ML 7641 Homework 1

Yuewen Zheng

2021-02-21

1 Linear Algebra

1.1 Determinant and Inverse of Matrix

1. The determinant of M

$$
\begin{aligned}
|M| &= r \times \begin{bmatrix} r & 3 \\ 4 & 1 \end{bmatrix} - 0 \times \begin{bmatrix} 0.5 & 3 \\ -9 & 1 \end{bmatrix} + 1 \times \begin{bmatrix} 0.5 & r \\ -9 & 4 \end{bmatrix} \\
&= r(r-12) + 1(2-(-9r)) \\
&= r^2 - 12r + 2 + 9r \\
&= r^2 - 3r + 2
\end{aligned}
$$

2. The value(s) of r for which M^{-1} does not exist and the means in terms of rank and singularity for values of r.
According to the properties of the matrix, M^{-1} does not exist When $|M| = 0$

$$|M| = 0$$

$$r^2 - 3r + 2 = 0$$

$$r = 1, 2$$

When $r = 1, 2$, the determinant of M is 0, which means $rank(M) = 0$, and M is a singular matrix.

3. Calculate M^{-1} for $r = 3$ and show that $MM^{-1} = I$.
When $r = 3$,

$$M = \begin{bmatrix} 3 & 0 & 1 \\ 0.5 & 3 & 3 \\ -9 & 4 & 1 \end{bmatrix}$$

$$Adj(M) = \begin{pmatrix} +\begin{vmatrix} 3 & 3 \\ 4 & 1 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 4 & 1 \end{vmatrix} & +\begin{vmatrix} 0 & 1 \\ 3 & 3 \end{vmatrix} \\ \\ -\begin{vmatrix} 0.5 & 3 \\ -9 & 1 \end{vmatrix} & +\begin{vmatrix} 3 & 1 \\ -9 & 1 \end{vmatrix} & -\begin{vmatrix} 3 & 1 \\ 0.5 & 3 \end{vmatrix} \\ \\ +\begin{vmatrix} 0.5 & 3 \\ -9 & 4 \end{vmatrix} & -\begin{vmatrix} 3 & 0 \\ -9 & 4 \end{vmatrix} & +\begin{vmatrix} 3 & 0 \\ 0.5 & 3 \end{vmatrix} \end{pmatrix} = \begin{bmatrix} -9 & 4 & -3 \\ -27.5 & 12 & -8.5 \\ 29 & -12 & 9 \end{bmatrix}$$

$$|M| = r^2 - 3r + 2 = 9 - 9 + 2 = 2$$

$$M^{-1} = \frac{\text{adj}(M)}{\det(M)} = \frac{\begin{bmatrix} -9 & 4 & -3 \\ -27.5 & 12 & -8.5 \\ 29 & -12 & 9 \end{bmatrix}}{2} = \begin{bmatrix} -4.5 & 2 & -1.5 \\ -13.75 & 6 & -4.25 \\ 14.5 & -6 & 4.5 \end{bmatrix}$$

Thus,

$$MM^{-1} = \begin{bmatrix} 3 & 0 & 1 \\ 0.5 & 3 & 3 \\ -9 & 4 & 1 \end{bmatrix} \begin{bmatrix} -4.5 & 2 & -1.5 \\ -13.75 & 6 & -4.25 \\ 14.5 & -6 & 4.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4. Find the determinant of M^{-1} for $r = 3$. What is the relationship between the determinant of M and the determinant of M^{-1}?

