Regression and Classification

Aug-Nov 2019

Assignment 1

Consider linear model:

$$y = X\beta + \epsilon,$$

where $y = (y_1, y_2, \cdots, y_n)$, $X = ((x_{ij}))_{n\times p}$, $\beta = (\beta_1, \beta_2, \cdots, \beta_p)$ and $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$.

1. Show that the least squares method guarantees at least one solution.

2. Under what condition, the least square method will have a unique solution.

3. If the condition for a unique solution in the least square methods does not meet, what happens?

4. State and prove the Gauss-Markov Theorem.

5. If error structure, in linear models, follows $N(0, \sigma^2)$, then find the sampling distribution of the regression coefficients $\hat{\beta}$, where
$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

6. Briefly explain why feature extraction is essential?

7. What is the total model space with p many features? Explain.

8. What is complexity of step-wise feature selection? Explain.

9. We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain $p + 1$ models, containing $0, 1, 2, \cdots, p$ predictors. Which of the three models with k predictors has the smallest training RSS?

10. In this exercise, we will generate simulated data, and will then use this data to perform best subset selection.
    (a) Use the `rnorm()` function to generate a predictor X of length $n = 100$, as well as a noise vector ϵ of length $n = 100$.
    (b) Generate a response vector Y of length $n = 100$ according to the model
    $$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon,$$
    where $\beta_0, \beta_1, \beta_2, \beta_3$ are constants of your choice.
    (c) Use the `regsubsets()` function to perform best subset selection in order to choose the best model containing the predictors, $X, X^2, \cdots, X^{10}$. What is the best model obtained according to C_p, BIC or adjusted R^2?
    (d) What are the models selected by forward and backward selection mthod?
    (e) Fit LASSO model and compare the four models.

11. What do you mean by multi-collinearity?