(xij ) (j = 1, 2, . . . , k) ,

which is a regularized regression process:

USE EQUATION 6 in paper