STOR 565 Final Exam

Rui Li

TOTAL POINTS

65 / 100

QUESTION 1

1 Honor Code Pledge **0 / 0**

QUESTION 2

2 Question 1 **6 / 8**

QUESTION 3

3 Question 2 **7 / 8**

QUESTION 4

4 Question 3 **12 / 12**

QUESTION 5

5 Question 4 **7 / 8**

QUESTION 6

6 Question 5 **8 / 8**

QUESTION 7

Question 6 10 pts

7.1 Part a **1.5 / 2**

7.2 Part b **2 / 2**

7.3 Part c **1.5 / 3**

7.4 Part d **2 / 3**

QUESTION 8

8 Question 7 **4 / 8**

QUESTION 9

9 Question 8 **0 / 8**

QUESTION 10

Question 9 8 pts

10.1 Part a **0.5 / 2**

10.2 Part b **0.5 / 2**

10.3 Part c **0 / 2**

10.4 Part d **2 / 2**

QUESTION 11

Question 10 14 pts

11.1 Part a **3 / 14**

QUESTION 12

12 Question 11 **8 / 8**

Honor code pledge

— Rui Li

1 Honor Code Pledge 0 / 0

Rui Li 730287790

1. From the lecture notes of Order Relations for Maxima and Minima,

$$\min_{1\le i\le n} a_i = -\max_{1\le i\le n}(-a_i)$$

$$\min_{1\le i\le n} b_i = -\max_{1\le i\le n}(-b_i)$$

Thus $\min\limits_{1\le i\le n} a_i - \min\limits_{1\le i\le n} b_i$

$$= -\max_{1\le i\le n}(-a_i) - \left(-\max_{1\le i\le n}(-b_i)\right)$$

$$= \max_{1\le i\le n}(-b_i) - \max_{1\le i\le n}(-a_i)$$

$$\le \max_{1\le i\le n}\left[(-b_i)-(-a_i)\right]$$

$$= \max_{1\le i\le n}(a_i - b_i)$$

Therefore, the upper bound on $\min\limits_{1\le i\le n} a_i - \min\limits_{1\le i\le n} b_i$ is $\max\limits_{1\le i\le n}(a_i - b_i)$.

2 Question 1 6 / 8

2. Define $u = (u_1, u_2, u_3, \ldots, u_d)^t \in \mathbb{R}^d$
$v = (v_1, v_2, v_3, \ldots, v_d)^t \in \mathbb{R}^d$

The general sufficient condition satisfy $\langle u, v \rangle = \|u\| \|v\|$ is:

$$\frac{u_1}{v_1} = \frac{u_2}{v_2} = \frac{u_3}{v_3} = \cdots = \frac{u_d}{v_d} = c \in \mathbb{R} \quad — ①.$$

Proof:

$$\langle u, v \rangle = u^t v = u_1 v_1 + u_2 v_2 + \cdots + u_d v_d$$
$$\|u\| = (u_1^2 + u_2^2 + \cdots + u_d^2)^{\frac{1}{2}}$$
$$\|v\| = (v_1^2 + v_2^2 + \cdots + v_d^2)^{\frac{1}{2}}$$

From the condition ①, we can see

$$\|u\|\|v\| = \left(\sum_{i=1}^d u_i^2\right)^{\frac{1}{2}} \cdot \left(\sum_{i=1}^d v_i^2\right)^{\frac{1}{2}}$$
$$= \left(\sum_{i=1}^d (c \cdot v_i)^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^d (v_i^2)\right)^{\frac{1}{2}}$$
$$= c \cdot \sum_{i=1}^d v_i^2$$

$$\langle u, v \rangle = \sum_{i=1}^d u_i v_i = \sum_{i=1}^d (c \cdot v_i) \cdot v_i$$
$$= \left(\sum_{i=1}^d v_i^2\right) \cdot c$$

Therefore, $\|u\|\|v\| = \langle u, v \rangle$ under the condition ① which is provided above.

Done.

