CSE 847 Home Assignment 1

Submitted by: Ritam Guha (MSU ID: guharita)

Date: February 9, 2021

---

1 Introduction

Questions in the textbook Pattern Recognition and Machine Learning:

Response:

1. (10 points) Page 58, Question 1.3

   1.1) In the question, we have three colored boxes: red, blue and green. The distribution of different fruits in the given boxes and the probabilities of selecting the boxes are shown in Table 1.

Box Color	Probability of Selection	Apples	Oranges	Limes
Red	0.2	3	4	3
Blue	0.2	1	1	0
Green	0.6	3	3	4

Table 1: Representation of fruit distribution among Red, Blue and Green boxes.

The probability of selecting an apple is:

$P(F = a)$

$= \sum_{x \epsilon \{r,g,b\}} P(F = a|B = x)P(B = x)$

$= (3/10) \times 0.2 + (1/2) \times 0.2 + (3/10) * 0.6$

$= (0.3 \times 0.2) + (0.5 \times 0.2) + (0.3 \times 0.6)$

$= 0.06 + 0.1 + 0.18$

$= 0.34$ **(Ans-1.1.1)**

The second part of the question asks about the conditional probability of selecting a green box if the removed fruit was an orange. So, we have to find the probability: $P(B = g|F = o)$.

$P(B = g|F = o)$
$= \dfrac{P(F = o|B = g)P(B = g)}{P(F = o)}$ (using Bayes' Theorem)
$= \dfrac{(3/10) \times 0.6}{(4/10) \times 0.2 + (1/2) \times 0.2 + (3/10) \times 0.6}$
$= \dfrac{0.3 \times 0.6}{0.4 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6}$
$= \dfrac{0.18}{0.08 + 0.1 + 0.18}$
$= \dfrac{0.18}{0.36}$
$=0.5$ **(Ans-1.1.2)**

---

2. (10 points) Page 59, Question 1.6

1.2) **To Proof:** If two variables x and y are independent, their covariance is zero.

Proof: The expression for covariance between two variables can be represent as follows:

$cov[x, y]$

$= E_{x,y}[\{x - E[x]\}\{y - E[y]\}]$

$= E_{x,y}[xy - xE[y] - yE[x] + E[x]E[y]]$

$= E_{x,y}[xy] - 2E[x]E[y] + E[x]E[y]$

$= E_{x,y}[xy] - E[x]E[y]$

$= \sum_{x,y} xyP(x, y) - \sum_x xP(x) \times \sum_y yP(y)$

$= \sum_x xP(x) \times \sum_y yP(y) - \sum_x xP(x) \times \sum_y yP(y)$

(For independent variables $x, y : P(x, y) = P(x)P(y)$)

$= 0$

Thus, it is proved that covariance of two independent variables is always zero.

---

3. (20 points) Page 59, Question 1.11

1.3) The log likelihood can be represented by the following expression:

$$\ln P(x|\mu, \sigma^2) = -\frac{1}{2\sigma^2} \textstyle\sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln (2\pi)$$

In order to find maximum log likelihood, we need to differentiate this expression with respect to μ and σ^2 and set the respective results of the derivations equal to 0.

Differentiating the expression w.r.t μ, we get:

$$\frac{\partial \ln P(x|\mu, \sigma^2)}{\partial \mu} = 0$$
$$\Rightarrow -\frac{1}{2\sigma^2} \times \sum_{n=1}^{N} 2(x_n - \mu)(-1) = 0$$
$$\Rightarrow \sum_{n=1}^{N} (\mu - x_n) = 0$$
$$\Rightarrow N \times \mu - \sum_{n=1}^{N} x_n = 0$$
$$\Rightarrow \mu = \frac{1}{N} \sum_{n=1}^{N} x_n$$

Hence, for maximum likelihood the value of μ should be: $\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n$

Differentiating the expression w.r.t σ^2, we get:

$$\frac{\partial \ln P(x|\mu, \sigma^2)}{\partial \sigma^2} = 0$$
$$\Rightarrow -\frac{1}{2} \times \frac{-1}{\sigma^4} \sum_{n=1}^{N} (x_n - \mu_{ML})^2 - \frac{N}{2\sigma^2} = 0$$
$$\Rightarrow \frac{1}{2\sigma^4} (\sum_{n=1}^{N} (x_n - \mu_{ML})^2 - N\sigma^2) = 0$$
$$\Rightarrow N\sigma^2 = \sum_{n=1}^{N} (x_n - \mu_{ML})^2$$

$$\Rightarrow \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$$

Hence, for maximum likelihood the value of σ^2 should be: $\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$

