Hyuntaek Oh

Xiaoli Fern

AI 534_001

Oct. 2, 2024

Written Homework Assignment 0 (WA0)

Linear algebra

(a) **Transpose and Associative Property, Positive Semi-definite matrices [2pt]** Define a matrix $B = bb^T$, where $b \in R^{d \times 1}$ is a column vector that is not all-zero. Show that B is a positive semi-definite matrix.

[Hint: To show that B is positive semi-definite, we need to show that B is symmetric, and for any vector $x \in R^{d \times 1}$, $x^T Bx \geq 0$. For the latter, try to get $x^T Bx$ to look like the product of two identical scalars. Note that $b^T x = (x^T b)^T$, that $a^T = a$ for scalar value a, and that matrix multiplication is associative.]

There are two steps to solve this problem based on the hint: the first is that B is symmetric, and the second is that for any vector $x \in R^{d \times 1}$, the condition $x^T Bx \geq 0$ should hold.

1) Prove that B is symmetric

B is symmetric if B has same matrix after transposing it.

$$(B)^T = (bb^T)^T = b^T b = B$$

After transposing B, the matrix B^T is identical to the matrix B, and thus B is symmetric.

2) For any vector $x \in R^{d \times 1}$, the condition $x^T Bx \geq 0$

By using one of the rules of matrix calculations, we can make two groups of matrices like…

$$\begin{aligned} x^T Bx &= x^T (bb^T) x \\ &= (x^T b)(b^T x) \end{aligned}$$

We know that transpose of $b^T x$ is same as

$$b^T x = (x^T b)^T$$

So,

$$\begin{aligned} x^T B x &= x^T(bb^T)x \\ &= (x^T b)(b^T x) \\ &= (b^T x)^2 \end{aligned}$$

For any x, $(b^T x)^2$ is greater than 0 or equal to 0.

Thus, B is a positive semi-definite matrix.

(b) **Solving systems of linear equations with matrix inverse. [2pt]** Consider the following set of linear equations:

$$\begin{cases} x_1 + x_2 - x_3 - x_4 = 1 \\ 2x_1 + 5x_2 - 7x_3 - 5x_4 = -2 \\ 2x_1 - x_2 + x_3 + 3x_4 = 4 \\ 5x_1 + 2x_2 - 4x_3 - 2x_4 = 6 \end{cases}$$

(a) (1 pt) Please express the system of equations as $\boldsymbol{Ax} = \boldsymbol{b}$ by specifying the matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$

A set of linear equations above can be expressed like

$$\begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix}$$

$$\boldsymbol{Ax} = \boldsymbol{b}$$

(b) (1 pt) Solve for $\boldsymbol{Ax} = \boldsymbol{b}$ by using the matrix inverse of A (you can use software to compute the inverse)

$$A^{-1} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & -2 \end{bmatrix}^{-1} = \frac{1}{det(A)} \times C^T = \frac{1}{det(A)} \times \begin{bmatrix} C_{11} & \cdots & C_{14} \\ \vdots & \ddots & \vdots \\ C_{41} & \cdots & C_{44} \end{bmatrix}^T$$

$$\begin{aligned} det(A) = (+1) \times \begin{vmatrix} 5 & -7 & -5 \\ -1 & 1 & 3 \\ 2 & -4 & -2 \end{vmatrix} - (+1) \times \begin{vmatrix} 2 & -7 & -5 \\ 2 & 1 & 3 \\ 5 & -4 & -2 \end{vmatrix} \\ + (-1) \times \begin{vmatrix} 2 & 5 & -5 \\ 2 & -1 & 3 \\ 5 & 2 & -2 \end{vmatrix} - (-1) \times \begin{vmatrix} 2 & 5 & -7 \\ 2 & -1 & 1 \\ 5 & 2 & -4 \end{vmatrix} = 24 \end{aligned}$$