3 Question 2 7 / 8

3. a. $\eta(x) = \frac{1}{3}$

$\phi^*(x) = \mathbb{I}(\eta(x) \geq \frac{1}{2}) = \mathbb{I}(\frac{1}{3} \geq \frac{1}{2}) = 0.$

$R(\phi^*(x)) = E\min(\eta(x), 1-\eta(x))$

$= E\min(\frac{1}{3}, \frac{2}{3}) = E(\frac{1}{3}) = \frac{1}{3}$

$\pi_1 = \int_1^{+\infty} \eta(x) \cdot f(x)\, dx$

$= \int_1^{+\infty} \frac{1}{3} \cdot x^{-2}\, dx$

$= -\frac{1}{3} x^{-1} \Big|_1^{+\infty}$

$= 0 - (-\frac{1}{3}) = \frac{1}{3}$

$\pi_0 = \int_1^{+\infty} (1-\eta(x)) \cdot f(x)\, dx$

$= \int_1^{+\infty} \frac{2}{3} \cdot x^{-2}\, dx$

$= -\frac{2}{3} x^{-1} \Big|_1^{+\infty}$

$= 0 - (-\frac{2}{3}) = \frac{2}{3}$

b. $\eta(x) = \frac{1}{x}$

$\phi^*(x) = \mathbb{I}(\eta(x) \geq \frac{1}{2}) = \mathbb{I}(\frac{1}{x} \geq \frac{1}{2}) = \mathbb{I}(X \leq 2)$

(In this problem $x \geq 1$)

$R(\phi^*(x)) = E\min(\eta(x), 1-\eta(x))$

$= E\min(\frac{1}{x}, 1-\frac{1}{x}) = E\min(\frac{1}{x}, \frac{x-1}{x})$

let $\frac{1}{x} \leq 1-\frac{1}{x}$, $\Rightarrow x \geq 2$.

Thus, $R(\phi^*(x)) = E\min(\frac{1}{x}, \frac{x-1}{x})$

$= \int_2^{+\infty} \frac{1}{x} \cdot f(x)\, dx + \int_1^2 \frac{x-1}{x} f(x)\, dx.$

$= \int_2^{+\infty} x^{-3}\, dx + \int_1^2 x^{-2} - x^{-3}\, dx.$

$= -\frac{1}{2}x^{-2}\Big|_2^{+\infty} + (-x^{-1} + \frac{1}{2}x^{-2})\Big|_1^2$

$= 0 - (-\frac{1}{2} \cdot 2^{-2}) + (-2^{-1} + \frac{1}{2} \cdot 2^{-2}) - (-1^{-1} + \frac{1}{2} \cdot 1^{-2})$

$= \frac{1}{4}$

$\pi_1 = \int_1^{+\infty} \eta(x) \cdot f(x)\, dx = \int_1^{+\infty} \frac{1}{x} \cdot \frac{1}{x^2}\, dx = -\frac{1}{2}x^{-2}\Big|_1^{+\infty} = \frac{1}{2}$

$\pi_0 = \int_1^{+\infty} (1-\eta(x)) f(x)\, dx = \int_1^{+\infty} x^{-2} - x^{-3}\, dX = -X^{-1} + \frac{1}{2}X^{-2}\Big|_1^{+\infty} = \frac{1}{2}$

Scanned with CamScanner

4 Question 3 **12 / 12**

4. $f(x) = x \cdot e^{-x}$ for $x \geq 0$.

$f'(x) = e^{-x} - x \cdot e^{-x} = e^{-x} \cdot (1-x)$

when $x \geq 1$, $f'(x) \leq 0$, $f(x)$ decreases
when $0 \leq x < 1$, $f'(x) > 0$, $f(x)$ increases.

That is to say $f(x)$ get the maximum value at $x = 1$

$\max f(x) = f(1) = e^{-1}$, $\arg\max x = 1$

$f(x)$ will get the minimum value at 0, or $+\infty$.