---

2 Linear Algebra I

1. (10 points) Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}, C = \begin{bmatrix} 3 & -1 & 3 \\ 4 & 1 & 5 \\ 2 & 1 & 3 \end{bmatrix}, D = \begin{bmatrix} 2 & -4 & 5 \\ 0 & 1 & 4 \\ 3 & 2 & 1 \end{bmatrix}, E = \begin{bmatrix} 3 & -2 \\ 2 & 4 \end{bmatrix}.$$

If possible, compute the following:

Response:

(a) $(2A)^T$

$$= \left(2 \times \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix}\right)^T = \begin{bmatrix} 2 & 4 & 6 \\ 4 & 2 & 8 \end{bmatrix}^T = \begin{bmatrix} 2 & 4 \\ 4 & 2 \\ 6 & 8 \end{bmatrix}$$

(b) $(A - B)^T$

The dimension of matrix A is: 2×3

The dimension of matrix B is: 3×2

For subtraction operator, the operand matrices should be same dimensional which is not the case over here. So, this operation is not possible.

(c) $(3B^T - 2A)^T$

$$= \left(3 \times \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}^T - 2 \times \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix}\right)^T$$

$$= \left(3 \times \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix} - 2 \times \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix}\right)^T$$

$$= \left(\begin{bmatrix} 3 & 6 & 9 \\ 0 & 3 & 6 \end{bmatrix} - \begin{bmatrix} 2 & 4 & 6 \\ 4 & 2 & 8 \end{bmatrix}\right)^T$$

$$= \begin{bmatrix} 1 & 2 & 3 \\ -4 & 1 & -2 \end{bmatrix}^T$$

$$= \begin{bmatrix} 1 & -4 \\ 2 & 1 \\ 3 & -2 \end{bmatrix}$$

(d) $(-A)^T E$

$$= \begin{bmatrix} -1 & -2 & -3 \\ -2 & -1 & -4 \end{bmatrix}^T \begin{bmatrix} 3 & -2 \\ 2 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -2 \\ -2 & -1 \\ -3 & -4 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 2 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} (-1) \times 3 + (-2) \times 2 & (-1) \times (-2) + (-2) \times 4 \\ (-2) \times 3 + (-1) \times 2 & (-2) \times (-2) + (-1) \times 4 \\ (-3) \times 3 + (-4) \times 2 & (-3) \times (-2) + (-4) \times 4 \end{bmatrix}$$

$$= \begin{bmatrix} -7 & -6 \\ -8 & 0 \\ -17 & -10 \end{bmatrix}$$

(e) $(C + D^T + E)^T$

The dimension of matrix C is: 3×3

The dimension of matrix D^T is: 3×3

The dimension of matrix E is: 2×2

For summation, all the matrices involved in the operation should be same dimensional matrices, but we can see in the operation that E is not same dimensional as the other two matrices. So, it's not possible to compute the expression.

---

2. (10 points) Which of the following are subspace of $\mathbb{R}^2$? Justify your answer.

Response:

S is a subspace of $\mathbb{R}^m$, if and only if $\alpha s_1 + \beta s_2 \in S$, for any $s_1, s_2 \in S$ and any scalars α, β. This definition of subspace can be used to verify whether the following spaces are subspaces of $\mathbb{R}^2$. For each of the spaces, two points will be considered with the coordinates $s_1 = (x_1, y_1)$ and $s_2 = (x_2, y_2)$. Then for these spaces to become valid subspaces of $\mathbb{R}^2$, $\alpha s_1 + \beta s_2$ should belong to the spaces.

(a) $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 0\}$

According to the definition of the space,

$x_1^2 + y_1^2 = 0$

$x_2^2 + y_2^2 = 0$

These equations are possible, only when $s_1 = (0, 0)$ and $s_2 = (0, 0)$ [sum of two squared numbers is equal to zero only when the numbers are zeros]. So, this space consists of only one point (0,0).

