STA 35C Statistical Data Science III

Practice midterm exam 2 – Set A

Instructor: Dogyoon Song

Name: _____ **Student ID:** _____

Instructions: This midterm exam is **closed-book**, except for the permitted note sheet described below. You may bring a pen or pencil, one letter-sized sheet of *hand-written* notes (both sides), and a *non-graphing* calculator. No other materials (e.g., textbooks) are allowed. For a timed attempt, allow 50 minutes. Problems 1–4 total **120 points**; Problem 5 offers *up to 10 bonus points*. Once you receive this exam, please confirm that you have all 6 pages.

- Make sure to clearly write your name and student ID above.
- For Problem 1, give only the requested answers; no explanations are required. For Problems 2–5, show the relevant steps for partial credit. Sentence limits apply to explanations, not displayed calculations.
- If necessary, round numerical answers to three decimal places; fractions, logarithms, and exponentials may be left unevaluated.
- Complete Problems 1–4 before attempting the optional bonus questions in Problem 5.

Problem	**Score**
Problem 1	
Problem 2	
Problem 3	
Problem 4	
Problem 5 (bonus)	
Total	

Problem 1 (30 points). Basic concepts

Give only the requested answers. No explanations are required.

(a) **(5 points)** Which component of the usual squared prediction error decomposition measures **sensitivity of the fitted prediction to the training sample**? Circle one: **squared bias / variance of the fitted prediction / irreducible noise**.

(b) **(5 points)** Three fixed tuning values are evaluated by five-fold cross-validation, using direct refitting for each tuning value and fold. How many model fits are required, **excluding the final fit on all data**?

(c) **(5 points)** An ordinary bootstrap resample draws n times independently with replacement from n observed rows. Give the probability that **one specified row is never drawn**. You may leave a power unevaluated.

(d) **(5 points)** Forward stepwise selection starts with six candidate predictors and has already included two of them. How many **candidate additions** are compared at the next step?

(e) **(5 points)** Ridge regression is fitted on a fixed dataset with response mean $\overline{y}$. Its intercept is unpenalized. As $\lambda \to \infty$, what **constant fitted prediction** does it approach?

(f) **(5 points)** All 10 null hypotheses are true. Their tests are independent, and each rejects with probability **exactly** 0.05. Give the probability of **at least one false rejection**. Leave your answer as an expression if preferred.

Problem 2 (30 points). Prediction error and resampling

Each part concerns a separate analysis.

(a) (10 points) At a fixed input x_0, $f(x_0) = 3$. Across training samples, $\widehat{f}(x_0)$ takes the values $1, 2, 3$, each with probability $1/3$; these are its **only possible values**. A new response is $Y_0 = 3 + \varepsilon_0$, where ε_0 is independent of the training sample, with mean 0 and variance 1.

Compute $\mathbb{E}[(Y_0 - \widehat{f}(x_0))^2]$, showing the mean and variance of $\widehat{f}(x_0)$ in your calculation.

(b) (10 points) Ten observations are partitioned into three validation folds. Each model fit uses only the observations outside its validation fold. The held-out results are:

Validation fold	1	2	3
Number of held-out observations	2	3	5
Mean squared prediction error within the fold	1	2	5

Compute the CV MSE per observation, pooling all held-out predictions.

(c) (10 points) Five paired bootstrap resamples of an i.i.d. regression dataset give the following slope estimates:

$$1, \quad 2, \quad 2, \quad 2, \quad 3.$$

Compute the bootstrap estimate of the original slope estimator's standard error, using the sample standard deviation of these five estimates (divisor $5 - 1$). These five replicates are supplied only to illustrate the calculation.

Problem 3 (30 points). Subset search and ridge penalization

For parts (a)–(b), the following RSS values come from least-squares fits on the **same training data**. Every model includes an intercept; $\varnothing$ denotes no predictors.

Predictors	RSS	Predictors	RSS
$\varnothing$	240	X_1, X_2	65
X_1	72	X_1, X_3	70
X_2	135	X_2, X_3	30
X_3	198	X_1, X_2, X_3	30

(a) (10 points) For each subset size k, **write the predictor subset with the smallest RSS** in the table below. Use $\varnothing$ for $k = 0$.

Number of predictors k	Best subset at this size
0	
1	
2	
3	

(b) (10 points) Starting with the intercept-only model, apply **forward stepwise selection using training RSS** until exactly two predictors are included. Report:

(i) the selected pair, showing which predictor is added at each step;
(ii) a **one-sentence explanation** of why it differs from the best two-predictor subset in part (a).

(c) (10 points) For a **separate dataset** with two standardized predictors, consider

$$Q(b_0, b_1, b_2) = \mathrm{RSS}(b_0, b_1, b_2) + 2(b_1^2 + b_2^2).$$

The intercept is not penalized. Two candidates give:

Candidate	b_0	b_1	b_2	Training RSS
A	5	3	1	8
B	5	1	1	12

Report the objective value for each candidate and identify which candidate has the smaller value. Compare only these two candidates; do not optimize over other coefficient vectors.

Problem 4 (30 points). Reviewing a bike-rental forecast

A team predicts daily bike-rental demand. Its numerical analyses and a proposed workflow are described below.

(a) (10 points) Three ridge penalty values are evaluated using the **same CV folds**. The equations below are the corresponding fits on all development data; the intercept is unpenalized.

Penalty λ	CV MSE	Fitted equation
0	6.4	$\widehat{y} = 10 + 4x$
2	5.1	$\widehat{y} = 12 + 2x$
6	5.8	$\widehat{y} = 13 + x$

Report both quantities:

(i) the λ with the smallest CV MSE;
(ii) that setting's fitted prediction at $x = 2$.

(b) (10 points) A separate study produces the following valid p-values for four prespecified tests. These are the **entire testing family**, already in increasing order.

Null hypothesis	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$	$H_{0,4}$
p-value	0.010	0.016	0.030	0.200

At family-wise level 0.05, report the rejection set for each method, showing its cutoff comparisons:

(i) **Bonferroni** (5 points);
(ii) **Holm** (5 points).

(c) (10 points) For an alternative prediction procedure, an analyst selects the two predictors with the largest absolute sample correlations with the response using **all development observations**. The analyst then holds these two predictors fixed during five-fold CV and claims that this assesses the complete selection-and-fitting procedure.

Describe the corrected procedure within each CV fold in **at most two sentences**, specifying which observations are used for selection, fitting, and evaluation.

Problem 5 (10 bonus points). Optional bonus questions

These questions are independent of the required questions and of each other.

(a) (5 points) An ordinary bootstrap resample draws n times independently with replacement from n observed rows. Rows are treated as distinct by their **indices**, even if their values agree. Let D be the number of distinct row indices represented in the resample.

Find $\mathbb{E}[D]$ conditional on the original dataset and justify your answer. You may leave a power unevaluated.

(b) (5 points) Two predictor columns are identical: $x_{i1} = x_{i2}$ for every training observation i. For $\lambda > 0$, ridge regression minimizes

$$\sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} b_j^2.$$

Must every minimizer satisfy $b_1 = b_2$? Give a brief algebraic argument or a counterexample.