$f(0) = 0 \cdot e^{-0} = 0$

$$\lim_{x \to +\infty} f(x) = \lim_{x \to +\infty} \frac{x}{e^x}$$

$$= \lim_{x \to +\infty} \frac{(x)'}{(e^x)'} \text{ , (L'Hôpital's rule)}$$

$$= \lim_{x \to +\infty} \frac{1}{e^x} = 0$$

Thus, $\min f(x) = f(0) = \lim_{x \to +\infty} f(x) = 0$

$\arg\min f(x) = 0, +\infty$.

Therefore, $\min f(x) = 0$, $\arg\min f(x) = 0, +\infty$
$\max f(x) = e^{-1}$, $\arg\max f(x) = 1$.

Scanned with CamScanner

5 Question 4 7 / 8

5. Becaus, X and Y are independent,

X^2 is independent of Y^2,

X^2 is independent of $\frac{1}{Y^2}$.

Thus.

$$E[X^2/Y^2] = E[X^2 \cdot (\tfrac{1}{Y^2})] = E[X^2] \cdot E[\tfrac{1}{Y^2}]$$

Define $h(u) = \frac{1}{u}$, $u > 0$

$$h'(u) = -u^{-2}$$

$$h''(u) = 2u^{-3} > 0.$$

Therefore, $h(u)$ is a convex function.

According to the Jensen's Inequality,

$$h(E(Y^2)) \le Eh(Y^2)$$

$$\frac{1}{E(Y^2)} \le E\left(\frac{1}{Y^2}\right)$$

Thus, $E[X^2/Y^2] = E(X^2) \cdot E(\frac{1}{Y^2})$

$$\ge E(X^2) \cdot \frac{1}{E(Y^2)} = E(X^2)/E(Y^2)$$

Therefore, $E(X^2/Y^2) \ge EX^2/EY^2$

6 Question 5 8 / 8

6. a. $X = \begin{bmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{bmatrix}$, let design matrix X with ith row x_i^t

b. let $X = [x_{\cdot 1}, x_{\cdot 2}, x_{\cdot 3}, \cdots, x_{\cdot n}]$,
where $x_{\cdot j}$ is the jth column of X.

For $\forall\, c_1, c_2, \cdots, c_n \in \mathbb{R}$,
let $c_1 x_{\cdot 1} + c_2 x_{\cdot 2} + \cdots + c_n x_{\cdot n} = 0$ — ①
Multiply $x_{\cdot 1}$ to both sides of ①

Then $c_1 (x_{\cdot 1})^2 + \sum_{i=2}^{n} c_i (x_{\cdot i})(x_{\cdot 1}) = 0$

Because $x_{\cdot i}, x_{\cdot 1}$ are orthogonal $(i \neq 1)$

Thus, $(x_{\cdot i})(x_{\cdot 1}) = 0 \Rightarrow c_1 (x_{\cdot 1})^2 = 0$
$\Rightarrow c_1 = 0, \ (x_{\cdot 1} \neq 0)$

Similarly, we can multiply $(x_{\cdot i})$ to both sides of ①
and get $c_1 = c_2 = \cdots = c_n = 0$ — ②

Since ② is the only solution of ①

Then all columns of X are linearly independent
showing that X has full rank.

7.1 Part a 1.5 / 2

6. a. $X = \begin{bmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{bmatrix}$, let design matrix X with ith row X_i^t

b. let $X = [x_{\cdot 1}, x_{\cdot 2}, x_{\cdot 3}, \dots, x_{\cdot n}]$,
where $x_{\cdot j}$ is the jth column of X.

For $\forall\, c_1, c_2, \dots, c_n \in \mathbb{R}$,
let $c_1 x_{\cdot 1} + c_2 x_{\cdot 2} + \dots + c_n x_{\cdot n} = 0$ — ①
Multiply $x_{\cdot 1}$ to both sides of ①

Then $c_1 (x_{\cdot 1})^2 + \sum_{i=2}^{n} c_i (x_{\cdot i})(x_{\cdot 1}) = 0$

Because $x_{\cdot i}, x_{\cdot 1}$ are orthogonal $(i \neq 1)$

Thus, $(x_{\cdot i})(x_{\cdot 1}) = 0 \Rightarrow c_1 (x_{\cdot 1})^2 = 0$
$\Rightarrow c_1 = 0, \ (x_{\cdot 1} \neq 0)$

Similarly, we can multiply $(x_{\cdot i})$ to both sides of ①
and get $c_1 = c_2 = \dots = c_n = 0$ — ②

Since ② is the only solution of ①

Then all columns of X are linearly independent
showing that X has full rank.

