AI 669: Machine Learning

Assignment 0

Name: Nam Nguyen

ID: 934-422-327

I. Linear Algebra

1.

a)

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$[3 \times 2] \qquad [3 \times 3]$$

The dimension of the first matrix is 3×2

The dimension of the second matrix is 3×3

Hence, we cannot compute this matrix product

b)

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 5 \\ 10 & 9 & 11 \\ 16 & 15 & 17 \end{bmatrix}$$

$$[3 \times 3] \qquad [3 \times 3] \qquad [3 \times 3]$$

c)

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 4 & 1 & -1 & -4 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 1 & -1 \\ 2 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ -21 & 2 \end{bmatrix}$$

$$[2 \times 4] \qquad [4 \times 2] \qquad [2 \times 2]$$

d) $\underline{x} = [1 \ 0 \ 1]^T = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

- $|\underline{x}|_2 = \sqrt{1^2 + 0^2 + 1^2} = \sqrt{2}$

- $\underline{x}^T \underline{x} = [1 \ 0 \ 1] \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = (1)(1) + (0)(0) + (1)(1)$

  $= 2$

- $\underline{x}\,\underline{x}^T = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} [1 \ 0 \ 1] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

2.

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

a) $\underset{\sim}{A} \, \underline{x} = \underline{b}$

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

$$\Rightarrow \quad \underset{\sim}{A} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & -2 \end{bmatrix}; \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}; \quad \underline{b} = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix}$$

b)

$$\underset{\sim}{A}\,\underline{x} = \underline{b}$$

Since $\underset{\sim}{A}$ is invertible matrix, we have

$$\underline{x} = \underset{\sim}{A}^{-1}\,\underline{b} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1.83 \\ -0.33 \\ 0.75 \\ -0.25 \end{bmatrix}$$

II. Vector Calculus

1.

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

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

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

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

$$\Rightarrow f'(x) = e^{\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)} \cdot \left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)'$$

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

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

2.

a) $f(z) = \log(1+z)$, $z = \underline{x}^T \underline{x}$, $\underline{x} \in \mathbb{R}^D$

$= \log(1 + \underline{x}^T \underline{x})$

$\Rightarrow \nabla_{\underline{x}} f = \dfrac{1}{1+\underline{x}^T\underline{x}} \nabla_{\underline{x}}(1 + \underline{x}^T \underline{x}) = \dfrac{2\underline{x}}{1+\underline{x}^T\underline{x}}$

b) $f(z) = e^{(-\frac{1}{2} z)}$

$z = g(\underline{y}) = \underline{y}^T \underset{\sim}{S}^{-1} \underline{y}$;

$\underline{y} = h(\underline{x}) = \underline{x} - \underline{\mu}$

$\begin{cases} \underline{x}, \underline{\mu} \in \mathbb{R}^D \\ \underset{\sim}{S} \in \mathbb{R}^{D \times D} \end{cases}$

$\underset{\sim}{S}$ is symmetric

$\Rightarrow f(z) = e^{(-\frac{1}{2} \cdot \underline{y}^T \underset{\sim}{S}^{-1} \underline{y})}$

$= e^{(-\frac{1}{2} \cdot (\underline{x} - \underline{\mu})^T \underset{\sim}{S}^{-1} (\underline{x} - \underline{\mu}))}$

$\Rightarrow \nabla_{\underline{x}} f = e^{(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \underset{\sim}{S}^{-1} (\underline{x} - \underline{\mu}))}$

$\cdot \nabla_{\underline{x}} \left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \underset{\sim}{S}^{-1} (\underline{x} - \underline{\mu})\right)$

$= e^{(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \underset{\sim}{S}^{-1} (\underline{x} - \underline{\mu}))}$

$\cdot \left(-\frac{1}{2}\right) 2 \underset{\sim}{S}^{-1} (\underline{x} - \mu)$

$$= - \underset{\sim}{S}^{-1} (\underline{x} - \underline{\mu}) \, e^{\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \underset{\sim}{S}^{-1} (\underline{x} - \underline{\mu})\right)}$$

III. Probability

1.

