臺灣大學數學系

九十五學年度博士班入學考試題

迴歸分析

Jun, 2006

1. (15) Suppose $Y = X\beta + \varepsilon$, where $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{cov}(\varepsilon) = diag(c_1, \cdots, c_n)\sigma^2$.
   (a) Give three examples and specify $\mathrm{cov}(\varepsilon)$ in each case.
   (b) If $c_i^{-1} = \eta(Ey_i)$ for some unknown function η, how to estimate β, σ^2 and c_i's?

2. (20) Consider
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \beta_3(x_{i1} - x)\delta_i + \beta_4(x_{i1} - x)^2\delta_i + \varepsilon,\ i = 1, \cdots, n,$$
where $\delta_i = I(x_{i1} > x)$ and ε_i, $i = 1, \cdots, n$, are uncorrelated random errors. Give a method to estimate the regression curve when x is not known.

3. (25) Suppose $Y = X\beta+\varepsilon$, where X is $n\times k$ dimensional, β is a k-vector, $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{cov}(\varepsilon) = \sigma^2 I$, is the correct model. Let $X = (X_1, X_2)$ and $\beta' = (\beta_1', \beta_2')$, where X_1 is $n \times p$ dimensional and β_1 is a p-vector, $p < k$. Suppose we fit the model $Y = X_1\beta_1 + \varepsilon$ and obtain the least squares estimator b_p of β_1, the estimator s_p^2 of σ^2, the residual sum of squares RSS_p and the prediction $\hat{Y}_p$ of Y.
   (a) Compare the bias and covariance matrix of b_1 with those of $b^{(1)}$, LSE of β_1 via the true model.
   (b) Compute the bias of s_p^2.
   (c) Compute TMSE($\hat{Y}_p$), trace of MSE($\hat{Y}_p$)=cov($\hat{Y}_p$) + Bias($\hat{Y}_p$)Bias($\hat{Y}_p$)', and conclude that TMSE($\hat{Y}_p$)$/\sigma^2$ can be estimated by Mallow's C_p
$$C_p = \frac{RSS_p}{s^2} - (n - 2p),$$
   where s^2 is the estimator of σ^2 when fitting the correct model.

4. (10) Give two variable selection procedures for general linear model $Y = X\beta+\varepsilon$.

5. (30) Consider the general linear model $Y = X\beta + \varepsilon$, where $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{cov}(\varepsilon) = \sigma^2 I$.
   (a) Describe multicollinearity of X and show its effect on least squares estimation of β.

(b) Describe the following three remedies of multicollineary and show how they reduce the problem.

   i. incomplete principal component regression
   ii. ridge regression
   iii. add additional observations