STA 35C Statistical Data Science III

Midterm 2 practice A solutions

Instructor: Dogyoon Song

Problem 1. Basic concepts

(a) **Variance of the fitted prediction** across training samples.

(b) $3 \times 5 = \mathbf{15}$ fits, excluding the final fit on all data.

(c) Each draw misses the specified row with probability $1 - 1/n$. Independence of the draws gives $(\mathbf{1 - 1/n})^{\mathbf{n}}$.

(d) $6 - 2 = \mathbf{4}$ candidate additions.

(e) The limiting constant fitted prediction is $\overline{\boldsymbol{y}}$: the penalized slopes approach zero and the unpenalized intercept approaches the response mean.

(f) Independence and the stated exact false-rejection probability give

$$P(\text{at least one false rejection}) = 1 - P(\text{none}) = \mathbf{1 - 0.95^{10} \approx 0.401}.$$

Problem 2. Prediction error and resampling

(a) **Moments and prediction error.** Let $Z = \widehat{f}(x_0)$. The given distribution has

$$\mathbb{E}[Z] = \frac{1+2+3}{3} = 2, \qquad \mathrm{Var}(Z) = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = \frac{2}{3}.$$

These are distributional moments, so the divisor is 3, not $3 - 1$. Using $f(x_0) = 3$ and independent new-noise variance 1,

$$\mathbb{E}[(Y_0 - \widehat{f}(x_0))^2] = (2-3)^2 + \frac{2}{3} + 1 = \frac{\mathbf{8}}{\mathbf{3}} \approx \mathbf{2.667}.$$

(b) **Unequal-fold CV MSE.** Convert each fold mean to a sum before pooling:

$$\text{CV MSE} = \frac{2 \times 1 + 3 \times 2 + 5 \times 5}{2+3+5} = \frac{33}{10} = \mathbf{3.3}.$$

This weights every held-out observation equally; an unweighted average of the three fold MSEs does not.

(c) **Bootstrap standard error.** The mean of the five estimates is $\overline{\beta}^* = 2$, and

$$\sum_{b=1}^{5} (\widehat{\beta}_b^* - 2)^2 = 1 + 0 + 0 + 0 + 1 = 2.$$

Thus

$$\widehat{\mathrm{SE}}_{\mathrm{boot}} = \sqrt{\frac{2}{5-1}} = \frac{\mathbf{1}}{\sqrt{\mathbf{2}}} \approx \mathbf{0.707}.$$

Problem 3. Subset search and ridge penalization

(a) Best subset at each size. Minimize RSS separately among the subsets of each fixed size:

Number of predictors k	Best subset at this size	RSS
0	$\varnothing$	240
1	$\{X_1\}$	72
2	$\{X_2, X_3\}$	30
3	$\{X_1, X_2, X_3\}$	30

These are candidates at fixed sizes, not a final choice of model size.

(b) Forward-selected pair. First add X_1 because $72 < 135 < 198$. Next compare only the additions to $\{X_1\}$:

$$\mathrm{RSS}(X_1, X_2) = 65 < 70 = \mathrm{RSS}(X_1, X_3).$$

Add X_2, obtaining $\boldsymbol{\{X_1, X_2\}}$. Forward selection cannot remove its first choice X_1, so it never considers the better pair $\{X_2, X_3\}$ at the two-predictor stage.

(c) Ridge objective comparison. Excluding the intercept from the penalty,

$$Q_A = 8 + 2(3^2 + 1^2) = 8 + 20 = 28,$$
$$Q_B = 12 + 2(1^2 + 1^2) = 12 + 4 = 16.$$

Candidate B has the smaller objective, $16 < 28$, among these two candidates.

Problem 4. Reviewing a bike-rental forecast

(a) Tuning and prediction. The smallest CV MSE is 5.1, attained at $\boldsymbol{\lambda = 2}$. Its final fitted equation is $\widehat{y} = 12 + 2x$, giving

$$\widehat{y}(2) = 12 + 2(2) = \mathbf{16}.$$

(b) Bonferroni. The cutoff is $0.05/4 = 0.0125$. Only 0.010 is at most this cutoff, so the rejection set is $\boldsymbol{\{H_{0,1}\}}$.

Holm. Compare the ordered p-values with $\alpha/(m - i + 1)$, stopping at the first failure:

Rank i	Ordered p-value	Holm cutoff	Comparison
1	0.010	$0.05/4 = 0.0125$	Pass
2	0.016	$0.05/3 \approx 0.016667$	Pass
3	0.030	$0.05/2 = 0.025$	Fail; stop
4	0.200	0.05	Not reached

The Holm rejection set is $\boldsymbol{\{H_{0,1}, H_{0,2}\}}$.

(c) Corrected fold-wise procedure. Within each CV fold, select the two predictors with the largest absolute sample correlations with the response using only that fold's training observations, then fit the model on those same observations. Use the held-out responses only to evaluate the resulting predictions, and repeat selection and fitting separately within each fold's training data.

Problem 5. Optional bonus questions

(a) Expected number of distinct row indices. Let I_i indicate that row index i appears at least once. Conditional on the original data,

$$D = \sum_{i=1}^{n} I_i, \qquad \mathbb{E}[I_i] = P(I_i = 1) = 1 - \left(1 - \frac{1}{n}\right)^n.$$

By linearity of expectation,

$$\mathbb{E}[D] = \sum_{i=1}^{n} \mathbb{E}[I_i] = \boldsymbol{n}\left[\mathbf{1} - \left(\mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{n}}\right)^{\boldsymbol{n}}\right].$$

Independence of the inclusion indicators I_i is not needed.

(b) Yes; every ridge minimizer has $b_1 = b_2$. Suppose instead that $b_1 \neq b_2$, and replace both by $a = (b_1+b_2)/2$, leaving all other coefficients unchanged. Since $x_{i1} = x_{i2}$, this preserves each fitted value:

$$b_1 x_{i1} + b_2 x_{i2} = (b_1 + b_2) x_{i1} = a x_{i1} + a x_{i2}.$$

The RSS is therefore unchanged, but the penalty decreases by

$$\lambda\big(b_1^2 + b_2^2 - 2a^2\big) = \frac{\lambda}{2}(b_1 - b_2)^2 > 0.$$

This contradicts optimality; thus unequal coefficients cannot occur at a minimizer.