HTML_2023_HW0

tags: `Personal`

Combinatorics and Probability

P1

As we know, the description implements the pascal's rule, therefore, the result should be

$$
\begin{aligned}
\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} &= \frac{(n-1)!}{((n-1)-k)!k!} + \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!} \\
&= \frac{(n-1)!}{(n-1-k)!k!} + \frac{(n-1)!}{(n-k)!(k-1)!} \\
\text{simplifying} \Leftrightarrow & \ (n-1)-(k-1) = n-k \\
&= \frac{(n-1)!(n-k)}{(n-k)!k!} + \frac{(n-1)!k}{(n-k)!k!} \\
\text{Since } \frac{1}{(n-1-k)!} &\Rightarrow \frac{n-k}{(n-k)(n-k-1)(n-k-2)\ldots 3*2*1} \\
&= \frac{(n-1)!k + (n-1)!(n-k)}{(n-k)!k!} \\
&= \frac{(n-1)!(k+(n-k))}{(n-k)!k!} = \frac{(n-1)!n}{(n-k)!k!} \\
&= \frac{n!}{(n-k)!k!} = \binom{n}{k} \\
\therefore C(N,K) &= \frac{N!}{K!(N-K)!} \text{ for } N \geq 1 \text{ and } 0 \leq K \leq N
\end{aligned}
$$

As a result, we should choose $[a]$ as our solution.
Ref: [Show that C(n,k)=C(n−1,k)+C(n−1,k−1) [duplicate]](https://math.stackexchange.com/questions/780627/show-that-cn-k-cn-1-k-cn-1-k-1)
(https://math.stackexchange.com/questions/780627/show-that-cn-k-cn-1-k-cn-1-k-1)

P2

There are 2^{10} possible outcomes if 10 fair coins are tossed, and the possible outcomes of selecting exactly 4 heads from 10 coins are $\binom{10}{4}$, therefore the probability of getting exactly 4 heads when flipping 10 fair coins is

$$P(\text{head}=4)=\frac{C_4^{10}}{2^{10}}=\frac{\frac{10!}{4!(10-4)!}}{1024}=\frac{210}{1024}\approx 0.2$$

As a result, we should choose $[c]$ as our solution.

P3

The probability that all 3 tosses resulted in the heads is $\frac{1}{8}$.
The probability that one of the tosses resulted in head is $\frac{1}{8}$.
The probability of all possible outcomes after knowing one of the tosses resulted in the head is $1-\frac{1}{8}=\frac{7}{8}$
Therefore, the probability that all three tosses resulted in heads after knowing one of the tosses resulted in the head is $\frac{\frac{1}{8}}{\frac{7}{8}}=\frac{1}{7}$

As a result, we should choose $[d]$ as our solution.

P4

The probability of getting $x=1$ is $\frac{1}{8}$.
The probability of getting $x=-1$ is $\frac{1}{4}$.
The probability of getting $|x|=1$ is $\frac{1}{8}+\frac{1}{4}=\frac{3}{8}$.

$$\therefore \frac{P(x=-1)}{P(|x|=1)}=\frac{\frac{1}{4}}{\frac{3}{8}}=\frac{2}{3}$$

As a result, we should choose $[e]$ as our solution.

P5

$$
\begin{aligned}
\sigma_x^2 &= \frac{1}{N-1}\sum_{n=1}^{N}(x_n-\bar{x})^2 \\
&= \frac{1}{N-1}(\sum_{n=1}^{N}x_n^2 - 2\bar{x}\sum_{n=1}^{N}x_n + \sum_{n=1}^{N}(\bar{x})^2) \\
\text{Since } \bar{x} &\Rightarrow \frac{1}{N}\sum_{n=1}^{N}x_n \Rightarrow N\bar{x} = \sum_{n=1}^{N}x_n \\
&\Rightarrow 2\bar{x}\sum_{n=1}^{N}x_n = 2n(\bar{x})^2 = 2\sum_{n=1}^{N}(\bar{x})^2 \\
&= \frac{1}{N-1}(\sum_{n=1}^{N}x_n^2 - 2\sum_{n=1}^{N}(\bar{x})^2 + \sum_{n=1}^{N}(\bar{x})^2) \\
&= \frac{1}{N-1}(\sum_{n=1}^{N}x_n^2 - \sum_{n=1}^{N}(\bar{x})^2) \\
&= \frac{1}{N-1}\sum_{N=1}^{N}(x_n^2-\bar{x}^2)
\end{aligned}
$$

As a result, we should choose $[b]$ as our solution.