   When r is 3, according to the problem 3 the determinant of M is 2.

$$\begin{aligned} |M^{-1}| &= -4.5 \times \begin{bmatrix} 6 & -4.25 \\ -6 & 4.5 \end{bmatrix} - 2 \times \begin{bmatrix} -13.75 & -4.25 \\ 14.5 & 4.5 \end{bmatrix} - 1.5 \times \begin{bmatrix} -13.75 & 6 \\ 14.5 & -6 \end{bmatrix} \\ &= -4.5 \times 1.5 - 2 \times (-0.25) - 1.5 \times (-4.5) \\ &= 0.5 \end{aligned}$$

   Thus, the determinant of M^{-1} is 0.5. The relationship between the determinant of M and the determinant of M^{-1} is that $det(M) = \frac{1}{det(M^{-1})}$

1.2 Characteristic Equation

Consider the eigenvalue problem:

$$Ax = \lambda x, x \neq 0$$

where x is a non-zero eigenvector and λ is eigenvalue of A. Prove that the determinant $|A - \Lambda I| = 0$.
Multiply the identity matrix I on the both sides of the equation, then we get:

$$IAx = I\lambda x$$

$$Ax = (\lambda I)x$$

Then,

$$(A - \lambda I)x = 0$$

If $B = A - \lambda I$ is a inverse matrix, then we could get:

$$x = B^{-1}0$$

Which means x is a zero vector.
However, x is a non-zero eigenvector, so $B = A - \lambda I$ is not a inverse matrix. It means that the determinant of $B = A - \lambda I$ is zero, $|A - \lambda I| = 0$.

1.3 Eigenvalues and Eigenvectors

1. The eigenvalues of A in terms of z.
   Organize the equation:

$$Ax = \lambda x$$

   which x is the eigenvector, λ is the eigenvalue

$$(A - \lambda I)x = 0$$

$$\begin{aligned} det(A - \lambda I) &= 0 \\ det(A - \lambda I) &= \begin{vmatrix} z - \lambda & 2 \\ 3 & z - \lambda \end{vmatrix} \\ &= (z - \lambda)^2 - 6 \\ \lambda &= z \pm \sqrt{6} \end{aligned}$$

2. Find the normalized eigenvectors of matrix A.

$$\begin{aligned} A - \lambda I &= \begin{bmatrix} \pm\sqrt{6} & 2 \\ 3 & \pm\sqrt{6} \end{bmatrix} \\ (A - \lambda I)x &= 0 \\ \begin{bmatrix} \pm\sqrt{6} & 2 \\ 3 & \pm\sqrt{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= 0 \end{aligned}$$

   From this equation, we could get:

$$3x_1 \pm \sqrt{6}x_2 = 0$$

   Due to the normalized eigenvectors, which means:

$$x_1^2 + x_2^2 = 1$$

   Thus,

$$x = \begin{bmatrix} \frac{\sqrt{10}}{5} \\ \frac{\sqrt{15}}{5} \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{10}}{5} \\ \frac{\sqrt{15}}{5} \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{10}}{5} \\ -\frac{\sqrt{15}}{5} \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{10}}{5} \\ -\frac{\sqrt{15}}{5} \end{bmatrix}$$

2 Expectation, Co-variance, Independence, and Bayes' Rule

1. Calculate the Expectation and Variance of X

$$\begin{aligned}
E(X) &= 0 \times P(x=0) + 1 \times P(x=1) \\
&= 0(1-p) + 1 \times p \\
&= p \\
E(X^2) &= 0 \times P(x^2=0) + 1 \times P(x^2=1) \\
&= 0(1-p) + 1 \times p \\
&= p \\
Var(X) &= E(X^2) - E(X)^2 = p - p^2
\end{aligned}$$

2. Calculate the Expectation and Variance of Z.
According to the question, Z is a variable following the Binomial distribution with probability p. The probability density function for Z is given by:

$$P(Z=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where k is the number of x = 1

$$\begin{aligned}
E(Z) &= E(Y_1 + Y_2 + \cdots + Y_n) \\
&= E(Y_1) + E(Y_2) + \cdots + E(Y_n) \\
&= nE(X) \\
&= np
\end{aligned}$$

Due to each element of Y is independent of the other elements, the Variance of Z is:

$$\begin{aligned}
Var(Z) &= Var(Y_1 + Y_2 + \cdots + Y_n) \\
&= Var(Y_1) + Var(Y_2) + \cdots + Var(Y_n) \\
&= nVar(X) \\
&= np(1-p)
\end{aligned}$$

