CS181 Pset 0 Solutions

Due: Never

The goal of these problems is to quickly check your readiness for CS 181. If you find these problems challenging please check out the Section 0 document and/or reach out to one of the TFs. Even if you find this easy, we still recommend reading through the Section 0 notes for review.

Problem 1

Given the matrix $\mathbf{X}$ and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{1}$$

(a) Expand $\mathbf{Xy} + \mathbf{z}$

(b) Expand $\mathbf{y^T X y}$

Solution:

(a)

$$\mathbf{Xy} + \mathbf{z} = \begin{pmatrix} x_{11}y_1 + x_{12}y_2 \\ x_{21}y_1 + x_{22}y_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_{11}y_1 + x_{12}y_2 + z_1 \\ x_{21}y_1 + x_{22}y_2 + z_2 \end{pmatrix}$$

(b)

$$\begin{aligned} \mathbf{y^T X y} &= \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} x_{11}y_1 + x_{21}y_2 & x_{12}y_1 + x_{22}y_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= x_{11}y_1^2 + x_{21}y_1y_2 + x_{12}y_1y_2 + x_{22}y_2^2 \end{aligned}$$

Problem 2

Assume matrix $\mathbf{X}$ has dimensionality (or shape) $(n \times d)$, and vector $\mathbf{w}$ has shape $(d \times 1)$.

(a) What shape is $\mathbf{y} = \mathbf{Xw}$?

(b) What shape is $(\mathbf{X}^T\mathbf{X})^{-1}$?

(c) Using y from part (a), what shape is $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$?

(d) Assume vector $\mathbf{w'} = \mathbf{w}^T$. What shape is $\mathbf{y'} = \mathbf{Xw'}^T$?

Solution:

(a) $(n \times 1)$

(b) $\mathbf{X}^T\mathbf{X}$ has shape $(d \times d)$ and so $(\mathbf{X}^T\mathbf{X})^{-1}$ has shape $(d \times d)$

(c) $\mathbf{X}^T\mathbf{y}$ has shape $(d \times 1)$ and so $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ has shape $(d \times 1)$

(d) Transposing a matrix twice returns the original matrix so we have $\mathbf{y'} = \mathbf{Xw}$, which has shape $(n \times 1)$

Problem 3

Write $\mathbf{u} = \mathbf{u}^{\parallel} + \mathbf{u}^{\perp}$ where $\mathbf{u}^{\parallel} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ is the projection of $\mathbf{u}$ onto $\mathbf{v}$. Verify that $\langle \mathbf{u}^{\parallel}, \mathbf{u}^{\perp}\rangle = 0$ and that $\mathbf{u} = \mathbf{u}^{\parallel}$ if and only if $\mathbf{u}$ is a scaled multiple of $\mathbf{v}$.

Solution: We have $\mathbf{u}^{\perp} = \mathbf{u} - \mathbf{u}^{\parallel} = \mathbf{u} - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$. Then

$$\langle \mathbf{u}^{\parallel}, \mathbf{u}^{\perp}\rangle = \left\langle \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}, \mathbf{u} - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} \right\rangle = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle} \left\langle \mathbf{v}, \mathbf{u} - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} \right\rangle =$$

$$\frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle} \left(\langle \mathbf{v},\mathbf{u}\rangle - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\langle \mathbf{v},\mathbf{v}\rangle \right) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}(\langle \mathbf{v},\mathbf{u}\rangle - \langle \mathbf{u},\mathbf{v}\rangle) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}(\langle \mathbf{v},\mathbf{u}\rangle - \langle \mathbf{v},\mathbf{u}\rangle) = 0,$$

where we note that $\langle \mathbf{v},\mathbf{u}\rangle = \langle \mathbf{u},\mathbf{v}\rangle$ since $\mathbf{u}$ and $\mathbf{v}$ are real vectors.

If $\mathbf{u} = \mathbf{u}^{\parallel} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ then $\mathbf{u}$ is a scaled multiple of $\mathbf{v}$. For the other direction suppose $\mathbf{u} = c\mathbf{v}$ for some $c \in \mathbb{R}$. Then $\langle \mathbf{u},\mathbf{v}\rangle = \langle c\mathbf{v},\mathbf{v}\rangle = c\langle \mathbf{v},\mathbf{v}\rangle \implies \mathbf{u}^{\parallel} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} = \frac{c\langle \mathbf{v},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} = c\mathbf{v} = \mathbf{u}$.

Problem 4

For an invertible matrix $\mathbf{A}$ show that $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$ where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$.