$$[C_{ij}] = +\begin{vmatrix} 5 & -7 & -5 \\ -1 & 1 & 3 \\ 2 & -4 & -2 \end{vmatrix} - \begin{vmatrix} 2 & -7 & -5 \\ 2 & 1 & 3 \\ 5 & -4 & -2 \end{vmatrix} + \begin{vmatrix} 2 & 5 & -5 \\ 2 & -1 & 3 \\ 5 & 2 & -2 \end{vmatrix} - \begin{vmatrix} 2 & 5 & -7 \\ 2 & -1 & 1 \\ 5 & 2 & -4 \end{vmatrix}$$
$$- \begin{vmatrix} 1 & -1 & -1 \\ -1 & 1 & 3 \\ 2 & -4 & -2 \end{vmatrix} + \begin{vmatrix} 1 & -1 & -1 \\ 2 & 1 & 3 \\ 5 & -4 & -2 \end{vmatrix} - \begin{vmatrix} 1 & 1 & -1 \\ 2 & -1 & 3 \\ 5 & 2 & -2 \end{vmatrix}$$
$$+ \begin{vmatrix} 1 & 1 & -1 \\ 2 & -1 & 1 \\ 5 & 2 & -4 \end{vmatrix} + \begin{vmatrix} 1 & -1 & -1 \\ 5 & -7 & -5 \\ 2 & -4 & -2 \end{vmatrix} - \begin{vmatrix} 1 & -1 & -1 \\ 2 & -7 & -5 \\ 5 & -4 & -2 \end{vmatrix}$$
$$+ \begin{vmatrix} 1 & 1 & -1 \\ 2 & 5 & -5 \\ 5 & 2 & -2 \end{vmatrix} - \begin{vmatrix} 1 & 1 & -1 \\ 2 & 5 & -7 \\ 5 & 2 & -4 \end{vmatrix} - \begin{vmatrix} 1 & -1 & -1 \\ 5 & -7 & -5 \\ -1 & 1 & 3 \end{vmatrix}$$
$$+ \begin{vmatrix} 1 & -1 & -1 \\ 2 & -7 & -5 \\ 2 & 1 & 3 \end{vmatrix} - \begin{vmatrix} 1 & 1 & -1 \\ 2 & 5 & -7 \\ 2 & -1 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 1 & -1 \\ 2 & 5 & -7 \\ 2 & -1 & 1 \end{vmatrix}$$
$$= \begin{bmatrix} 12 & 48 & 42 & -6 \\ -4 & 4 & -6 & 6 \\ 0 & 12 & 0 & 12 \\ 4 & -16 & -6 & -6 \end{bmatrix}$$

After transposing the matrix C,

$$C^T = \begin{bmatrix} 12 & -4 & 0 & 4 \\ 48 & 4 & 12 & -16 \\ 42 & -6 & 0 & -6 \\ -6 & 6 & 12 & -6 \end{bmatrix}$$

Now we can calculate the matrix inverse of A

$$A^{-1} = \frac{1}{24} \times \begin{bmatrix} 12 & -4 & 0 & 4 \\ 48 & 4 & 12 & -16 \\ 42 & -6 & 0 & -6 \\ -6 & 6 & 12 & -6 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/6 & 0 & 1/6 \\ 2 & 1/6 & 1/2 & -2/3 \\ 7/4 & -1/4 & 0 & -1/4 \\ -1/4 & 1/4 & 1/2 & -1/4 \end{bmatrix}$$

Thus,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/6 & 0 & 1/6 \\ 2 & 1/6 & 1/2 & -2/3 \\ 7/4 & -1/4 & 0 & -1/4 \\ -1/4 & 1/4 & 1/2 & -1/4 \end{bmatrix} \times \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 11/6 \\ -1/3 \\ 3/4 \\ -1/4 \end{bmatrix}$$

