Summary on Basic Dimension Reduction Techniques

Henry Yuchi

July 7, 2017

1 Introduction

Dimension reduction attempts to describe a function of many variables with fewer ones by reducing the number of independent variables. In this week I have been looking into four basic methods, which are the linear model, the quadratic model, SIR and SAVE respectively.

Linear Model The linear model is a straightforward technique, generating a new variable based on a linear combination of all independent variables.

$$y = \mathbf{a}^{\mathrm{T}}\mathbf{x} + b$$

Hence this linear combination serves as the input variable for output/observation. It reduces the relationship to a first order one if exists by solving a least squares problem $A\mathbf{w} = \mathbf{b}$.

- For each set of inputs, take their values as entries in a row in matrix A.
- The vector $\mathbf{w}$ is the unknowns to be solved for this least squares problem, which contains the coefficients for all first order terms and the bias.
- The observations (outputs) are put in the vector $\mathbf{b}$ accordingly.
- By finding the LSS, the vector a and b are recovered. Note vector b is the constant bias in the model, which means when constructing matrix A there should be a column of 1 taking the bias terms.

2 Quadratic Model

The quadratic model is similar to the linear model, except that the model is quadratic and produces two sets of variables after dimension reduction.

$$y = \mathbf{x}^{\mathrm{T}}B\mathbf{x} + \mathbf{c}^{\mathrm{T}}\mathbf{x} + d$$

Similarly the model can be solved by carrying out a least squares problem $A\mathbf{w} = \mathbf{b}$.

- For each set of input variables, put them in the same row in matrix A, including a column of 1s for bias.
- Each observation/output is put in the vector b.
- The coefficients are set as unknowns for the least squares problem in w. This vector contains the coefficients for all the square terms for each independent variable, the coefficients for all second order cross terms, the ones for all first order terms and the bias.
- By finding the LSS the output can be shown as a function of two variables. Hence the output can be viewed as a function based on two variables.

3 Sliced Inverse Regression (SIR)

Sliced inverse regression approximates the central DRS by using the expectation of the inverse regression.

- First standardize the input variables/predictors.
  $\mu = \mathbf{E}[x]$ and $\Sigma_x = Cov[x]$ where the covariance matrix has a Cholesky factorization result of $\Sigma_x = Cov[x]$=$\mathbf{L}\mathbf{L}^{\mathrm{T}}$.
  Hence the standardized predictor $\mathbf{z} = \mathbf{L}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ has an expectation of zeros and identity covariance matrix $\mathbf{E}[z] = 0$ and $\mathrm{Cov}[z] = \mathbf{I}$.

- Secondly partition the observations into R bins. Usually the bin widths are uniform.
  $y_{min} = \bar{y}_0 < \bar{y}_1 < ... < \bar{y}_{R-1} < \bar{y}_R = y_{max}$ The corresponding inputs are partitioned into these R bins as well.

- For each bin compute the sample means for the predictors.
  $\hat{\boldsymbol{\mu}}(r) = \frac{1}{N_r}\Sigma\mathbf{x}_i$ where i is the bin number and N_r is the number of input points in that bin.

- Then compute the weighted covariance matrix.
  $\hat{\mathbf{D}}_{SIR} = \frac{1}{N}\Sigma N_r \hat{\boldsymbol{\mu}}(r)\hat{\boldsymbol{\mu}}(r)^{\mathrm{T}}$ which is a summation of all R bins.

- Compute the eigen decomposition via SVD.
  $\hat{\mathbf{D}}_{SIR} = \hat{\mathbf{W}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{W}}^{\mathrm{T}}$ Then select the n largest eigenvalues and the corresponding eigenvectors.

- To show the results in a 2D plot the eigenvectors with two largest eigenvalues should be picked and the calculated $w_1^T x$ and $w_2^T x$ are the two axes.

4 Sliced Average Variance Estimation (SAVE)

Similarly the SAVE technique also divides the observation points into a number of bins and then carry out the sample means and variances.

- First standardize the input variables/predictors.
  $\mu = \mathbf{E}[x]$ and $\Sigma_x = Cov[x]$ where the covariance matrix has a Cholesky factorization result of $\Sigma_x = Cov[x]$=$\mathbf{L}\mathbf{L}^{\mathrm{T}}$.
  Hence the standardized predictor $\mathbf{z} = \mathbf{L}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ has an expectation of zeros and identity covariance matrix $\mathbf{E}[z] = 0$ and $\mathrm{Cov}[z] = \mathbf{I}$.

- Secondly partition the observations into R bins. Usually the bin widths are uniform.
  $y_{min} = \bar{y}_0 < \bar{y}_1 < ... < \bar{y}_{R-1} < \bar{y}_R = y_{max}$ The corresponding inputs are partitioned into these R bins as well.

- For each bin compute the sample means for the predictors.
  $\hat{\boldsymbol{\mu}}(r) = \frac{1}{N_r}\Sigma\mathbf{x}_i$ where i is the bin number and N_r is the number of input points in that bin.

- Then compute the sample covariance of the predictors for the points in each bin.
  $\hat{\boldsymbol{\Sigma}}(r) = \frac{1}{N_r - 1}\Sigma(\mathbf{x}_i - \hat{\boldsymbol{\mu}}(r))(\mathbf{x}_i - \hat{\boldsymbol{\mu}}(r))^{\mathrm{T}}$

- From the sample covariance matrices we can obtain
  $\hat{\mathbf{D}}_{SAVE} = \frac{1}{N}\Sigma N_r(\mathbf{I} - \hat{\boldsymbol{\Sigma}}(r))^2$ which is a summation of all R bins.

- Finally we can compute the eigendecomposition
  $\hat{\mathbf{D}}_{SAVE} = \hat{\mathbf{W}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{W}}^{\mathrm{T}}$
  The largest two eigenvalues are found and the corresponding eigenvectors are taken out. As before, to show the results in a 2D plot the eigenvectors with two largest eigenvalues should be picked and the calculated $w_1^T x$ and $w_2^T x$ are the two axes.