7.2 Part b 2 / 2

c. $\hat{\beta}^{OLS} = (X^t X)^{-1} X^t y$

$\hat{\beta}^{Ridge}_{\lambda} = (X^t X + \lambda I_p)^{-1} X^t y$

(Full rank of X ensures $X^t X$ is invertible).

d. $\hat{\beta}^{OLS} > \hat{\beta}^{Ridge}_{\lambda}$

7.3 Part c 1.5 / 3

c. $\hat{\beta}^{OLS} = (X^t X)^{-1} X^t y$

$\hat{\beta}_\lambda^{Ridge} = (X^t X + \lambda I_p)^{-1} X^t y$

(Full rank of X ensures $X^t X$ is invertible).

d. $\hat{\beta}^{OLS} > \hat{\beta}_\lambda^{Ridge}$

Scanned with CamScanner

7.4 Part d 2 / 3

Let $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{N\times d}$.

7. The condition of the program is

$C = \{x : \|Bx\|^3 \le t\}$

For $\forall\ x, y \in C$, and $\alpha \in [0,1]$.

$\|B(\alpha x + (1-\alpha)y)\|^3$

$\le (\alpha\|Bx\| + (1-\alpha)\|By\|)^3$

$= \alpha^3 \|Bx\|^3 + 3\alpha^2(1-\alpha)\|Bx\|^2\|By\| + 3\alpha(1-\alpha)^2\|Bx\|\|By\|^2$
$+ (1-\alpha)^3\|By\|^3$

$\le (\alpha^3 + 3\alpha^2(1-\alpha) + 3\alpha(1-\alpha)^2 + (1-\alpha)^3) \cdot t$

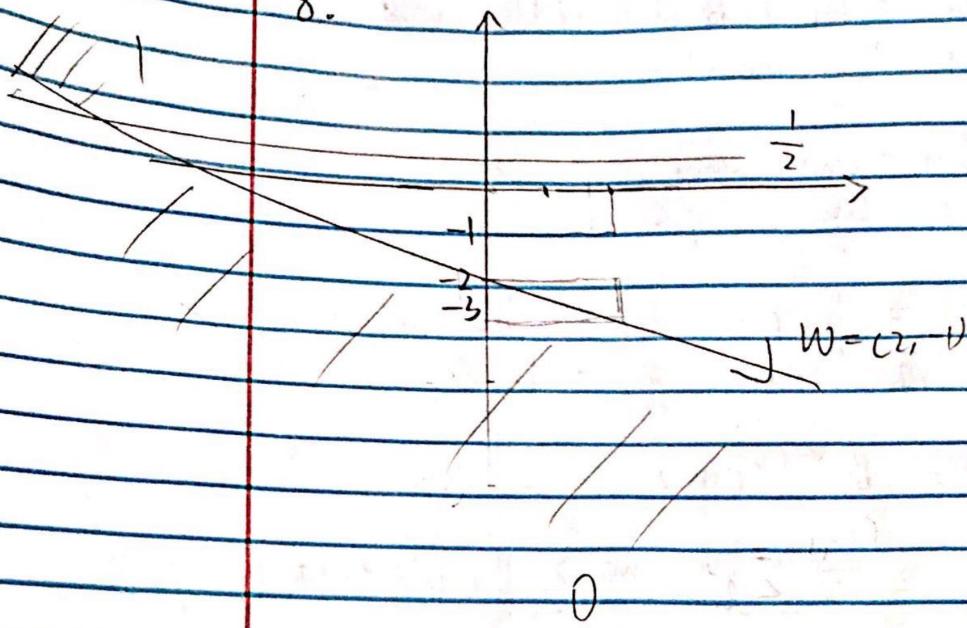
$= (\alpha + (1-\alpha))^3 t = t.$

Thus, $\alpha x + (1-\alpha) y \in C$. That is to say C is a convex set for any $B \in \mathbb{R}^{N\times d}$.

