CSE 847 (Spring 2021): Machine Learning— Homework 1

Instructor: Jiayu Zhou

1 Introduction

Questions in the textbook Pattern Recognition and Machine Learning:

1. Page 58, Question 1.3

   Solution: Let us denote apples, oranges and limes by a, o and l respectively.

   The marginal probability of selecting an apple is given by:

$$\begin{aligned} p(a) &= p(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g) \\ &= 0.3 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6 = 0.34 \end{aligned}$$

   The conditional probability $p(g|o)$ is given by:

$$\begin{aligned} p(g|o) &= \frac{p(o|g)p(g)}{p(o)} = \frac{p(o|g)p(g)}{p(o|r)p(r) + p(o|b)p(b) + p(o|g)p(g)} \\ &= \frac{0.3 \times 0.6}{0.4 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6} = 0.5 \end{aligned}$$

2. Page 59, Question 1.6 **Solution:** The definition of covariance is $\mathrm{cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$. Given the fact that $p(x, y) = p(x)p(y)$ when x and y are independent, we obtain:

$$\mathbb{E}[xy] = \sum_x \sum_y p(x, y)xy = \sum_x p(x)x \sum_y p(y)y = \mathbb{E}[x]\mathbb{E}[y]$$

3. Page 59, Question 1.11

   Solution: Let

$$\ell = \ln p(\mathbf{X}|\mu, sigma^2) = -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi).$$

   By standard rules of differentiation we get

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{n=1}^{N} (x_n - \mu).$$

   Set this to zero, and we get

$$\frac{1}{\sigma^2} \sum_{n=1}^{N} x_n = \frac{1}{\sigma^2} N\mu \Rightarrow \mu = \frac{1}{N} \sum_{n=1}^{N} x_n$$

   Similarly

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2(\sigma^2)^2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \frac{1}{\sigma^2}.$$

Setting to zero, we get

$$\frac{N}{2}\frac{1}{\sigma^2} = \frac{1}{2(\sigma^2)^2}\sum_{n=1}^{N}(x_n - \mu)^2.$$

Substitute μ_{ML}, we then get

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

2 Linear Algebra I

1. 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:

(a) $(2A)^T$

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

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

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

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

Solution:

(a) $(2A)^T = \begin{bmatrix} 2 & 4 \\ 4 & 2 \\ 6 & 8 \end{bmatrix}$

(b) Not possible, dimensions do not agree.

(c) $(3B^T - 2A)^T = \begin{bmatrix} 1 & -4 \\ 2 & 1 \\ 3 & -2 \end{bmatrix}$

(d) $(-A)^T E = \begin{bmatrix} -7 & -6 \\ -8 & 0 \\ -17 & -10 \end{bmatrix}$

(e) Not possible, dimensions do not agree.

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

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

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

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

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

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

Solution:

(a) It is a subspace of $\mathbb{R}^2$, as the set can be equivalently expressed as $\{(0,0)\}$.

(b) Not a subspace of $\mathbb{R}^2$. The set can be equivalently expressed as $S_1 \cup S_2, S_1 = \{(x,y) \in \mathbb{R}^2 | x = y\}, S_1 = \{(x,y) \in \mathbb{R}^2 | x = -y\}$. The combination of the point $s_1 \in S_1$ and $s_2 \in S_2$ does not necessarily belong to $S_1 \cup S_2$. Example:

$$(1,1) \in S_1, (1,-1) \in S_2, 1 \times (1,1) + 1 \times (1,-1) \notin S_1 \cup S_2.$$

(c) Not a subspace of $\mathbb{R}^2$. The combination of two points from this set does not necessarily belong to this set.

(d) Is a subspace of $\mathbb{R}^2$. Take arbitrary two points from the set (a,a) and (b,b) and arbitrary two scalars α and β. The combination $\alpha \times (a,a) + \beta \times (b,b) = (\alpha a + \beta b, \alpha a + \beta b)$ is still in the original set.

(e) Not a subspace of $\mathbb{R}^2$. Take a counter example $1 \times (2,1) + 1 \times (3,2) = (5,3)$, which does not belong to the original set.

3. 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.

Solution: We can verify

$$AB = \begin{bmatrix} -4 & 7 \\ 0 & 5 \end{bmatrix}, BA = \begin{bmatrix} -1 & 2 \\ 9 & 2 \end{bmatrix}.$$

Therefore $AB \neq BA$.

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

Solution: Denote the matrix A by

$$A = \begin{bmatrix} a^1 \\ \vdots \\ a^i \\ \vdots \\ a^m \end{bmatrix} \in \mathbb{R}^{m \times n}$$

where a_i is a row vector. We can express AB as

$$A = \begin{bmatrix} a^1 \\ \vdots \\ a^i \\ \vdots \\ a^m \end{bmatrix} \times B = \begin{bmatrix} a^1 B \\ \vdots \\ a^i B \\ \vdots \\ a^m B \end{bmatrix}.$$

Since one of $\{a^i\}$ is a zero vector, AB has a zero row-vector.

(b) 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.

Solution: Denote the matrix A by $A = [a_1 \dots a_i \dots a_n]$, where $\{a_i\}$ is a column vector. Since BA can be expressed as

$$BA = B \times [a_1 \dots a_i \dots a_n] = [Ba_1 \dots Ba_i \dots Ba_n]$$

BA has a column vector of zeros.

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

Proof:

Assume that $A \in \mathbf{R}^{n \times r}$ and $B \in \mathbf{R}^{r \times m}$ and denote $C = AB$. The proof consists of two parts.

- We prove $\text{rank}(C) \leq \text{rank}(A)$. Denote the matrices A and B respectively by

$$A = [a_1, \dots, a_r], B = [b_1, \dots, b_m],$$

where $a_i(i = 1, \dots, r)$ and $b_i(j = 1, \dots, m)$ are column vectors of length n and r, respectively. Let the j-th column in B be

$$b_j = \begin{bmatrix} b_{j1} \\ b_{j2} \\ \vdots b_{jr} \end{bmatrix}.$$

The j-th column of C can be expressed as

$$c_j = Ab_j = [a_1, a_2, \dots, a_r] \times \begin{bmatrix} b_{j1} \\ b_{j2} \\ \vdots \\ b_{jr} \end{bmatrix},$$

which is a linear combination of the r columns of A. Therefore we have $\text{rank}(C) \leq \text{rank}(A)$.

- Similarly, we can prove $\text{rank}(C) \leq \text{rank}(B)$