Homework 5

Jack Kai Lim

May 15, 2022

Problem 1

a)

As the equation for a one-dimensional linear regression is, $y = w_1 + w_2 x$ we can get the design matrix to be,

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Then using the equation $\vec{y} = X\vec{w}$ we get the response vector to be,

$$\vec{y} = \begin{bmatrix} w_1 + w_2 x_1 \\ w_1 + w_2 x_2 \\ \vdots \\ w_1 + w_2 x_n \end{bmatrix}$$

And from there we can get the normal equations matrix using $X^T\vec{y}$,

$$\begin{bmatrix} \sum_{i=1}^n w_1 + w_2 x_i \\ \sum_{i=1}^n x_i(w_1 + w_2 x_i) \end{bmatrix}$$

b)

Let the linear form of the equation be $H_0 = ax + b$ where a is the gradient and b is the intercept and rewriting the normal equation we can get,

$$X^T X\vec{w} = X^T\vec{y}$$
$$\begin{bmatrix} \sum_{i=1}^n w_1 + w_2 x_i \\ \sum_{i=1}^n x_i(w_1 + w_{\_}2x_i) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{bmatrix}$$

So we can take the first equation to get b

$$\sum_{i=1}^{n} w_1 + w_2 x_i = \sum_{i=1}^{n} y_i$$
$$\sum_{i=1}^{n} w_1 = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} w_2 x_i$$

Set $\sum_{i=1}^{n} w_1 = b$, $\sum_{i=1}^{n} y_i = H_0$ and $\sum_{i=1}^{n} w_2 x_i = ax$ where w_1 is the gradient of the linear line,

$$b = H_0 - ax$$

Taking the other equation,

$$\sum_{i=1}^{n} x_i(w_1 + w_2 x_i) = \sum_{i=1}^{n} y_i x_i$$
$$\sum_{i=1}^{n} w_2 x_i^2 = \sum_{i=1}^{n} x_i(y_i - w_1)$$
$$\sum_{i=1}^{n} w_2 = \frac{\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} w_1}{\sum_{i=1}^{n} x_i}$$

Now let $\sum_{i=1}^{n} w_2 = a$, $\sum_{i=1}^{n} y_i = H_0$, $\sum_{i=1}^{n} w_1 = b$ and $\sum_{i=1}^{n} x_i = x$ then we get,

$$a = \frac{H_0 - b}{x}$$

Problem 2

a)

We get the design matrix to be,

$$\begin{bmatrix} 1 & 1 & 7 \\ 1 & 2 & 14 \\ 1 & 3 & 21 \\ 1 & 4 & 28 \end{bmatrix}$$

Therefore we can get the matrix of $X^T X$,

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 7 & 14 & 21 & 28 \end{bmatrix} \begin{bmatrix} 1 & 1 & 7 \\ 1 & 2 & 14 \\ 1 & 3 & 21 \\ 1 & 4 & 28 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 70 \\ 10 & 30 & 210 \\ 70 & 210 & 1470 \end{bmatrix}$$

Therefore we can get the rank of the matrix as using RREF we get,

$$RREF(X^T X) = \begin{bmatrix} 4 & 10 & 70 \\ 0 & 5 & 35 \\ 0 & 0 & 0 \end{bmatrix}$$

As there are 2 pivots, we can see that the matrix has a rank of 2.

b)

We cannot calculate unique values of $\vec{w}$ in this case, as $X^T X$ is not full rank, it can not be inverted therefore we cannot calculate $\vec{w} = (X^T X)^{-1} X^T \vec{y}$ as $X^T X$ is non invertible as it isn't full rank

c)

To solve the questions we are going to use the fact that $\|\vec{w}\|^2 = \vec{w} \cdot \vec{w} = \vec{w}^T \vec{w}$ and $\frac{d}{d\vec{w}}(\vec{v} \cdot \vec{w}) = \vec{v}$ and finally the results we proved in homework 4 $\frac{d}{d\vec{w}}(\vec{w}^T X^T x \vec{w}) = 2X^T X \vec{w}$.