3. Are Z and X independent?
The conclusion is that Z and X is dependent, the prove is below:

$$\begin{aligned}
P(Z=k, X=1) &= P(Z=k|X=1)P(X=1) \\
&= \binom{n-1}{k-1} p^{k-1}(1-p)^{n-1-k+1} \\
&= \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} \\
P(Z=k)P(X=1) &= \binom{n}{k} p^k (1-p)^{n-k} \times p \\
&= \frac{n!}{k!(n-k)!} p^{k+1}(1-p)^{n-k}
\end{aligned}$$

Thus,

$$P(Z=k)P(X=1) = \frac{np^2}{k} P(Z=k, X=1)$$

For each k, it is impossible for $np^2 = k$, which means Z and X are dependent.

4. Find $P(X = 1|Z = k)$ using Bayes' Rule. Simplify your answer.

$$\begin{aligned} P(X = 1|Z = k) &= \frac{P(X = 1, Z = k)}{P(Z = k)} \\ &= \frac{\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k}}{\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}} \\ &= \frac{k}{np} \end{aligned}$$

5. Calculate $Cov(X, Z)$. Are X and Z uncorrelated or not? If they are correlated, interpret the sign of $Cov(X, Z)$.
   For $Z = Y_1 + Y_2 + \cdots + Y_n = X + Y_2 + \cdots + Y_n$, it is obvious that X is correlated with Z.

$$\begin{aligned} E(XZ) &= E[X(Y_1 + Y_2 + \cdots + Y_n)] \\ &= E(Y_1^2 + Y_1Y_2 + \cdots + Y_1Y_n) \\ &= E(Y_1^2) + E(Y_1)E(Y_2) + \cdots + E(Y_1)E(Y_n) \\ &= p + (n-1)p^2 \\ Cov(X, Z) &= E(XZ) - E(X)E(Z) \\ &= p + (n-1)p^2 - p \times np \\ &= p - p^2 \end{aligned}$$

3 Optimization

1. Specify the Lagrange function.
   According to the question, the Lagrange function is:

$$\begin{aligned} L(x, y, a, b) &= f(x, y) - a[g_1(x, y) - 4] - b[g_2(x, y) - 4] \\ &= 2x^2 + 3xy - a(\frac{1}{2}x^2 - y - 4) - b(y - 4) \end{aligned}$$

   where $a,\ b \geq 0$

2. List the KTT conditions.

$$\begin{aligned} \frac{\partial L}{\partial x} &= 4x - ax + 3y = 0 \\ \frac{\partial L}{\partial y} &= a + 3x - b = 0 \\ a &\geq 0 \\ b &\geq 0 \\ a(\frac{1}{2}x^2 - y - 4) &= 0 \\ b(y - 4) &= 0 \end{aligned}$$

3. 4 possibilities formed could be

   (a)Both constraint are active, the optimal solution that lies at the intersection point of two constraints, so

$$a(\frac{1}{2}x^2 - y - 4) = 0 \text{ a} > 0$$

   Therefore,

$$\frac{1}{2}x^2 - y - 4 = 0$$

   And

$$b(y - 4) = 0 \text{ b} > 0$$

   Therefore,

$$y = 4$$

   and to solve the equal we could gain that $x = 4, -4$ and $y = 4$ Take the results to the KTT condition, we could obtain that

$$\begin{cases} x = 4 \quad a = 7, b = 19 \\ x = -4 \quad a = 1, b = -11 \ [invalid] \end{cases}$$

   (b)Both constraint are inactive, any of the constraint lines do not pass through the optimal point. We could get $a = 0$ and $b = 0$, so

$$\begin{cases} \frac{1}{2}x^2 - y - 4 \leq 0 \\ y - 4 \leq 0 \end{cases}$$

   Take the a, b to the KTT condition, we could obtain that

$$\begin{cases} x = 0 \\ y = 0 \end{cases}$$

   (c) The first constraint is active and the second constraint is inactive, therefore, we could get

$$\begin{cases} \frac{1}{2}x^2 - y - 4 = 0 \text{ a} > 0 \\ y - 4 \leq 0 \text{ b} = 0 \end{cases}$$