Solution: We have $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ so $|\mathbf{A}\mathbf{A}^{-1}| = |\mathbf{A}| \cdot |\mathbf{A}^{-1}| = |\mathbf{I}| = 1 \implies |\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$, where we use the fact that the determinant factors over products and that $|\mathbf{A}| \neq 0$ since $\mathbf{A}$ is invertible.

Problem 5

Solve the following vector/matrix calculus problems. In all of the below, $\mathbf{x}$ and $\mathbf{w}$ are column vectors (i.e. $n \times 1$ vectors). It may be helpful to refer to *The Matrix Cookbook* by Petersen and Pedersen, specifically sections 2.4, 2.6, and 2.7.

(a) Let $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$. Find $\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\delta}{\delta \mathbf{x}} f(\mathbf{x})$.

Hint: As a first step, you can expand $\mathbf{x}^T\mathbf{x} = (x_1^2 + x_2^2 + ... + x_n^2)$, where $\mathbf{x} = (x_1, ..., x_n)$.

(b) Let $f(\mathbf{w}) = (1 - \mathbf{w}^T\mathbf{x})^2$. Find $\nabla_{\mathbf{w}} f(\mathbf{w}) = \frac{\delta}{\delta \mathbf{w}} f(\mathbf{w})$.

(c) Let $\mathbf{A}$ be a symmetric n-by-n matrix. If $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{w}^T\mathbf{x}$, find $\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\delta}{\delta \mathbf{x}} f(\mathbf{x})$.

Solution:

(a)

$$\begin{aligned}
\nabla(\mathbf{x}^T\mathbf{x}) &= \nabla(x_1^2 + x_2^2 + \ldots + x_n^2) \\
&= \nabla(x_1^2 + x_2^2 + \ldots + x_n^2) \\
&= \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix} \\
&= 2\mathbf{x}
\end{aligned}$$

(b) By the chain rule:

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{w}} &= 2(1 - \mathbf{w}^T\mathbf{x})\frac{\partial}{\partial \mathbf{w}}(1 - \mathbf{w}^T\mathbf{x}) \\
&= -2(1 - \mathbf{w}^T\mathbf{x})\mathbf{x}
\end{aligned}$$

(c) The partial of $\mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x}$ with respect to x_i is:

$$\begin{aligned}
\frac{\partial}{\partial x_i}\mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x} &= \frac{\partial}{\partial x_i}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_i x_j \\
&= \sum_{k \neq i} a_{ik}x_k + \sum_{j \neq i} a_{ji}x_j + 2a_{ii}x_i \\
&= \sum_{k=1}^{n} a_{ik}x_k + \sum_{j=1}^{n} a_{ji}x_j \\
&= \sum_{k=1}^{n} a_{ik}x_k + \sum_{k=1}^{n} a_{ik}x_k \ \text{ since } A \text{ is symmetric} \\
&= 2\sum_{k=1}^{n} a_{ik}x_k
\end{aligned}$$

This is the ith row that results from multiplying $2Ax$. Thus $\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x}$ is $2Ax$.

Since $\frac{\partial}{\partial \mathbf{x}}\mathbf{w}^\mathbf{T}\mathbf{x}$ is $\mathbf{w}$, the total answer is:

$$\mathbf{A}\mathbf{x} + \mathbf{w} \tag{2}$$

Problem 6

Solve the following:

(a) Verify that $\text{Var}(aX + b) = a^2\text{Var}(X)$.

Hint: As a first step, you can expand $Var(aX+b)$ using the definition of variance. Simplify using properties of expectations.

(b) Suppose that $X_1, ..., X_n$ are i.i.d with mean μ and variance σ^2. Let $\bar{X}$ be the mean $\frac{1}{n}\sum_i^n X_i$. Find $E(\bar{X})$ and $Var(\bar{X})$.

Solution:

(a) We have

$$\text{Var}(aX + b) = \mathbb{E}[(aX + b)^2] - (\mathbb{E}[aX + b])^2 = \mathbb{E}[a^2X^2 + 2abX + b^2] - (a\mathbb{E}[x] + b)^2 =$$

$$a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2(\mathbb{E}[X])^2 - 2ab\mathbb{E}[X] - b^2 = a^2(\mathbb{E}[X^2] - (\mathbb{E}[X])^2) = a^2\text{Var}(X).$$

(b)

$$E(\bar{X}) = E(\frac{1}{n}\sum_i^n X_i) = \frac{1}{n}\sum_i^n E(X_i) = \frac{1}{n}\sum_i^n \mu = \frac{1}{n}\cdot n\cdot\mu = \mu$$

$$\text{Var}(\bar{X}) = \text{Var}(\frac{1}{n}\sum_i^n X_i) = \frac{1}{n^2}\sum_i^n \text{Var}(X_i) = \frac{1}{n^2}\sum_i^n \sigma^2 = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$$

Problem 7

Prove or come up with counterexamples for the following statements:

(a) Random variables A and B are conditionally independent given C. Does this imply that A and B are (unconditionally) independent?

