Midterm Exam 1 Review

- 25 multiple choice (select all that apply) questions using scantron (Bring a pencil!!)
- Covers up to (and including) today lecture up to (and including) sec 6.2 in the course notes
- Formula sheet will be provided and is online (no cheat sheet)
- You can bring a calculator
- Similar to assignment questions and examples in course notes

- study tips:
  1) review assignments
  2) review exercises in course notes
  3) understand everything on the formula sheet
  4) do practice midterm exam (online)

- bring 1) pencil 2) eraser 3) calculator
  4) student ID

BE ON TIME!

University of Alberta

GENERAL PURPOSE ANSWER SHEET

NAME (Surname followed by space then Given name)

BUCHLER DIETER

IMPORTANT DIRECTIONS FOR MARKING ANSWERS

Use HB **pencil only**.
Do NOT use ink or ballpoint pens.
Make heavy black marks that fill the circle completely.
Erase cleanly any answer you wish to change.

CORRECT MARK

INCORRECT MARKS

Please note that question numbers 101 to 250 appear on the back

IDENTIFICATION NUMBER								
A	B	C	D	E	F	G	H	I
0	1	2	3	4	5	6	7	8

SPECIAL CODES					
J	K	L	M	N	O
2					

SA1		
P	Q	R

SA2		
S	T	U

1 A B C D E	11 A B C D E	21 A B C D E	31 A B C D E	41 A B C D E
2 A B C D E	12 A B C D E	22 A B C D E	32 A B C D E	42 A B C D E
3 A B C D E	13 A B C D E	23 A B C D E	33 A B C D E	43 A B C D E
4 A B C D E	14 A B C D E	24 A B C D E	34 A B C D E	44 A B C D E
5 A B C D E	15 A B C D E	25 A B C D E	35 A B C D E	45 A B C D E
6 A B C D E	16 A B C D E	26 A B C D E	36 A B C D E	46 A B C D E
7 A B C D E	17 A B C D E	27 A B C D E	37 A B C D E	47 A B C D E
8 A B C D E	18 A B C D E	28 A B C D E	38 A B C D E	48 A B C D E
9 A B C D E	19 A B C D E	29 A B C D E	39 A B C D E	49 A B C D E
10 A B C D E	20 A B C D E	30 A B C D E	40 A B C D E	50 A B C D E
51 A B C D E	61 A B C D E	71 A B C D E	81 A B C D E	91 A B C D E
52 A B C D E	62 A B C D E	72 A B C D E	82 A B C D E	92 A B C D E
53 A B C D E	63 A B C D E	73 A B C D E	83 A B C D E	93 A B C D E
54 A B C D E	64 A B C D E	74 A B C D E	84 A B C D E	94 A B C D E
55 A B C D E	65 A B C D E	75 A B C D E	85 A B C D E	95 A B C D E
56 A B C D E	66 A B C D E	76 A B C D E	86 A B C D E	96 A B C D E
57 A B C D E	67 A B C D E	77 A B C D E	87 A B C D E	97 A B C D E
58 A B C D E	68 A B C D E	78 A B C D E	88 A B C D E	98 A B C D E
59 A B C D E	69 A B C D E	79 A B C D E	89 A B C D E	99 A B C D E
60 A B C D E	70 A B C D E	80 A B C D E	90 A B C D E	100 A B C D E

Math Review

Sets and notation: $\{0,1,2\}, \mathbb{N}, \mathbb{R}$

$\in, \subset, \notin, \not\subset, \cup, \cap, \setminus, \text{set}^c$

set builder notation:

$\{x \in \mathbb{N} \mid x < 3\} = \{0,1,2\}$

$\{f \mid f: \mathbb{R} \to \mathbb{R}\}$

Cartesian products (set of tuples):

$\mathcal{X} \times \mathcal{Y} = \{(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$

$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(a,b,c) \mid a,b,c \in \mathbb{R}\}$

