Statistical Learning

Midterm 2

1. For a given data set with 30 explanatory variables the residual sums of squares from the least squares method and ridge regression are equal to : 5.5 and 15 respectively. For the ridge regression the trace of $X(X'X + \gamma I)^{-1}X'$ is equal to 15. The standard deviation of the regression error term is equal to 1. Which of these two methods yields the best estimated prediction error ?

2. Consider the orthogonal design $X'X = I$ and the elastic net in the form

$$\hat{\beta} = argmin_{\beta}||Y - X\beta||_2^2 + 2||\beta||_2^2 + 3||\beta||_1 \ .$$

   Derive the analytical form of the solution as the function of the least squares estimator.

3. Consider the orthogonal design $X'X = I$ and the LASSO estimator in the form

$$\hat{\beta} = argmin_{\beta}\frac{1}{2}||Y - X\beta||_2^2 + 3||\beta||_1 \ .$$

   What is the

   - probability that a given false predictor will be selected (i.e. type I error)
   - the power of identifying the true predictor with $\beta = 3$.

4. The covariance matrix of the three dimensional normal distribution has the form

$$\Sigma = \begin{bmatrix} 6 & -4 & 2 \\ -4 & 8 & -4 \\ 2 & -4 & 6 \end{bmatrix} \ . \tag{1}$$

   a) Which of the coordinates are independent ?

   b) Which of the coordinate variables are conditionally independent [given the remaining variable] ?

   c) Draw a graphical model of this distribution.

5 Consider the following covariance matrix of the bivariate normal vector $\mathbf{X} = (X_1, X_2)$:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

   a) Find the eigenvalues and the eigenvectors of this covariance matrix.

   b) Find the first principal component of this distribution.

   c) What proportion of the total variation is explained by the first principal component ?

6 Explain sparse PCA - what is the goal and how it can be achieved.