Ref: How 1n∑ni=1X2i−X¯2=∑ni=1(Xi−X¯)2n (https://math.stackexchange.com/questions/416581/how-frac1n-sum-i-1n-x-i2-bar-x2-frac-sum-i-1n-x-i-bar-x)

P6

The tightest possible range of $P(A \cup B)$ happens when

1. A is a subset of B, $A \subseteq B \Rightarrow P(A|B) = 1$
2. A and B are disjoint, $P(A|B) = 0$

$\therefore$ The range of $P(A \cup B) = [P(B), P(A) + P(B)] = [0.4, 0.7]$

As a result, we should choose $[e]$ as our solution.

Linear Algebra

P7

The rank of a matrix is the number of linearly independent rows.

$$\begin{bmatrix}1 & 2 & 1\\1 & 0 & 3\\1 & 1 & 2\end{bmatrix} \Rightarrow \begin{bmatrix}1 & 2 & 1\\0 & -2 & 2\\0 & -1 & 1\end{bmatrix} \Rightarrow \begin{bmatrix}1 & 2 & 1\\0 & -2 & 2\\0 & 0 & 0\end{bmatrix}$$

There are one zero columns, two non-zero columns, therefore, the rank of the matrix is 2.

As a result, we should choose $[c]$ as our solution.
Ref: How to Find the Rank of a Matrix (with echelon form) (https://www.youtube.com/watch?v=cSj82GG6MX4)

P8

$$\begin{aligned}
\text{Let } M &= \begin{bmatrix}0 & 2 & 4\\2 & 4 & 2\\3 & 3 & 1\end{bmatrix}\\
det(M) &= 0\times4\times1 + 2\times2\times3 + 4\times2\times3\\
&- \ 4\times4\times3 - 2\times2\times1 - 0\times2\times3\\
&= -16\\
Adj(M) &= \begin{bmatrix}+\begin{vmatrix}4 & 2\\3 & 1\end{vmatrix} & -\begin{vmatrix}2 & 4\\3 & 1\end{vmatrix} & +\begin{vmatrix}2 & 4\\4 & 2\end{vmatrix}\\ -\begin{vmatrix}2 & 2\\3 & 1\end{vmatrix} & +\begin{vmatrix}0 & 4\\3 & 1\end{vmatrix} & -\begin{vmatrix}0 & 4\\2 & 2\end{vmatrix}\\ +\begin{vmatrix}2 & 4\\3 & 3\end{vmatrix} & -\begin{vmatrix}0 & 2\\3 & 3\end{vmatrix} & +\begin{vmatrix}0 & 2\\2 & 4\end{vmatrix}\end{bmatrix}\\
&= \begin{bmatrix}-2 & 10 & -12\\4 & -12 & 8\\-6 & 6 & -4\end{bmatrix}\\
M^{-1} &= \frac{1}{det(M)} \times Adj(M)\\
&= \begin{bmatrix}\frac{1}{8} & -\frac{5}{8} & \frac{3}{4}\\-\frac{1}{4} & \frac{3}{4} & -\frac{1}{2}\\\frac{3}{8} & -\frac{3}{8} & \frac{1}{4}\end{bmatrix}
\end{aligned}$$