Take the $y = \frac{1}{2}x^2 - 4, b = 0$ to the KTT condition, we could obtain that

$$\begin{cases} 4x + 3\left(\frac{1}{2}x^2 - 4\right) - ax = 0 \\ 3x + a = 0 \end{cases}$$

Therefore,

$$x = -\frac{4 + 2\sqrt{58}}{9} = -2.14, \quad a = \frac{4 + 2\sqrt{58}}{3} = 6.41, \quad y = \frac{4(31 + 2\sqrt{58})}{81} - 4 = -1.72$$

(c) The first constraint is inactive and the second constraint is active, therefore, we could get

$$\begin{cases} \frac{1}{2}x^2 - y - 4 \leq 0 \text{ a} = 0 \\ y - 4 = 0 \text{ b} > 0 \end{cases}$$

Take the $y = 4, a = 0$ to the KTT condition, we could obtain that

$$\begin{cases} x = -3 \\ b = -9[invalid] \end{cases}$$

4. All candidate point is

$$\begin{cases} x = 4, \ y = 4 \quad a = 7, b = 19 \\ x = 0, \ y = 0 \quad a = 0, b = 0 \\ x = -2.14, \ y = -1.72 \quad a = 6.41, b = 0 \end{cases}$$

5. Take the all candidate point for $f(x, y)$

$$\begin{cases} f(4, 4) = 80 \quad \text{a=7,b=19} \\ f(0, 0) = 0 \quad \text{a=0,b=0} \\ f(-2.14, -1.72) = 20.14 \quad a = 6.41, b = 0 \end{cases}$$

for those candidate point, the maximum value is 80 when $x = 4$, and $y = 4$ To find the at this point, the $L(x, y)$ is concave or convex we need to take the derivative for $a = 7, \quad b = 19$

$$\frac{\partial L}{\partial x} = 3y - 3x$$

Take the second derivative, we could gain

$$\frac{\partial^2 L}{\partial x^2} = -3$$

For y,

$$\frac{\partial L}{\partial y} = 3x - 12$$

Take the second derivative, we could gain

$$\frac{\partial^2 L}{\partial y^2} = 0$$

Therefore, at this point, the $L(x, y)$ is concave.

4 Maximum Likelihood

4.1 Discrete Example

1. When you roll the die 6 times and each side appears exactly once, the likelihood of the result will use number of success divide the total number:
$$\theta = \frac{1}{6}$$

2. The maximum likelihood estimation(MLE): Based on Bernoulli Distribution, we took into two conditions, one is for when dice landing on 1 and another is for dice landing on other numbers.
$$P(X = x) = \begin{cases} 0 & \text{x=1} \\ 1 - \theta & \text{x=other} \end{cases}$$
For Bernoulli: objective function:
$$f(x_i; \theta) = \theta(1 - \theta)^5, x_i \in \{1, other\}$$
$$L(\theta) = p(X = 1, X = 2, X = 3, X = 4, X = 5, X = 6)$$
Based on the i.i.d assumption
$$L(\theta) = \prod_{i=1}^{6} \theta(1 - \theta)^5 = \underset{\theta}{\text{argmax}} \theta(1 - \theta)^5$$
On here, we covert multiplication into summation by taking the log
$$\log L(\theta) = l(\theta) = \log \theta + 5 \log(1 - \theta)$$
Take the derivative of the θ, we can gain,
$$\frac{1}{\theta} - \frac{5}{1 - \theta} = 0$$
Therefore,
$$\theta = \frac{1}{6}$$