$$\begin{aligned} R_\lambda(\vec{w}) &= \frac{1}{n}\|\vec{y} - X\vec{w}\|^2 + \lambda \|\vec{w}\|^2 \\ &= \frac{1}{n}(\vec{y} - X\vec{w})^T (\vec{y} - X\vec{w}) + \lambda(\vec{w}^T \vec{w}) \\ &= \frac{1}{n}(\vec{y}^T \vec{y} - X^T \vec{y} \cdot \vec{w} - X^T \vec{y} \cdot \vec{w} + \vec{w}^T X^T X \vec{w}) + \lambda(\vec{w}^T \vec{w}) \\ &= \frac{1}{n}(\vec{y} \cdot \vec{y} - 2X^T \vec{y} \cdot \vec{w} + \vec{w}^T X^T X \vec{w}) + \lambda(\vec{w} \cdot \vec{w}) \end{aligned}$$

Now we can get the derivative of the equation to get the equation of the gradient of the function,

$$\begin{aligned}
\frac{d}{d\vec{w}}R_\lambda &= \frac{d}{d\vec{w}}\frac{1}{n}(\vec{y}\cdot\vec{y} - 2X^T\vec{y}\cdot\vec{w} + \vec{w}^TX^TX\vec{w}) + \frac{d}{d\vec{w}}\lambda(\vec{w}\cdot\vec{w}) \\
&= \frac{1}{n}(\frac{d}{d\vec{w}}\vec{y}\cdot\vec{y} - \frac{d}{d\vec{w}}2X^T\vec{y}\cdot\vec{w} + \frac{d}{d\vec{w}}\vec{w}^TX^TX\vec{w}) + \frac{d}{d\vec{w}}\lambda(\vec{w}\cdot\vec{w}) \\
&= \frac{1}{n}(-2X^T\vec{y} + 2X^TX\vec{w}) + 2\lambda\vec{w}
\end{aligned}$$

d)

To minimize the function we equate the derivative to 0 which gets us the normal equation to be,

$$X^TX\vec{w} + n\lambda\vec{w} = X^T\vec{y}$$

Next we notice that on the left hand side, that when the matrices are added, the constant $n\lambda$ are added on the diagonal of the square, symmetrical matrix of X^TX therefore we can get the equation to solve for a unique $\vec{w}$ as,

$$\vec{w} = (X^TX + n\lambda I)^{-1}X^T\vec{y}$$

Where I is the identity of dimension of the square matrix of X^TX. And therefore in order to solve this equation, the matrix of $X^TX + n\lambda I$ has to be full rank in order to be invertible, and as we add $n\lambda I$ we know that $X^TX + n\lambda I$ is always full rank and therefore always invertible. Hence, we can solve for a unique $\vec{w}$ in the dataset given,

$$\begin{aligned}
\vec{w} &= (X^TX + 4\lambda I)^{-1}X^T\vec{y} \\
&= \begin{bmatrix} 4+4\lambda & 10 & 70 \\ 10 & 30+4\lambda & 210 \\ 70 & 210 & 1470+4\lambda \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 7 & 14 & 21 & 28 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \end{bmatrix} \\
&= \frac{1}{2\lambda^2 + 752\lambda + 125} \begin{bmatrix} \frac{\lambda+375}{2} & -\frac{5}{4} & -\frac{35}{4} \\ -\frac{5}{4} & \frac{4\lambda^2+1474\lambda+245}{8\lambda} & -\frac{35(6\lambda+1)}{8\lambda} \\ -\frac{35}{4} & -\frac{35(6\lambda+1)}{8\lambda} & \frac{4\lambda^2 34\lambda+5}{8\lambda} \end{bmatrix} \begin{bmatrix} 100 \\ 300 \\ 2100 \end{bmatrix} \\
&= \frac{1}{2\lambda^2 + 752\lambda + 125} \begin{bmatrix} 50(\lambda+375) - 375 - 18375 \\ -125 + \frac{300(4\lambda^2+1474\lambda+245)}{8\lambda} - \frac{73500(6\lambda+1)}{8\lambda} \\ -875 - \frac{10500(6\lambda+1)}{8\lambda} + \frac{2100(4\lambda^2 34\lambda+5)}{8\lambda} \end{bmatrix}
\end{aligned}$$

Therefore in this case, as we know that $X^TX + n\lambda I$ is invertible we can say that for this dataset that there exist a unique $\vec{w}$ that is based of the value of lambda.

e)

As λ approaches 0, we only obtain a unique $\vec{w}$ if and only if the matrix of $X^T X$ is full rank / invertible. As when λ approaches 0 we get,

$$\lim_{\lambda \to 0} (X^T X + n\lambda I)^{-1} X^T \vec{y} = (X^T X)^{-1} X^T \vec{y}$$

Which is the normal equation for R_{sq}. Therefore in the case of the given dataset, relating back to part b, we know that we do not obtai a unique $\vec{w}$ in this case as $X^T X$ is not full rank.