8 Question 7 4 / 8

8.

$g(x) \geq \frac{1}{2}$



9 Question 8 0 / 8

9.

	1	1	1	0	0	1	0	0
	0	2	4	6	8	10	12	14
1-nn classifier =	1	1	1	0	0	0	0	0
3-nn classifier =	1	1	1	1	0	0	0	0

a. 0, 2, 4 are assign to class 1

b. 0, 2, 4, 6 are assign to class 1.

c. Training error of 1-nn rule =

$$\frac{1}{8} = \frac{1}{8}$$

d. Training error of 3-nn rule =

$$\frac{2}{8} = \frac{1}{4}.$$

10.1 Part a 0.5 / 2

9.

	1	1	1	0	0	1	0	0
	0	2	4	6	8	10	12	14
1-nn classifier =	1	1	1	0	0	0	0	0
3-nn classifier =	1	1	1	1	0	0	0	0

a. 0, 2, 4 are assign to class 1

b. 0, 2, 4, 6 are assign to class 1.

c. Training error of 1-nn rule =

$$\frac{1}{8} = \frac{1}{8}$$

d. Training error of 3-nn rule =

$$\frac{2}{8} = \frac{1}{4}.$$

10.2 Part b 0.5 / 2

9.

	1	1	1	0	0	1	0	0
	0	2	4	6	8	10	12	14
1-nn classifier =	1	1	1	0	0	0	0	0
3-nn classifier =	1	1	1	1	0	0	0	0

a. 0, 2, 4 are assign to class 1

b. 0, 2, 4, 6 are assign to class 1.

c. Training error of 1-nn rule =

$$\frac{1}{8} = \frac{1}{8}$$

d. Training error of 3-nn rule =

$$\frac{2}{8} = \frac{1}{4}.$$

10.3 Part c **0 / 2**

9.

	1	1	1	0	0	1	0	0
	0	2	4	6	8	10	12	14
1-nn classifier =	1	1	1	0	0	0	0	0
3-nn classifier =	1	1	1	1	0	0	0	0

a. 0, 2, 4 are assign to class 1

b. 0, 2, 4, 6 are assign to class 1.

c. Training error of 1-nn rule =

$$\frac{1}{8} = \frac{1}{8}$$

d. Training error of 3-nn rule =

$$\frac{2}{8} = \frac{1}{4}.$$

10.4 Part d 2 / 2

10\. a. Devide $D_n = D_{n1} \cup D_{n2} \cup \cdots \cup D_{n5}$ into 5-folds each has $\frac{n}{5}$ points.

Devide $D_m = D_{m1} \cup D_{m2} \cup \cdots \cup D_{m5}$ into 5-folds each has $\frac{m}{5}$ points

let $\hat{\phi}^{l}(x) = \phi(x, c = D_{nl})$ be the rule derived from training set D_{nl}.

b. $ER(\phi) = \hat{R}^{5-CV}(\phi)$

$$= \frac{1}{5} \sum_{l=1}^{5} \left(m^{-1} \sum_{m(l-1) < i \le ml} \mathbb{I}\left(\hat{\phi}^{l}(x_i) \neq y_i\right) \right)$$

Scanned with CamScanner

11.1 Part a 3 / 14

11. $Y = \langle 2V, U \rangle = (2V)^t U = 2V^t U$

$$E(Y) = E(2V^t U) = 2V^t \cdot EU = 0$$

$$Var(Y) = Var(2V^t U) = (2V^t) \cdot Var U \cdot (2V)$$

$$= 4V^t \Sigma V$$

Because U is multinormal, from the definition of multivariate Normal $\langle 2V, U \rangle = 2 \langle V, U \rangle$ is univariate normal.

Therefore, $Y \sim N(0, 4V^t \Sigma V)$

Scanned with CamScanner

12 Question 11 8 / 8