3. When you roll the die 6 times and side 1 appears 4 times, the MLE for θ is:
$$\theta = \frac{2}{3}$$

4.2 Normal distribution

1. We assume that the noise is Guassian with mean 0 and variance σ^2. The z is in normal distribution, for y' it is also a noisy value which is equal to $y + z$. Therefore, if we re-write the model, it will be:

$$y' \sim \mathcal{N}(mx + c,\, \sigma^2)$$

Therefore, we can write the likelihood of this sample at (x_1, y_1') as,

$$L(y_1'|x_1, m, c, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1' - (mx_1 + c))^2}{2\sigma^2}}$$

2. When there are given n values and make c=0, the likelihood will be like,

$$L(y_i'|x_i, m, \sigma^2) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\frac{\sum_{i=1}^{n}(y_i' - mx_i)^2}{2\sigma^2}}$$

3. The maximum likelihood estimator for m is, first covert the multiplication into summation by take log

$$\log L(y_i'|x_i, m, \sigma^2) = -n\log(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{\sum_{i=1}^{n}(y_i' - mx_i)^2}{2\sigma^2}$$

Then, make derivative for m,

$$\frac{\partial \log(L)}{\partial m} = 0$$

$$-\frac{2\sum_{i=1}^{n} x_i(y_i' - mx_i)}{\sigma^2} = -\frac{2\sum_{i=1}^{n} x_i y_i' + \sum_{i=1}^{n} x_i^2}{\sigma^2}$$

$$m = \frac{\sum_{i=1}^{n} x_i y_i'}{\sum_{i=1}^{n} x_i^2}$$

4.3 Exponential Distribution

1. Fin the P.D.F for above C.D.C, we need to take the derivative,

$$f(x) = \frac{d}{dx_i}P(X_i \leq x|\lambda_i) = \frac{d}{dx_i}1 - e^{-\lambda_i x} = -(-\lambda_i \frac{d}{dx_i}x)e^{-\lambda_i x} = \lambda_i e^{-\lambda_i x}$$

$$P(X_i \leq x|\lambda_i) = \begin{cases} 0 & \text{x} < 0 \\ \lambda_i e^{-\lambda_i x} & 0 \leq x \end{cases}$$

2. The expected value of distribution for a given i is

$$\int_0^\infty p(x)dx = \frac{1}{\lambda_i}$$

3. Find Maximum Likelihood Estimator for λ,

$$L(\lambda|x_1, x_2...x_n) = \prod_{i=1}^{\frac{n}{2}} \lambda \exp^{-\lambda x_i} \prod_{i=\frac{n}{1}+1}^{n} 2\lambda \exp^{-2\lambda x_i}$$

$$= \lambda^{\frac{n}{2}} \exp^{-} \lambda \sum_{i=1}^{\frac{n}{2}} x_i \times 2\lambda^n \exp^{-} 2\lambda \sum_{i=\frac{n}{2}+1}^{n} x_i$$

On here, we covert multiplication into summation by taking the log,

$$\ln L = \frac{n\ln(\lambda)}{2} - \lambda \sum_{i=1}^{\frac{n}{2}} x_i - (\ln(2) + n\ln(\lambda)) - 2\lambda \sum_{i=\frac{n}{2}+1}^{n} x_i$$

Take the derivative,

$$\frac{\partial \log(L)}{\partial \lambda} = 0$$

$$\frac{n}{2\lambda} - \sum_{i=1}^{\frac{n}{2}} x_i + \frac{n}{\lambda} - 2 \sum_{i=\frac{n}{2}+1}^{n} x_i = 0$$

Based on the n is even and all values are positive, we could gain,

$$\frac{3n}{2a} = \sum_{i=1}^{n} x_i + \sum_{i=\frac{n}{2}+1}^{n} x_i$$

$$\lambda = \frac{3n}{2(\sum_{i=1}^{n} x_i + \sum_{i=\frac{n}{2}+1}^{n} x_i)}$$

4. the MLE for λ for $n = 1$ variable,

$$\lambda_1 = \frac{1}{x}$$

The reciprocal of the expected value is this value.