(b) Random variables A and B are independent. Does this imply that A and B are conditionally independent given some random variable C?

Solution: (a) No! Suppose we have a fair coin C_1 and an unfair coin C_2 that has Heads on both sides. We will select 1 coin and then flip the coin twice. Let C be the event that we select C_1. Let A be be the event that first flip lands Heads and let B be the event that the second flip lands Heads. Given C we have that A and B are two separate flips of a fair coin, and so the flips are independent given C. However, suppose we do not know which coin has been selected. Then given A has occurred the probability of selecting coin C_1 is $1/3$ and that of selecting coin C_2 is $2/3$. But then $\mathbb{P}(B|A) = \mathbb{P}(B|C_1)\mathbb{P}(C_1|A) + \mathbb{P}(B|C_2)\mathbb{P}(C_2|A) = (1/2)\times 1/3 + (1)\times 2/3 = 5/6 \neq 3/4 = \mathbb{P}(B)$, so A and B are not independent.

(b) No! First, consider two fair, independent coin flips A, B. Let C be the event that $A = B$. On their own, A and B are independent but given C we can determine B from A or A from B (they are either perfectly correlated or perfectly anti-correlated), so A and B are not conditionally independent given C.

Problem 8

Suppose you undergo a test for a disease whose frequency in the population is 1% (i.e. the probability of any given person having the disease is 1%). The test for the disease has a 5% false positive rate (i.e. given that you don't have the disease, there's a 5% chance you test positive) and a 10% false negative rate (i.e. given that you do have the disease, there's a 10% chance that you test negative). Suppose you take the test and it comes back positive. What is the probability that you have the disease?

Solution: Let D be the event that you have the disease and T be the event that you test positive for the disease.
Applying Bayes' Rule and the law of total probability:

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T)} \\ &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\ &= \frac{0.90 \cdot 0.01}{0.90 \cdot 0.01 + 0.05 \cdot 0.99} \\ &\approx 0.15 \end{aligned}$$

Problem 9

Show that for random variables X, Y that $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$.

Solution: We have

$$\text{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 = \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 =$$

$$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

Problem 10

Using the probability density function of $X \sim \mathcal{N}(0,1)$ show that X has mean 0 and variance 1.

Hint: The PDF is $p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$. For the mean, you can reason about the properties of the PDF itself to get the answer without integration techniques. For the variance, use integration by parts with $u = x$ and $dv = xe^{-x^2/2}dx$ and the fact that the PDF itself integrates to 1.

Solution: For the mean, first write down the expression for the mean:

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}}$$

Notice that the PDF is an odd function, (i.e. $f(-x) = -f(x)$), and so the area from $-\infty$ to 0 exactly cancels out the area from 0 to ∞. Thus, $E(X) = 0$.
For the variance, we use LOTUS:

$$\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 = E(X^2) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} x^2 e^{-\frac{x^2}{2}} \text{ since } x^2e^{-x^2/2} \text{ is an even function}
\end{aligned}$$

Now use integration by parts with $u = x$ and $dv = xe^{-x^2/2}dx$, so $du = dx$ and $v = -e^{-x^2/2}$:

$$\begin{aligned}
Var(X) &= \frac{2}{\sqrt{2\pi}}(-xe^{-\frac{x^2}{2}}\Big|_0^{\infty} + \int_0^{\infty} e^{-\frac{x^2}{2}}) \\
&= \frac{2}{\sqrt{2\pi}}(0 + \frac{\sqrt{2\pi}}{2}) \\
&= 1
\end{aligned}$$

When going from the first to the second line, we use the fact that the second term is the normal PDF over half the support and without the $\frac{1}{\sqrt{2\pi}}$ term.

Problem 11

A random point (X, Y, Z) is chosen uniformly in the ball

$$B = \{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$$

(a) Find the joint PDF of (X, Y, Z).

(b) Find the joint PDF of (X, Y).

(c) Write an expression for the maginal PDF of X, as an integral.