Dot product: $\vec{x} \in \mathbb{R}^d, \vec{w} \in \mathbb{R}^d$

$\vec{x}^T \vec{w} = \left((x_1, \dots, x_d)^T (w_1, \dots, w_d)\right)^T = (x_1, \dots, x_d)(w_1, \dots, w_d)^T$

$= x_1 w_1 + \dots + x_d w_d$

$(x_1, \dots, x_d) \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}$

Functions:

$f: \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2$

$f: \mathbb{R}^2 \to \mathbb{R}, \quad f(\vec{x}) = f(x_1, x_2) = 2x_1 + 3x_2^2$

$\mathcal{A} : (\mathbb{R} \times \mathbb{R})^n \to \{f \mid f : \mathbb{R} \to \mathbb{R}\}, \quad n = 2$

$D = ((1,2),(3,4))$

$\mathcal{A}(D) = f$ where $f(x) = (4-2)x + 2 + 1$

Summation and Integration:

$\mathcal{X} = (x_1, x_2, x_3)$, $f(x) = x^2$

$$\sum_{x \in \mathcal{X}} x = \sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$$

$$\sum_{x \in \mathcal{X}} f(x) = \sum_{i=1}^{3} f(x_i) = f(x_1) + f(x_2) + f(x_3)$$

$\mathcal{Y} = [a, b]$, $f(y) = y^c$

$$\int_{\mathcal{Y}} f(y)\,dy = \int_a^b f(y)\,dy = \int_a^b y^c\,dy = \left.\frac{y^{c+1}}{c+1}\right|_a^b$$

$f(x,y) = xy$, $\mathcal{X} = (x_1, x_2, x_3)$, $\mathcal{Y} = [a, b]$

$$\int_{\mathcal{Y}} \sum_{x \in \mathcal{X}} f(x,y)\,dy = \int_a^b \sum_{i=1}^{3} x_i y\,dy$$

$\int^b$

$$= \int_a x_1 y + x_2 y + x_3 y \, dy$$

$$= \frac{x_1 y^2}{2}\Big|_a^b + \frac{x_2 y^2}{2}\Big|_a^b + \frac{x_3 y^2}{2}\Big|_a^b$$

(Partial) Derivatives:

$f: \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, $f'(x) = \frac{df}{dx}(x) = 2x$, $f''(x) = 2$

$f: \mathbb{R}^2 \to \mathbb{R}$, $f(\vec{x}) = f(x_1, x_2) = x_1^2 + x_2^2$

$\frac{\partial f}{\partial x_1}(x_1) = 2x_1$, $\frac{\partial f}{\partial x_2}(x_2) = 2x_2$ $\quad (2x_1, 2x_2)^T$

Common derivatives and properties on formula sheet

Probability

Outcome space and Events:

$\mathcal{X} = \{\underbrace{1,2,3,4,5,6}_{\text{(possible outcomes)}}\}$ outcome space

$E \subset \mathcal{X}$ event

Probability Distribution:

$\mathbb{P}$ takes as input events and outputs values in $[0,1]$

$\mathbb{P}(E)$ where $E \subset \mathcal{X}$ probability of event E

Random Variable:

$\underset{\text{(r.v.)}}{X} \in \underset{\text{(outcome space)}}{\mathcal{X}}$ and has distribution $\mathbb{P}$

$\mathbb{P}(X \in E) \overset{\text{def}}{=} \mathbb{P}(E)$ where $E \subset \mathcal{X}$

Common notation: If E contains a single outcome $E = \{x\}$ where $x \in \mathcal{X}$. Then $\mathbb{P}(X = x) \overset{\text{def}}{=} \mathbb{P}(X \in \{x\})$

$= \mathbb{P}(\{x\})$

If E is an interval:

$E = [a,b]$ then $\mathbb{P}(a \le X \le b) \stackrel{\text{def}}{=} \mathbb{P}(X \in [a,b])$

