Statistical Learning

Homework 3

Part 1: Conceptual Questions

Problem 1. Consider a regression of Y_i against a constant and X_i. Let $\hat{\beta}_0$, $\hat{\beta}_1$, and s^2 denote the estimated intercept, estimated slope parameter, and estimator of the variance of errors from that regression. Let T denote the t-statistic for testing H_0 that the slope parameter is zero in that regression. Let *pval* be the corresponding p-value. Now, let c_1 and c_2 be two constants ($c_2 \neq 0$). Define a new dependent variable and a new regressor as

$$Y_i^* = c_1 Y_i,$$
$$X_i^* = c_2 X_i.$$

Let $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and s_*^2 denote the estimated intercept, estimated slope parameter, and estimator of the variance of errors from the regression of Y_i^* against a constant and X_i^*. Let T^* denote the t-statistic for testing H_0 that the slope parameter in the regression of Y_i^* against a constant and X_i^* is zero. Let $pval^*$ be the corresponding p-value.

1. Find an expression for $\hat{\beta}_1^*$ in terms of $\hat{\beta}_1, c_1$, and c_2.
2. Find an expression for $\hat{\beta}_0^*$ in terms of $\hat{\beta}_0$ and c_1.
3. Find an expression for s_*^2 in terms of s^2 and c_1.
4. What is the relationship between T and T^*?
5. What is the relationship between *pval* and $pval^*$?

Problem 2. ISL (2nd edition) Page 219, Question 1.

Problem 3. ISL (2nd edition) Page 284, Question 5.

Problem 4. ISL (2nd edition) Page 285, Question 7. Read "Bayesian Interpretation for Ridge Regression and the Lasso" on Page 248.

Problem 5. Another resampling method is called jackknife, which is similar to LOOCV. Suppose that $\hat{\theta} = \varphi_n(Z_1, Z_2, ..., Z_n)$ is the estimator of an parameter θ. Denote $\hat{\theta}_{-j} = \varphi_{n-1}(Z_1, ..., Z_{j-1}, Z_{j+1}, ..., Z_n)$. $\hat{\theta}_{-j}$ is an estimator obtained by removing the j-th observation from the entire sample. The variation in $\left\{\hat{\theta}_{-j} : j = 1, ..., n\right\}$ should be informative about the population variance of $\hat{\theta}_n$. Denote $\bar{\hat{\theta}} = n^{-1}\sum_{j=1}^n \hat{\theta}_{-j}$. The Jackknife standard error is

$$\widehat{se}_{jk} = \sqrt{\frac{n-1}{n}\sum_{j=1}^n\left(\hat{\theta}_{-j} - \bar{\hat{\theta}}\right)^2}.$$

An approximate 95% confidence interval is $\left[\hat{\theta}_n - 2 \cdot \widehat{se}_{jk}, \hat{\theta}_n + 2 \cdot \widehat{se}_{jk}\right]$. Consider the following simple example: for i.i.d. random variables $X_1, X_2, ..., X_n$, where $X_i \sim \mathrm{N}(\theta, \sigma^2)$, $\hat{\theta}_n = n^{-1} \sum_{i=1}^{n} X_i$ is an estimator of θ. Argue that when n is large, $\Pr\left[\hat{\theta}_n - 2 \cdot \widehat{se}_{jk} \leq \theta \leq \hat{\theta}_n + 2 \cdot \widehat{se}_{jk}\right]$ is approximately 95% by showing that $(n-1) \sum_{j=1}^{n} \left(\hat{\theta}_{-j} - \overline{\hat{\theta}}\right)^2$ is equal to the sample variance.

Part 2: Applied Questions

Problem 6. ISL (2nd edition) Page 220, Question 5.

Problem 7. ISL (2nd edition) Page 221, Question 6.

Problem 8. ISL (2nd edition) Page 285, Question 8.

Problem 9. ISL (2nd edition) Page 286, Question 9 (a,b,c,d).