Problem 3

a)

As there are 4 boxes of uniform probability, this means that the probability of either one of the boxes having the jackpot is $\frac{1}{4}$ chance. Therefore the box I chose has a probability of $\frac{1}{4}$ to be the jackpot. And the sample space of this problem is given as,

$$S = \{box_1, box_2, box_3, box_4\}$$

Where each box has is assigned a probability of $\frac{1}{4}$

b)

This is essentially the Monte Hall problem so to solve this using conditional probability, we can calculate the probability that A = (Probability that the box you chose has the Jackpot) and B = (Probability that there is nothing in the magicians box) using that we calculate $P(A|B)$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4} \times \frac{1}{3}}{\frac{1}{3}} = \frac{1}{4}$$

This may seem counterintuitive, but essentially this proves that even if the magicians opens and reveals an empty box, your probability of winning does not change assuming you do not switch.

If you switch however, because the host says "Here is your million dollar chance." and that the host being the host knows the position of the Jackpot. And also the fact that you know one of the 3 other boxes is empty. This makes the probability of either of the other 2 boxes containing the prize to be $1 - P(A|B) = \frac{3}{4}$ therefore the probability of winning if you switch is equal to,

$$\frac{\frac{3}{4}}{\frac{1}{2}} = \frac{3}{8}$$

And $\frac{3}{8} > \frac{1}{4}$ therefore you should switch everytime the host ask you, as based of probability switching always gives you a better shot of winning.

Problem 4

a)

Using the know rule for 2 events, we can break each of the 3 events into 2 events and prove that they are all the same

$$\begin{aligned} P((A \cap B) \cap C) &= P(A \cap B) \times P(C|(A \cap B)) \\ &= P(A) \times P(B|A) \times P(C|A \cap B) \end{aligned}$$

We can further proof that this is true by taking the bracket around B and C and proving equality using the multiplication rule and that intersects are commutative,

$$\begin{aligned} P(A \cap (B \cap C)) &= P((B \cap C) \cap A) \\ &= P(B \cap C) \times P(A|B \cap C) \\ &= P(B \cap C) \times \frac{P(A \cap B \cap C)}{P(B \cap C)} \\ &= P(A \cap B \cap C) \end{aligned}$$

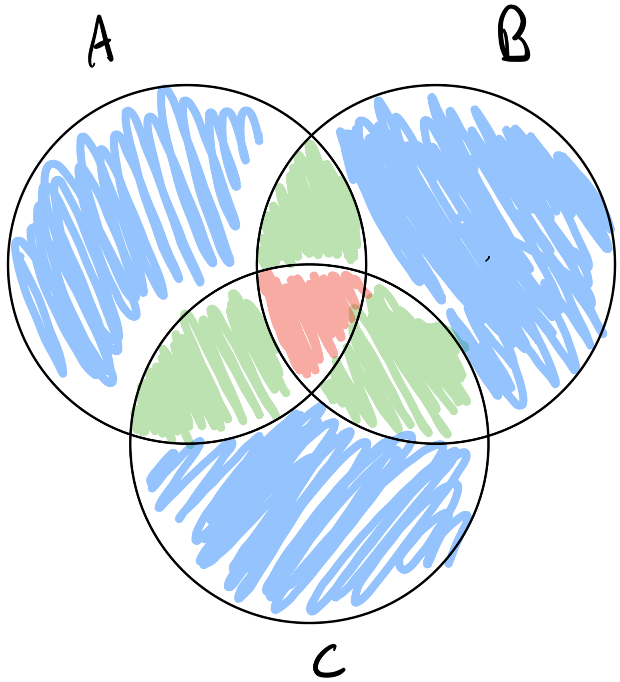
Therefore this shows equality proving the multiplication rule for 3 terms/events.

b)

Since we have that $P(A \cup B \cup C) = P((A \cup B) \cup C)$ and the distributive property of sets where $(A \cup B) \cap C = (A \cap C) \cup (B \cup C)$ and the given addition rule for 2 events, we can prove the addition rule for 3 events,

$$\begin{aligned} P((A \cup B) \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

c)