a) The marginal distribution,

$p(\cancel{x}) \cancel{=}$

- $p(X = x_1) = p(X = x_1, Y = y_1) + p(X = x_1, Y = y_2)$

$$+ p(X = x_1, Y = y_3)$$

$$= \sum_{j=1}^{3} p(X = x_1, Y = y_j) = 0.01 + 0.05 + 0.1$$

$$= 0.16$$

Similarly,

$p(X = x_2) = 0.02 + 0.1 + 0.05 = 0.17$

$p(X = x_3) = 0.03 + 0.05 + 0.03 = 0.11$

$p(X = x_4) = 0.1 + 0.07 + 0.05 = 0.22$

$p(X = x_5) = 0.1 + 0.2 + 0.04 = 0.34$

- $p(Y = y_1) = \sum_{i=1}^{\cancel{4}5} p(X = x_i, Y = y_1)$

$$= 0.01 + 0.02 + 0.03 + 0.1 + 0.1$$

$$= 0.26$$

Similarly,

$$p(Y=y_2) = 0.05 + 0.1 + 0.05 + 0.07 + 0.2 = 0.47$$

$$p(Y=y_3) = 0.1 + 0.05 + 0.03 + 0.05 + 0.04 = 0.27$$

b) The conditional distribution,

- $p(x \mid Y=y_1) = p(X=x_1 \mid Y=y_1) + p(X=x_2 \mid Y=y_1) + p(X=x_3 \mid Y=y_1) + p(X=x_4 \mid Y=y_1) + p(X=x_5 \mid Y=y_1)$

$$= \frac{p(X=x_1, Y=y_1)}{p(Y=y_1)} + \frac{p(X=x_2, Y=y_1)}{p(Y=y_1)} + \frac{p(X=x_3, Y=y_1)}{p(Y=y_1)} + \frac{p(X=x_4, Y=y_1)}{p(Y=y_1)} + \frac{p(X=x_5, Y=y_1)}{p(Y=y_1)}$$

$$= \frac{1}{0.26}\left(0.01 + 0.02 + 0.03 + 0.1 + 0.1\right)$$

$$= 1.$$

- $p(y \mid X=x_3) = \dfrac{\sum_{j=1}^{3} p(Y=y_j, X=x_3)}{p(X=x_3)} = 1.$

2.

First coin is fair $\begin{cases} P_{1,\,\text{head}} = \frac{1}{2} \\ P_{1,\,\text{tail}} = \frac{1}{2} \end{cases}$

Second coin is biased $\begin{cases} P_{2,\,\text{head}} = \frac{1}{10} \\ P_{2,\,\text{tail}} = \frac{9}{10} \end{cases}$

a) The probability that we picked the fair coin is $\frac{1}{2}$

b) The probability of the first toss being head is

$$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{\cancel{2}\;10}\right) = 0.3$$

c)

We have,

$$P(\text{fair coin} \mid \text{both tosses are heads}) = \frac{P(\text{both tosses are heads} \mid \text{fair coin}) \cdot P(\text{fair coin})}{P(\text{both tosses are heads})}$$

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

$$P(\text{both toss are heads} \mid \text{fair coin}) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \left(\frac{1}{4}\right)$$

$$P(\text{both toss are heads}) = \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{100}\right) = 0.13$$

$$\Rightarrow P(\text{fair coin} \mid \text{both tosses are heads}) = \frac{(1/4)(1/2)}{0.13} = 0.9615$$

8

d)

- Similar to section c, we have

P (biased coin | both tosses are heads)

$$= \frac{P(\text{both tosses are heads} \mid \text{biased coin})\, P(\text{biased coin})}{P(\text{both tosses are heads})}$$

$$= \frac{\left(\frac{1}{10}\right)\left(\frac{1}{10}\right) \cdot \left(\frac{1}{2}\right)}{0.13} = 0.0385$$

- If both tosses are heads, the probability that the third coin toss is head, is

$P(\text{fair coin} \mid \text{both tosses are heads})\,(1/2)$

$+ \; P(\text{biased coin} \mid \text{both tosses are heads})\,(1/10)$

$= (0.9615) \cdot (1/2) + (0.0385)\,(1/10) = 0.4846$