Dr. Boqiang Huang

Mathematical Foundations of Data Analysis II — SS 2019

Problem Set 04

Published on 07.05.2019 To be collected on 14.05.2019

Problem 1: (10 points)
Consider the general linear regression model:

$$\mathrm{Y} = \beta_0 + \sum_{j=1}^{p} \beta_j \mathrm{X}_j + \epsilon, \tag{1}$$

where X_j, ϵ and Y are the predictors, the zero-mean error and the corresponding response.
(a) Show that the model (1) is equivalent to the following one:

$$\mathrm{Y} - \mathrm{E}[\mathrm{Y}] = \sum_{j=1}^{p} \beta_j \left(\mathrm{X}_j - \mathrm{E}[\mathrm{X}_j]\right) + \epsilon, \tag{2}$$

where the intercept β_0 can be omitted;
(b) Suppose we have $n, n > p$, mean removed observations that denoted as $\boldsymbol{y} = [y_1, \dots, y_n]^T$, and $X = [x_{i,j}]_{n\times p} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_p]$ where $\boldsymbol{x}_j = [x_{1j}, \dots, x_{nj}]^T, i = 1, \dots, n, j = 1, \dots, p$. Here, the mean removal implies that $\sum_{i=1}^n y_i = 0$ and $\sum_{i=1}^n x_{ij} = 0, j = 1, \dots, p$. Then (2) could be represented by

$$\boldsymbol{y} = X\beta + \varepsilon, \tag{3}$$

where $\beta := [\beta_1, \dots, \beta_p]^T, \varepsilon := [\varepsilon_1, \dots, \varepsilon_n]^T$. To obtain the coefficient estimate $\hat{\beta}$, we may solve the following optimization problem

$$\hat{\beta}(s) = \arg\min_{\|\beta\|_{\ell_2}^2 \le s} \|\boldsymbol{y} - X\beta\|_{\ell_2}^2, \tag{4}$$

where s is a predefined positive threshold for the ℓ_2 norm of β. Please show that the coefficient estimate $\hat{\beta}$ can be equivalently derived as

$$\hat{\beta}(\lambda) = \left(X^T X + \lambda I_{p\times p}\right)^{-1} X^T \boldsymbol{y}, \tag{5}$$

if we transform the constrained optimization problem in (4) to the unconstrained one with penalty term using the augmented Lagrangian method as follows

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \|\boldsymbol{y} - X\beta\|_{\ell_2}^2 + \lambda \|\beta\|_{\ell_2}^2, \tag{6}$$

where λ is a suitably chosen Lagrange-multiplier.
(c) With the singular value decomposition of $X, X = U_{n\times n}\Sigma_{n\times p}V_{p\times p}^T$, where both U and V are unitary matrices, show that the ridge regression fit $\hat{\boldsymbol{y}}(\lambda) := X\hat{\beta}(\lambda)$ is just a linear combination of shrank response components y_i

$$\hat{\boldsymbol{y}}(\lambda) = \sum_{j=1}^{p} \frac{\sigma_j^2}{\sigma_j^2 + \lambda} \boldsymbol{u}_j \boldsymbol{u}_j^T \boldsymbol{y}, \tag{7}$$

where σ_j is the jth diagonal entry of Σ, and $\boldsymbol{u}_j$ is the jth column vector of U.

Problem 2: (5 points)
Consider the same linear regression model shown in (1) - (3), the lasso problem can be represented as

$$\hat{\beta}(s) = \arg\min_{\|\beta\|_{\ell_1} \le s} \|\boldsymbol{y} - X\beta\|_{\ell_2}^2, \tag{8}$$

or

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \|\boldsymbol{y} - X\beta\|_{\ell_2}^2 + \lambda \|\beta\|_{\ell_1}, \tag{9}$$

where s is a predefined positive threshold for the ℓ_1 norm of β, and λ is a suitably chosen Lagrange-multiplier. Assume that matrix X is the orthonormal design matrix, i.e. $X^T X = I_{p\times p} = (X^T X)^{-1}$, show that the lasso estimator is

$$\hat{\beta}_j(\lambda) = \mathrm{sign}(\hat{\beta}_j)\left(|\hat{\beta}_j| - \frac{1}{2}\lambda\right)_+, \tag{10}$$

where $\hat{\beta} = \left(X^T X\right)^{-1} X^T \boldsymbol{y} = X^T \boldsymbol{y}$ is the ordinary least squares estimator of β and $\hat{\beta}_j$ its jth element, and $f(x) = (x)_+ = \max\{x, 0\}$.

Problem 3: (5 points)
Please derive and explain the formula of the principle component analysis.