The diagonal of $M^{-1} = [\frac{1}{8}, \frac{3}{4}, \frac{1}{4}]$

As a result, we should choose $[d]$ as our solution.
Ref: 三階逆矩陣公式 (https://ccjou.wordpress.com/2012/10/04/%E4%B8%89%E9%9A%8E%E9%80%86%E7%9F%A9%E9%99%A3%E5%85%AC%E5%BC%8F/)
Ref: Main Diagonal (https://chortle.ccsu.edu/vectorlessons/vmch13/vmch13_17.html)

P9

$$\text{Let } M = \begin{bmatrix} 2023 & 1 & 1 \\ 2 & 2024 & 2 \\ -1 & -1 & 2021 \end{bmatrix}$$

From the definition of the eigenvetor v corresponding to the eigenvalue λ we have $Mv = \lambda v$, then $Mv = \lambda v = (M - \lambda \mathcal{I})v = 0$
The equation has a nonzero solution if and only if $det(M - \lambda \mathcal{I}) = 0$

$$\begin{aligned} det(M - \lambda \mathcal{I}) &= \begin{vmatrix} 2023-\lambda & 1 & 1 \\ 2 & 2024-\lambda & 2 \\ -1 & -1 & 2021-\lambda \end{vmatrix} \\ &= (2023-\lambda) \times (2024-\lambda) \times (2021-\lambda) + 1 \times 2 \times (-1) + 1 \times 2 \times (-1) \\ &- \ 1 \times (2024-\lambda) \times (-1) - (-1) \times 2 \times (2023-\lambda) - (2021-\lambda) \times 2 \times 1 \\ &= (2023-\lambda) \times (2024-\lambda) \times (2021-\lambda) + (2024-\lambda) \\ &= -(\lambda - 2024)((\lambda - 2023) \times (\lambda - 2021) + 1) \\ &= -(\lambda - 2024)(\lambda - 2022)^2 = 0 \\ &\Rightarrow \lambda = 2024 \text{ or } 2022 \end{aligned}$$

As a result, we should choose $[e]$ as our solution.

P10

$$\begin{aligned} \text{Let } M &= U\Sigma V^T \\ &\Rightarrow \text{where } M \in \mathbb{R}^{m\times n}, U \in \mathbb{R}^{m\times m}, \Sigma \in \mathbb{R}^{m\times n} V \in \mathbb{R}^{m\times n} \\ &\Rightarrow \text{where } UU^T = U^TU = I_m, VV^T = V^TV = I_n \\ \text{Define } M^\dagger &= V\Sigma^\dagger U^T, \text{where } \Sigma^\dagger[j][i] = \frac{1}{\Sigma[i][j]} \text{ when } \Sigma[i][j] \neq 0 \\ &\Rightarrow \text{where } \Sigma^\dagger \in \mathbb{R}^{m\times n} \\ \therefore MM^\dagger M &= U\Sigma V^T V\Sigma^\dagger U^T U\Sigma V^T \\ &= U(\Sigma\Sigma^\dagger)\Sigma V^T \\ &= U(I_m)\Sigma V^T = U\Sigma V^T \end{aligned}$$

If M is invertible (nonsingular) matrix, then $m = n$, and $\Sigma^\dagger = \Sigma^{-1}$

$$\begin{aligned} M^\dagger &= V\Sigma^\dagger U^T = V\Sigma^{-1}U^T \\ &= (U\Sigma V^T)^{-1} \\ &= M^{-1} \\ \therefore MM^\dagger M &= MM^{-1}M = M \end{aligned}$$

As a result, we should choose $[e]$ as our solution.

P11

Recall that a matrix M is positive semi-definite if $x^T M x > 0$ for all x

[a]

$$\forall x \Rightarrow x^T Z^T Z x = (Z^T x)^T (Z^T x) = \|Z^T x\|_2^2 \geq 0$$

[b]

A real symmetric matrix S whose eigenvalues are all non-negative is also always positive semi-definite by the Spectral Theorem. There is an orthogonal matrix Q such that $S = Q^T \Lambda Q$ with $\Lambda = \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}$. If x is any nonzero vector, then $y := Qx \neq 0$ and

$$\begin{aligned} x^T S x = x^T (Q^T \Lambda Q) &= (x^T Q^T)\Lambda(Qx) = y^T \Lambda y \\ &= \sum_{i=1}^{n} \lambda_i y_i^2 > 0 \end{aligned}$$

[c]

Since $x^T 0 x = 0$ for all x, it proves that an all-zero square matrix satisfies the requirement for positive semi-definiteness.

