Midterm Exam 1 Review

- Exam info in eClass announcement

Study tips:

- Review assignments
- Review exercises in course notes
- Understand everything on the formula sheet

Math Review

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

$\in$, $\subset$, $\notin$, $\not\subset$, $\cup$, $\cap$, $\setminus$, 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 (Sets 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: $\vec{x} \in \mathbb{R}^d$, $\vec{w} \in \mathbb{R}^d$

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

$$= x_1 w_1 + \dots + x_d w_d$$

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:

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

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

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

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

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

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

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

$\int^b$

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

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

(Partial) Derivatives:

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

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

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

Common derivatives and properties on formula sheet

Probability

Outcome space and Events:

$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ 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:

$X \in \mathcal{X}$ and has distribution $\mathbb{P}$

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

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

If E is an interval:

$E = [a,b]$ then $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \in [a,b])$

$E = [a,\infty)$ then $\mathbb{P}(X \geq a) = \mathbb{P}(X \in [a,\infty))$

$E = (-\infty, b]$ then $\mathbb{P}(X \leq b) = \mathbb{P}(X \in (-\infty, b])$

Discrete r.v.:

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

Continuous r.v.:

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

Calculating Probabilities:

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

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

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

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

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

Product Rule: $p(x, y) = p(y|x)\,p(x) = p(x|y)\,p(y)$

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 $\mathcal{Y}$

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

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

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

Expected loss

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

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 $\mathbb{P}$

$(X_1, \dots, X_n)$

Sample mean estimate:

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

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 a or b

Twice doffentiable: 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

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