That gives us, $x_1 = 0, y_1 = 0, x_2 = 0, y_2 = 0$.

The coordinates of $\alpha s_1 + \beta s_2$ are $(\alpha x_1 + \beta x_2, \alpha y_1 + \beta y_2)$.

$(\alpha x_1 + \beta x_2)^2 + (\alpha y_1 + \beta y_2)^2$

$= 0 + 0$

$= 0$

So, $\alpha s_1 + \beta s_2 \in \mathbb{R}^2$ **which makes this space a subspace of $\mathbb{R}^2$**.

(b) $\{(x, y) \in \mathbb{R}^2 | x^2 - y^2 = 0\}$

From the definition of the space,

$x_1^2 - y_1^2 = 0$

$x_2^2 - y_2^2 = 0$

$(\alpha x_1 + \beta x_2)^2 - (\alpha y_1 + \beta y_2)^2$

$= \alpha^2(x_1^2 - y_1^2) + \beta^2(x_2^2 - y_2^2) + 2\alpha\beta(x_1x_2 - y_1y_2)$

$= 0 + 0 + 2\alpha\beta(x_1x_2 - y_1y_2)$

$= 2\alpha\beta(x_1x_2 - y_1y_2)$

$\neq 0\ \forall s_1, s_2 \in S$ and for any scalars α, β

So, $\alpha s_1 + \beta s_2 \notin \mathbb{R}^2$. **This space is not a subspace of $\mathbb{R}^2$.**

For example, if we consider $s_1 = (5, -5), s_2 = (4, 4), \alpha = 0.2, \beta = 0.5$, we will get $\alpha s_1 + \beta s_2 = (3, 1)$ which does not belong to the space.

(c) $\{(x, y) \in \mathbb{R}^2 | x^2 - y = 0\}$ According to the definition of the space,

$x_1^2 - y_1 = 0$

$x_2^2 - y_2 = 0$

$(\alpha x_1 + \beta x_2)^2 - (\alpha y_1 + \beta y_2)$

$= \alpha^2 x_1^2 + \beta^2 x_2^2 + 2\alpha\beta x_1 x_2 - \alpha y_1 - \beta y_2$

$\neq 0\ \forall s_1, s_2 \in S$ and for any scalars α, β

So, $\alpha s_1 + \beta s_2 \notin \mathbb{R}^2$. **This space is not a subspace of $\mathbb{R}^2$.**

For example, if we consider $s_1 = (5, 25), s_2 = (-5, 25), \alpha = 0.2, \beta = 0.2$, we will get $\alpha s_1 + \beta s_2 = (0, 10)$ which does not belong to the space.

(d) $\{(x, y) \in \mathbb{R}^2 | x - y = 0\}$

According to the definition of the space,

$x_1 - y_1 = 0$

$x_2 - y_2 = 0$

$(\alpha x_1 + \beta x_2) - (\alpha y_1 + \beta y_2)$

$= \alpha(x_1 - y_1) + \beta(x_2 - y_2)$

$= 0 + 0$

$= 0$

So, $\alpha s_1 + \beta s_2 \in \mathbb{R}^2$ **which makes this space a subspace of $\mathbb{R}^2$**.

(e) $\{(x, y) \in \mathbb{R}^2 | x - y = 1\}$

According to the definition of the space,

$x_1 - y_1 = 1$

$x_2 - y_2 = 1$

$(\alpha x_1 + \beta x_2) - (\alpha y_1 + \beta y_2)$

$= \alpha(x_1 - y_1) + \beta(x_2 - y_2)$

$= \alpha + \beta$

$\neq 1 \ \forall s_1, s_2 \in S$ and for any scalars α, β

So, $\alpha s_1 + \beta s_2 \notin \mathbb{R}^2$. **This space is not a subspace of $\mathbb{R}^2$.**

For example, if we consider $s_1 = (5, 4), s_2 = (8, 7), \alpha = 0.2, \beta = 0.4$, we will get $\alpha s_1 + \beta s_2 = (4.2, 3.6)$ which does not belong to the space.

---

3. (10 points) Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix}.$$

Is $AB = BA$? Justify your answer.