Looking at the diagram if we want to get the union of everything, it is essentially the same as trying to get the the entire area of the venn diagrams. But in the case of adding the area of the 3 circles representing the venn diagram, we wil have areas which overlap, areas which get added more than once. So in the venn diagram above, the areas coloured blue are added once, the areas that are green are added twice and the area that is red is added 3 times. We can get the green area between A and B as,

$$P(A \cap B) - P(A \cap B \cap C)$$

The green area between B and C as,

$$P(B \cap C) - P(A \cap B \cap C)$$

The green area between A and C as,

$$P(A \cap C) - P(A \cap B \cap C)$$

And finally the red area as,

$$P(A \cap B \cap C)$$

As we want every single area to be added only once we get the equation,

$$P(A) + P(B) + P(C) - [P(A \cap B) - P(A \cap B \cap C)] - [P(B \cap C) - P(A \cap B \cap C)]$$

$$-[P(A \cap C) - P(A \cap B \cap C)] - 2P(A \cap B \cap C)$$

Which yields us the 3 event general addition rule of

$$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Problem 5

a)

First we calculate the probability of getting the first King of Hearts from the first deck, using combinatorics we get,

$$P(\text{Probability of getting deck with KOHs}) = \frac{\text{Sets with KOHs}}{\text{Total number of sets}} = \frac{C\binom{51}{12}}{C\binom{52}{13}}$$

Calculating the combinatorics we get,

$$\frac{C\binom{51}{12}}{C\binom{52}{13}} = \frac{\frac{51!}{12!(51-12)!}}{\frac{52!}{13!(52-13)!}} = \frac{1}{4}$$

Next, to get the probability of getting the second King of hearts as well, we can square the answer to get the probability which is,

$$(\frac{1}{4})^2 = \frac{1}{16}$$

Therefore the probability of getting a pair of King of Hearts is $\frac{1}{16}$.

b)

Using the compliment of you getting A specific card, the probability of either one of the other players receiving the first Aces of Hearts is $1 - \frac{1}{4} = \frac{3}{4}$. Then the probability for the same person getting the second Ace of Hearts is $\frac{1}{4}$ Therefore the probability of any other player getting a pair of Ace of Hearts is,

$$\frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$$

c)

To calculate the probability of getting no pairs, first we notice that the first distribution of the first deck does not matter as all the cards will be different regardless. Then for the second deck, we just have to pick the cards 13 times and not get the same card that you already have in your hand which equates to,

$$P(c) = 1 \times \frac{39}{52} \times \frac{38}{51} \times ... \times \frac{27}{39}$$

Which equals to 0.00032797302

d)

Let A = (Probability that you get the pair of KOHS) and B = (Probability some other player gets AOHS) then the probability we want to calculate is $P(B|A)$ therefore,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{16} \times \frac{3}{16}}{\frac{1}{16}} = \frac{3}{16}$$

e)

Let A = (The probability that you get a pair of King of Hearts and no Aces of Hearts) and B = (The probability that any of the other players get the pair of Ace of Hearts). Therefore we want to find the probability of $P(B|A)$,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{16} \times \frac{3}{16}}{\frac{1}{16} \times \frac{9}{16}} = \frac{1}{3}$$

The P(A) is found by mutliplying the probability that you get the pair of King of Hearts $\frac{1}{4}$ times the probability you don't get any aces which is $\frac{9}{16}$. As for the probability of you getting the pair of Kings and any other player getting the pair of aces is, $\frac{1}{16} \times \frac{3}{16}$

f)

Here we can use combinatorics in a similar fashion to when I used it in part a,

$$\frac{\text{Sets with KOHS}}{\text{Total number of sets}} = \frac{C\binom{102}{24}}{C\binom{104}{26}} = \frac{25}{412}$$

g)

You are more likely to get a pair of King of Hearts when the cards are distributed separately as compared to when they are together. My intuitive explaination for this is that in the case where they are distributed separately, you are looking for the probability that in 13 cards for one of them to be the King of Hearts and for that to happen twice while when they are together, you are looking for 2 cards in 26 cards instead. This intuitive makes the first method more likely than the second as in the first method, you are looking for 1 card in 13 draws from a deck of 52 cards which is very likely, and to have this decently likely event to happen twice. While in the second method, you are looking for 2 cards to appear in 26 draws in a deck of 104 cards, which is a lot less likely as for one, you are looking for the same card to appear twice, in the 26 cards draw and secondly you are drawing from a much larger deck.