Solution:

(a) Let $B = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$ be the closed unit ball. The volume of B is $\mathrm{vol}(B) = \int_B 1 dx\ dy\ dz = \frac{4}{3}\pi$. Since the distribution of (X, Y, Z) is uniform over B the PDF is then

$$f(x, y, z) = \frac{3}{4\pi} \cdot \chi((x, y, z) \in B)$$

where $\chi((x, y, z) \in B) = \begin{cases} 1 & (x, y, z) \in B \\ 0 & (x, y, z) \notin B \end{cases}$

(b) Let $C = \{(x, y) \mid x^2 + y^2 \leq 1\}$ be the unit circle. We have

$$f(x, y) = \int_{\mathbb{R}} f(x, y, z)\ dz = \frac{3}{4\pi} \int_{\mathbb{R}} \chi((x, y, z) \in B)\ dz = \frac{3}{4\pi} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} \chi((x, y) \in C)\ dz =$$

$$\frac{3}{2\pi} \sqrt{1 - x^2 - y^2} \cdot \chi((x, y) \in C).$$

(c) We have

$$f(x) = \int_{\mathbb{R}} f(x, y)\ dy = \frac{3}{2\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1 - x^2 - y^2} \cdot \chi(x \in [-1, 1])\ dy.$$

Problem 12

Suppose we randomly sample a Harvard College student from the undergraduate population. Let X be the indicator of the sampled individual concentrating in computer science, and let Y be the indicator of their working in the tech industry after graduation.

Suppose that the below table represented the joint PMF of X and Y:

	$Y = 1$	$Y = 0$
$X = 1$	$\frac{10}{100}$	$\frac{5}{100}$
$X = 0$	$\frac{15}{100}$	$\frac{70}{100}$

(a) Calculate marginal probability $P(Y = 1)$. In the context of this problem, what does this probability represent?

(b) Calculate conditional probability $P(Y = 1|X = 1)$. In the context of this problem, what does this probability represent?

(c) Are X and Y independent? Why or why not?

Solution:

(a) We have $P(Y = 1) = P(Y = 1, X = 1) + P(Y = 1, X = 0) = \frac{10}{100} + \frac{15}{100} = \frac{1}{4}$. This represents the probability that a Harvard student works in the tech industry after graduation.

(b) Similarly, we compute $P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 0) = \frac{10}{100} + \frac{5}{100} = \frac{3}{20}$. Then we compute $P(Y = 1|X = 1) = \frac{P(Y=1,X=1)}{P(X=1)} = \frac{10/100}{3/20} = \frac{2}{3}$. This represents the probability of a CS concentrator at Harvard working in the tech industry after graduation.

(c) X and Y are not independent since $P(Y = 1|X = 1) \neq P(Y = 1)$.

Problem 13

In her most recent work-from-home shopping spree, Nari decided to buy several house plants. She would like for them to grow as tall as possible, but needs your calculus help to understand how to best take care of them.

(a) After perusing the internet, Nari learns that the height y in mm of her Weeping Fig plant can be directly modeled as a function of the oz of water x she gives it each week:

$$y = -3x^2 + 72x + 70$$

Is this function concave, convex, or neither? Explain why or why not.

(b) Solve analytically for the critical points of this expression. For each critical point, use the second-derivative test to identify if each point is a local max, global max, local min, or global min.

(c) How many oz per week should Nari water her plant to maximize its height? With this much water how tall will her plant grow?

(d) Nari also has a Money Tree plant. The height y in mm of her Money Tree can be directly modeled as a function of the oz of water x she gives it per week:

$$y = -x^4 + 16x^3 - 93x^2 + 230x - 190$$

Is this function concave, convex, or neither? Explain why or why not.

Solution:

(a) It is concave since the 2nd derivative is $y'' = -6 < 0$.

(b) The first derivative is $y' = -6x + 72 = -6(x - 12)$. We have $y' = 0$ if and only if $x = 12$, so $x = 12$ is the only critical point. Since $y' > 0$ for $x < 12$ and $y' < 0$ for $x > 12$ we know that $x = 12$ is a local maximum. Since y is concave ($y'' < 0$) the point $x = 12$ is the global maximum.

(c) She should give her plant 12 oz of water a week for it to achieve the maximum height of 502 mm.

(d) Neither, the 2nd derivative is $y'' = -12x^2 + 96x - 186$, which is negative and positive depending on x.

Credits: Problems 11 and 12 were inspired by Exercise 7.19 and Example 7.1.5 in Blitzstein & Hwang's "Introduction to Probability".