Vector Calculus

(a) Derivatives.[2pt]. Compute the derivative $f'(x)$ for

(a) (1 pts) the logistic (aka sigmoid) function $f(x) = \frac{1}{1+\exp(-x)}$

$$f'(x) = \frac{d}{dx}(1 + e^{-x})^{-1} = (-1) \times \frac{1}{(1 + e^{-x})^2} \times \frac{d}{dx}(1 + e^{-x})$$

$$= (-1) \times \frac{1}{(1+e^{-x})^2} \times (0 + e^{-x}) \times \frac{d}{dx}(-x)$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

(b) (1 pts) $f(x) = \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

$$f'(x) = \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \times \left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)'$$

$$= \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \times \left(-\frac{((x-\mu)^2)'(2\sigma^2) - (x-\mu)^2(2\sigma^2)'}{(2\sigma^2)^2}\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \times \left(-\frac{x-\mu}{\sigma^2}\right)$$

(b) Gradients. [3pt] Compute the gradient $\nabla_x f$ of the following functions. Please clearly specify the dimension of the gradient.

(a) (1pt)

$$f(z) = \log(1+z)\,, z = x^T x, x \in R^D$$

$$f(z)' = \frac{d}{dz} f(z) = \frac{d}{dz}(\log(1+z)) = \frac{1}{1+z}$$

Compute gradient with respect to x.

$$\frac{d}{dz}(z) = \frac{d}{dz}(x^T x) = \frac{d}{dz}(||x||^2) = 2x$$

z is a scalar (dimension 1) since the matrix calculation shows that $(d \times 1) * (1 \times d) = 1 \; by \; 1$ is a scalar value.

Thus,

$$\nabla_x f(z) = \frac{2x}{1 + x^T x} (where \; the \; vector \; dimension \; is \; D \times 1)$$

(b) (2pt)

$$f(z) = \exp\left(-\frac{1}{2}z\right)$$
$$z = g(\boldsymbol{y}) = \boldsymbol{y}^T S^{-1} \boldsymbol{y}$$
$$\boldsymbol{y} = h(\boldsymbol{x}) = \boldsymbol{x} - \mu$$

where $x, \mu \in R^D,\ S = R^{D\times D}$ is a symmetric matrix

$$\frac{d}{dz} f(z) = \frac{d}{dz}\exp\left(-\frac{1}{2}z\right) = -\frac{1}{2}\exp\left(-\frac{1}{2}z\right)$$

Then, we can get $\nabla_y z$

$$\nabla_y z = 2S^{-1}y$$

z is a scalar value (dimension 1) because of quadratic form involving the matrix S^{-1} and vector y

$$\nabla_y f(z) = -\frac{1}{2} * \exp\left(-\frac{1}{2}z\right) \times 2S^{-1}y$$
$$= -\exp\left(-\frac{1}{2}y^T S^{-1} y\right) \times S^{-1}y$$

Due to $y = (x - \mu) \in R^D$, and thus

$$\nabla_x f(z) = -\exp\left(-\frac{1}{2}(x-\mu)^T S^{-1}(x-\mu)\right) * S^{-1}(x-\mu)$$

The dimension of the result is same as x, which is $R^{D\times 1}$

Probability

(a) **Joint, Marginal, and Conditional Probabilities [2pt]** Consider two discrete random variables X and Y with the following joint distribution:

Y	y_1	0.01	0.02	0.03	0.1	0.1
	y_2	0.05	0.1	0.05	0.07	0.2
	y_3	0.1	0.05	0.03	0.05	0.04
		x_1	x_2	x_3	x_4	x_5
				X		

Please compute:

(a) (1 pt) The Marginal distributions $p(x)$ and $p(y)$

$$p(x) = \begin{cases} p(x_1) = 0.01 + 0.05 + 0.1 = 0.16 \\ p(x_2) = 0.02 + 0.1 + 0.05 = 0.17 \\ p(x_3) = 0.03 + 0.05 + 0.03 = 0.11 \\ p(x_4) = 0.1 + 0.07 + 0.05 = 0.22 \\ p(x_5) = 0.1 + 0.2 + 0.04 = 0.34 \end{cases}$$

$$p(y) = \begin{cases} p(y_1) = 0.01 + 0.02 + 0.03 + 0.1 + 0.1 = 0.26 \\ p(y_2) = 0.05 + 0.1 + 0.05 + 0.07 + 0.2 = 0.47 \\ p(y_3) = 0.1 + 0.05 + 0.03 + 0.05 + 0.04 = 0.27 \end{cases}$$

(b) (1 pt) The Conditional distribution $p(x|Y = y_1)$ and $p(y|X = x_3)$

$$p(x_1|Y = y_1) = \frac{p(X = x_1, Y = y_1)}{p(y_1)} = \frac{0.01}{0.26} = \frac{1}{26}$$

$$p(x_2|Y = y_1) = \frac{p(X = x_2, Y = y_1)}{p(y_1)} = \frac{0.02}{0.26} = \frac{2}{26}$$

$$p(x_3|Y = y_1) = \frac{p(X = x_3, Y = y_1)}{p(y_1)} = \frac{0.03}{0.26} = \frac{3}{26}$$

$$p(x_4|Y = y_1) = \frac{p(X = x_4, Y = y_1)}{p(y_1)} = \frac{0.1}{0.26} = \frac{10}{26}$$

$$p(x_5|Y = y_1) = \frac{p(X = x_5, Y = y_1)}{p(y_1)} = \frac{0.1}{0.26} = \frac{10}{26}$$

$$p(y_1|X = x_3) = \frac{p(Y = y_1, X = x_3)}{p(x_3)} = \frac{0.03}{0.11} = \frac{3}{11}$$

$$p(y_2|X = x_3) = \frac{p(Y = y_2, X = x_3)}{p(x_3)} = \frac{0.05}{0.11} = \frac{5}{11}$$

$$p(y_3|X = x_3) = \frac{p(Y = y_3, X = x_3)}{p(x_3)} = \frac{0.03}{0.11} = \frac{3}{11}$$

(b) **Conditional probabilities, Marginalization and Bayes Rule [5pt]** Consider two coins, one is fair and the other one has a 1/10 probability for head. Now you randomly pick one of the coins, and toss it twice. Answer the following questions.

(a) (1pt) What is the probability that you picked the fair coin? What is the probability of the first toss being head?

Probability of picking up fair coin: $\frac{1}{2}$

Probability of the first toss being "Head": $\frac{1}{2}\left(\frac{1}{2} + \frac{1}{10}\right) = \frac{6}{20} (= \frac{3}{10})$

(b) (2pts) If both tosses are heads, what is the probability that you have chosen the fair coin (Hint: you should apply Bayes Rule for this)?

$P(fair|H_1H_2) = \frac{P(H_1H_2|fair) \times P(fair)}{P(H_1H_2)}$

$P(fair) = \frac{1}{2}$

$P(H_1H_2|fair) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$P(H_1H_2) = P(fair) \times P(H_1H_2|fair) + P(biased) \times P(H_1H_2|biased)$

$= \left(\frac{1}{2} \times \frac{1}{4}\right) + \left(\frac{1}{2} \times \frac{1}{100}\right) = \frac{26}{200} = \frac{13}{100}$

Thus,

$$P(fair|H_1H_2) = \frac{25}{26}$$

(c) (2pts) If both tosses are heads, what is the probability that the third coin toss will be head? (you should build on results of c)

$$P(H_3|H_1H_2) = P(H_3|fair, H_1H_2) \times P(fair|H_1H_2) + P(H_3|biased, H_1H_2) \times P(biased|H_1H_2)$$

$= \frac{1}{2} \times \frac{25}{26} + \frac{1}{10} \times (1 - \frac{25}{26})$

$$= \frac{126}{260} \left(= \frac{63}{130}\right)$$

(c) **Linearity of Expectation [2 pt]** A random variable x distributed according to a standard normal distribution (mean zero and unit variance) has the following probability density function (pdf):

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Using the properties of expectations, evaluate the following integral

$$\int_{-\infty}^{\infty} p(x)(ax^2 + bx + c)dx$$

[Hint: This is not a calculus question. The simple solution relies on linearity of expectation and the provided mean/variance of p(x).]