$E = [a,\infty)$ then $\mathbb{P}(X \ge a) \stackrel{\text{def}}{=} \mathbb{P}(X \in [a,\infty))$

$E = (-\infty,b]$ then $\mathbb{P}(X \le b) \stackrel{\text{def}}{=} \mathbb{P}(X \in (-\infty,b])$

$E \subset \mathcal{X} = \mathbb{R}$

<u>Discrete r.v.:</u>

$x \in \mathcal{X}$

Countable outcome space $\mathcal{X} = \{2,3,4,5\}$

$x \in \mathcal{X} = \mathbb{R}$

$\mathbb{P}(\{0,1,2\}) = 1$

<u>Continuous r.v.:</u>

$x \in \mathcal{X}$

Uncountable outcome space $\mathcal{X} = \mathbb{R}$

<u>Calculating Probabilities:</u>

If X is discrete: use pmf $p : \mathcal{X} \to [0,1]$

$$\mathbb{P}(X \in E) \stackrel{\text{def}}{=} \sum_{x \in E} p(x) \quad \text{where } E \subset \mathcal{X}$$

If Y is continuous: use pdf $p : \mathcal{X} \to [0,\infty)$

$$\mathbb{P}(X \in E) \stackrel{\text{def}}{=} \int_E p(x)\,dx \quad \text{where } E \subset \mathcal{X}$$

Commonly used discrete and continuous distributions on formula sheet.

Ex: Bernoulli, Normal, Laplace

Multivariate Probability:

$Z = (X, Y) \in \mathcal{X} \times \mathcal{Y}$ r.v.

On formula sheet $E_X \subset \mathcal{X}, E_Y \subset Y$

Joint: $\mathbb{P}(X \in E_X, Y \in E_Y)$

Marginal: $\mathbb{P}_X(X \in E_X)$, $\mathbb{P}_Y(Y \in E_Y)$

Conditional: $\mathbb{P}_{X|Y}(X \in E_X | Y = y)$, $\mathbb{P}_{Y|X}(Y \in E_Y | X = x)$

Product Rule: $p(x, y) = \underbrace{p(y|x)}_{P_{Y|X}(y|x)} p(x) = p(x|y) p(y)$

Bayes' rule

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Independence:

$X = (X_1, \ldots, X_n)$ $X_1, \ldots, X_n$ are independent if:

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n)$$

Functions of r.v.:

A function of a r.v. is a r.v.

If $X \in \mathcal{X}$ is a r.v. then:

$f: \mathcal{X} \to \mathcal{Y}$, $Y = f(X) = X^2$ is a r.v. with outcome space $\mathcal{Y}$

Ex: $\bar{X} = g(X_1, \ldots, X_n) = \frac{X_1 + \ldots + X_n}{n}$

Expectation and Variance:

$Z = (X, Y) \in \mathcal{X} \times \mathcal{Y}$ r.v.

On formula sheet with useful properties

Univariate: $\mathbb{E}[X]$

Function: $\mathbb{E}[f(X)]$

Multivariate: $\mathbb{E}[f(Z)] = \mathbb{E}[f(X, Y)]$

Conditional: $\mathbb{E}[f(Y) \mid X = x]$

Variance: $\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$

Supervised Learning

Dataset:

$D = ((\vec{X}_1, Y_1), \ldots, (\vec{X}_n, Y_n)) \in (\mathcal{X} \times \mathcal{Y})^n$ where

$(\vec{X}_i, Y_i) \sim \mathbb{P}_{\vec{X},Y}$ and independent for all $i \in \{1, \ldots, n\}$

$\mathcal{X} = \mathbb{R}^d$ feature vector (always $\mathbb{R}^d$)

$\mathcal{Y}$ Label or target

Learner:

$\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^n \Rightarrow \{f \mid f : \mathcal{X} \Rightarrow \mathcal{Y}\}$

Predictor:

$f : \mathcal{X} \Rightarrow \mathcal{Y}$

loss function

$\ell : \mathcal{Y} \times \mathcal{Y} \Rightarrow \mathbb{R}$ $\quad \ell(f(\vec{X}), Y)$

Expected loss

$L(f) = \mathbb{E}[\ell(f(\vec{X}), Y)]$ $\quad (\vec{X}, Y) \sim \mathbb{P}_{\vec{X},Y}$

Objective:

Define $\mathcal{A}$ such that $\mathbb{E}[L(\mathcal{A}(D))]$ is small

Regression:

If $\mathcal{Y}$ has a notion of order

Usually $\mathbb{R}$ or an interval $[a,b]$

Use squared or absolute loss

Classification:

$\mathcal{Y}$ does not have a notion of order

Usually finite set like $\{cat, dog, bird\}$

Use 0-1 loss

Learner: ERM

input dataset D

Estimation: Use D to estimate $L(f)$ for all $f \in \mathcal{F}$

call estimate $\hat{L}(f)$

Optimization: pick $\hat{f}$ as the $f \in \mathcal{F}$ that minimizes $\hat{L}(f)$

Estimation:

$X \in \mathcal{X}$ is a r.v. with distribution $\mathbb{P}$

Want to estimate $\mathbb{E}[X] = \mu$

Use n i.i.d. samples from $\mathbb{P}$

$(X_1, \ldots, X_n)$

Sample mean estimate:

$$\hat{\mu} = \bar{X} = g(X_1, \ldots, X_n) = \frac{X_1 + \cdots X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation and Variance:

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X] = \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n]$$

$$\mathrm{Var}[\bar{X}] = \frac{\mathrm{Var}[X]}{n} = \frac{\mathrm{Var}[X_1]}{n} = \cdots = \frac{\mathrm{Var}[X_n]}{n}$$

Optimization

$$\min_{w \in \mathcal{W}} g(w) = g(w^*) \quad \text{where} \quad w^* = \underset{w \in \mathcal{W}}{\operatorname{argmin}}\, g(w)$$

$$w^* = \underset{w \in \mathcal{W}}{\operatorname{argmin}}\, g(w) = \underset{w \in \mathcal{W}}{\operatorname{argmax}}\, -g(w)$$

$$\min_{w \in \mathcal{W}} g(w) = -\left(\max_{w \in \mathcal{W}} -g(w)\right)$$

Solving Optimization problems:

- If $\mathcal{W}$ discrete just compare $g(w)$ for all $w \in \mathcal{W}$
- If $\mathcal{W}$ continuous can use derivatives sometimes

Continuous Optimization:

If $g(w)$ is convex and twice differentiable then:

Cases: 1. If $\mathcal{W} = \mathbb{R}$ then w^* is the solution to $g'(w) = 0$

2. If $\mathcal{W} = [a,b]$ then w^* is the solution to $g'(w) = 0$ if this solution is in $[a,b]$. Otherwise, w^* is either a or b

Twice diffentiable: The second derivative of $g(w)$ written $g''(w)$ exists for all $w \in \mathcal{W}$

Convex: $g(w)$ is convex if $g''(w) \geq 0$ for all $w \in \mathcal{W}$

"Usually $g(w)$ is bowl shaped"

Multidimensional Minimization

If $\mathcal{W} = \mathbb{R}^d$ for $d > 1$, and $g(\vec{w})$ is convex

Note: it is more complicated to check if $g(\vec{w})$ is convex if $d > 1$. So, I will just tell you

Then we calculate

$$\vec{w}^* = (w_1^*, \ldots, w_d^*)^T = \underset{\vec{w} \in \mathcal{W}}{\operatorname{argmin}}\, g(\vec{w})$$

by setting w_j^* as the solution to

$$\frac{\partial g}{\partial w_j}(w_j) = 0 \quad \text{for all } j \in \{1, \dots, d\}$$

$$g(\vec{w}^*) = \min_{\vec{w} \in W} g(\vec{w})$$