Response:

AB

$$= \left(\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} 1 \times 2 + 2 \times (-3) & 1 \times (-1) + 2 \times 4 \\ 3 \times 2 + 2 \times (-3) & 3 \times (-1) + 2 \times 4 \end{bmatrix} \right)$$

$$= \begin{bmatrix} -4 & 7 \\ 0 & 5 \end{bmatrix}$$

BA

$$= \left(\begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} 2 \times 1 + (-1) \times 3 & 2 \times 2 + (-1) \times 2 \\ (-3) \times 1 + 4 \times 3 & (-3) \times 2 + 4 \times 2 \end{bmatrix} \right)$$

$$= \begin{bmatrix} -1 & 2 \\ 9 & 2 \end{bmatrix}$$

Comparing the values of AB and BA, it can be seen that $AB \neq BA$.

---

4. (a) (10 points) Let A be an $m \times n$ matrix with a row consisting entirely of zeros. Show that if B is an $n \times p$ matrix, then AB has a row of zeros.

Response: A is an $m \times n$ dimensional matrix, B is an $n \times p$ dimensional matrix. So, AB will be an $m \times p$ dimensional matrix.

If we consider a_{ij}, b_{ij} and p_{ij} to be the element of the i^th row and j^th column in A, B and P (where $P = AB$) respectively, then we have the following relation:

$p_{ij} = \sum_{k=1}^{n} a_{ik} \times b_{kj}$

So, if A has a row (say r) consisting entirely of zeros, then $a_{rj} = 0$ for $j = 1, \ldots, n$. Now, if we concentrate on the r^{th} row of P, then we will see:

$p_{rj} = \sum_{k=1}^{n} a_{rk} \times b_{kj}$

This will always be zero as $a_{rk} = 0$. Thus, we can see that if A has a row consisting entirely of zeros. AB will also have a row of zeros only.

(b) (10 points) Let A be an $m \times n$ matrix with a column consisting entirely of zeros, and let B be $p \times m$. Show that BA has a column of zeros.

Response: A is an $m \times n$ dimensional matrix, B is an $p \times m$ dimensional matrix. So, BA will be an $p \times n$ dimensional matrix.

If we consider a_{ij}, b_{ij} and p_{ij} to be the element of the i^th row and j^th column in A, B and P (where $P = BA$) respectively, then we have the following relation:

$p_{ij} = \sum_{k=1}^{m} b_{ik} \times a_{kj}$

So, if A has a column (say c) consisting entirely of zeros, then $a_{ic} = 0$ for $i = 1, \ldots, m$. Now, if we concentrate on the c^{th} column of P, then we will see:

$p_{ic} = \sum_{k=1}^{n} b_{ik} \times a_{kc}$

This will always be zero as $a_{kc} = 0$. Thus, we can see that if A has a column consisting entirely of zeros. BA will also have a column of zeros only.

---

5. (10 points) Let $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$ be two matrices. Show that $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$.

Response:

Let's consider A_i and B_j to be the i^{th} row of A and j^{th} column of B. Then we can write the expression AB in two ways:

$$AB = \begin{bmatrix} A_1B \\ A_2B \\ A_3B \\ \vdots \\ A_mB \end{bmatrix} = \begin{bmatrix} AB_1 & AB_2 & AB_3 & \ldots & AB_n \end{bmatrix}$$

In the first representation, where the product AB is represented as A's row-wise multiplication with B, we can see that each of the terms of the form A_iB can be treated as some scalar multiplication of B's column vectors. So, all of them can be represented as linear combinations of the independent column vectors in B, except the situations where $A_i = \vec{0}$. In this way, the number of independent columns in AB gets decreased in comparison to B.

$$rank(AB) \leq rank(B) \tag{1}$$

Similarly for the second representation, where the product is represented as A's multiplication with B's columns, we can consider each AB_j to be some scalar multiplications of A's row vectors. In this case, they can be represented as linear combinations of the independent row vectors in A, except the situations where $B_j = \vec{0}$.

$$rank(AB) \leq rank(A) \tag{2}$$

Combining the Equations 1 and 2, we can successfully conclude that the rank of the product AB should be less than or equal to the rank of A and B. So,

$$rank(AB) \leq min\{rank(A), rank(B)\}$$