We can apply one of the rules in Linearity of Expectation into the following integral

For example,

$$E[f(x) + g(x)] = E[f(x)] + E[g(x)]$$

$$\int_{-\infty}^{\infty} p(x) * (ax^2 + bx + c)dx$$

$$= \int_{-\infty}^{\infty} p(x) * (ax^2)dx + \int_{-\infty}^{\infty} p(x) * (bx)dx + \int_{-\infty}^{\infty} p(x) * (c)dx$$

$$= aE[x^2] + bE[x] + E[c]$$

$$= a * 1 + b * 0 + c$$

$$= a + c$$

$Var(x) = E[x^2] - (E[x])^2$ is equal to 1 since mean zero and unit variance.

Thus,

$(a + c)$ is answer.

(d) **Cumulative Density Functions / Calculus [2 pt]** X is a continuous random variable over the interval [0,1], show that the following function p is a valid probability density function (PDF) and derive the corresponding cumulative density function (CDF).

$$p(x) = \begin{cases} 4x & 0 \le x \le 1/2 \\ -4x + 4 & 1/2 \le x \le 1 \end{cases}$$

[Hint: Recall that a function is a valid PDF function if it integrates to 1: $\int_{-\infty}^{\infty} p(x)\ dx = 1$. And the cumulative density function (CDF) is defined as $C(x) = P(X \le x)$ or the probability that a sample from p is less than x – which can be computed as $C(x) = \int_{-\infty}^{\infty} p(x)\ dx$. This is a calculus question. But the PDF is a piece-wise linear function, hence it is straightforward.]

1) The following function p is a valid probability density function (PDF)

   For proof, we need to check $\int_0^{\frac{1}{2}} 4x\, dx + \int_{\frac{1}{2}}^{1} -4x + 4\, dx = 1$ since it is a condition to validate PDF.

$$[2x^2]_0^{\frac{1}{2}} + [-2x^2 + 4x]_{\frac{1}{2}}^{1} = \left(2 * \frac{1}{4}\right) + \left(-2 * 1^2 + 4 * 1 - \left(-2 * \frac{1}{4} + 4 * \frac{1}{2}\right)\right)$$

$$= \frac{1}{2} + \frac{1}{2} = 1$$

   Thus, the following function p is a valid probability density function.

2) Derive the corresponding cumulative density function (CDF).

   The conditions of CDF are (1)$C(x) = P(X \le x)$ and (2)$C(x) = \int_{-\infty}^{x} p(x)dx$.

   Then, we divide the range of $p(x)$ into two parts ($0 \le x \le \frac{1}{2}$, $\frac{1}{2} < x \le 1$) respectively.

   For $0 \le x \le \frac{1}{2}$,

$$C(x) = \int_0^x 4t\, dt = [2t^2]_0^x = 2x^2$$

   For $\frac{1}{2} < x \le 1$,

$$C(x) = \int_0^{\frac{1}{2}} 4t\, dt + \int_{\frac{1}{2}}^{x} -4t + 4\, dt = [2t^2]_0^{\frac{1}{2}} + [-2t^2 + 4t]_{\frac{1}{2}}^{x}$$

$$= 2 * \frac{1}{4} + (-2x^2 + 4x) - \left(-2 * \frac{1}{4} + 4 * \frac{1}{2}\right)$$

$$= -2x^2 + 4x - 1$$

Thus, we can derive the corresponding CDF.

$$C(x) = \begin{cases} 2x^2 & 0 \leq x \leq 1/2 \\ -2x^2 + 4x - 1 & 1/2 < x \leq 1 \end{cases}$$