5 Information Theory

5.1 Mutual Information and Entropy

(a) The joint entropy

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} P(x,y)\log_2[P(x,y)] = H(X) + H(Y|X) = 3.375$$

(b) The marginal entropies is,

$$H(X) = \sum P(x)\log_2\frac{1}{P(x)} = 1.75$$

$$H(Y) = \sum P(y)\log_2\frac{1}{P(y)} = 2$$

(c) The conditional entropy formal is

$$H(X|Y) = H(X,Y) - H(Y) = 3.375 - 2 = 1.375$$

Mutual Information between X and Y is

$$I(X;Y) = H(X) - H(X|Y) = 1.75 - 1.375 = 0.375$$

5.2 Entropy Proofs

(a)

$$H(X \mid Y)=\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2\left(\frac{1}{P\left(x_i \mid y_j\right)}\right)$$

$$I(X ; Y)=\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i \mid y_j)}{P(x_i))}$$

$$I(X ; Y)=\sum_{i=1}^{n} \sum_{j=1}^{m} p\left(x_i, y_j\right) \log_2\left(\frac{1}{p\left(x_i\right)}\right)-\sum_{i=1}^{n} \sum_{j=1}^{m} P\left(x_i, y_j\right) \log_2\left(\frac{1}{P\left(x_i \mid y_j\right)}\right)$$

On here the $H(X|Y)=\sum_{i=1}^{n} \sum_{j=1}^{m} P\left(x_i, y_j\right) \log_2\left(\frac{1}{P(x_i|y_j)}\right)$

Therefore

$$I(X ; Y)=\sum_{i=1}^{n} \sum_{j=1}^{m} P\left(x_i, y_j\right) \log_2\left(\frac{1}{P\left(x_i\right)}\right)-H(X \mid Y)$$

On here

$$\sum_{i=1}^{m} p(x_i, y_j)=P\left(x_i\right)$$

$$I(X ; Y)=\sum_{i=1}^{n} P\left(x_i\right) \log_2\left(\frac{1}{p\left(x_i\right)}\right)-H(X \mid Y)$$

On here $\sum_{i=1}^{n} P\left(x_i\right) \log_2\left(\frac{1}{p(x_i)}\right)=H(X)$

Therefore, $I(X, Y)=H(X)-H(X|Y)$. And according to the symmetric properties, $H(X)-H(X|Y)=H(Y)-H(Y|X)$, so $I(X, Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)$

(b) Use the formula for joint entropy,

$$H(X, Y)=H(X)+H(Y|X)$$

Use the formula on (a),

$$I(X, Y)=H(X)-H(X|Y)$$

so

$$H(X)=I(X, Y)+H(X|Y)$$

Therefore,

$$H(X, Y)=H(X)+H(Y|X)=I(X, Y)+H(X|Y)+H(Y|X)$$

5.3 Bonus for undergrads

(a) According to the conditional entropy formula, Given that $H(Y|Z) = 0$ could be write as

$$H(Y|Z) = \sum_{i}^{n} P(y_1, Z) \log_2 \frac{p(Z)}{p(y_i, Z)} = 0$$

Therefore, $\log_2 \frac{p(Z)}{p(y_i,Z)} = 0$, so $\frac{p(Z)}{p(y_i,Z)} = 1$ and $p(Z) = p(y, Z)$
Since for any value of Z such $P(Z) > 0$, therefore, for any value of y, there exist such $P(y, Z) > 0$
For any value of Z and Y,because

$$P_{Y|Z}\left(y_i \mid z_j\right) = \frac{P_{YZ}\left(y_i, z_j\right)}{P_z\left(z_j\right)}$$

, so $P(Y|Z)$ cannot equal to 0. Furthermore, due to

$$p(Z) = p(y, Z) = p(y_i, Z)p(y_i)$$

, there only exists one value of Y such that $P(Y|Z)$ cannot equal to 0.