[d]

$$\begin{aligned} &\text{Consider } A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, x = \begin{bmatrix} 1 & -1 \end{bmatrix}^T, \text{then} \\ &\Rightarrow x^T A x = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 \end{aligned}$$

As a result, we should choose $[d]$ as our solution.

Ref: Eigenvalues and eigenvectors of symmetric matrices
(https://inst.eecs.berkeley.edu/~ee127/sp21/livebook/l_sym_sed.html)
Ref: Proving that a symmetric matrix is positive definite iff all eigenvalues are positive

(https://math.stackexchange.com/questions/1434167/proving-that-a-symmetric-matrix-is-positive-definite-iff-all-eigenvalues-are-pos)
Ref: Lecture 4.9. Positive definite and semidefinite forms
(https://www.math.purdue.edu/~eremenko/dvi/lect4.9.pdf)
Ref: https://mast.queensu.ca/~br66/419/spectraltheoremproof.pdf
(https://mast.queensu.ca/~br66/419/spectraltheoremproof.pdf)

P12

By using Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
&|u^T x| \leq |u^T| \cdot |x| = |x| \\
\Rightarrow\ &|x| \leq u^T x \leq |x| \\
&\min u^T x = -|x|
\end{aligned}
$$

As a result, we should choose $[c]$ as our solution.

P13

There must exist two points x_1, x_2, while x_1 is in H_1, and x_2 is in H_2. Also, a pair of x_1, x_2 satisfies $\|x_1 - x_2\| = |3 - (-2)| = 5$. Let d be a vector perpendicular to H_1 and H_2 and $\|d\| = d$. Since $w \perp H_1$, let $d = cw$. Then we have

$$
\begin{aligned}
((x_2 - x_1) - cw)^T cw = 0 \Rightarrow w^T(x_1 - x_2) - c\|w\|^2 = 0 \\
\Rightarrow c = \frac{w^T(x_2 - x_1)}{\|w\|^2} \\
\text{Since } w^T(x_2 - x_1) = 3 - (-2) = 5 \Rightarrow c = \frac{5}{\|w\|^2} \\
\therefore d = \|d\| = |c|\|w\| = \frac{5}{\|w\|}
\end{aligned}
$$

As a result, we should choose $[b]$ as our solution.
Ref: Distance between two hyperplanes (https://www.cosmozhang.xyz/notes/svm_pre.pdf)
Ref: Distance between two hyperplanes (https://math.stackexchange.com/questions/1484272/distance-between-two-hyperplanes)

Calculus

P14

$$\begin{aligned} \text{Let } g(x,y) &= e^x + e^{2y} + e^{3xy^2} \\ \therefore \frac{\partial g(x,y)}{\partial y} &= \frac{\partial}{\partial y}(e^x) + \frac{\partial}{\partial y}(e^{2y}) + \frac{\partial}{\partial y}(e^{3xy^2}) \\ &= 0 + (\frac{\partial}{\partial y}(2y))e^{2y} + (\frac{\partial}{\partial y}(3xy^2))e^{3xy^2} \\ &= 2e^{2y} + 6xye^{3xy^2} \end{aligned}$$

As a result, we should choose $[b]$ as our solution.

P15

$$\begin{aligned} \text{Let } f(x,y) &= xy, x(u,v) = cos(u+v), y(u,v) = sin(u-v) \\ \therefore \frac{\partial f}{\partial v} &= \frac{\partial}{\partial v}(xy) = \frac{\partial}{\partial v}(cos(u+v)sin(u-v)) \\ \text{Since product rule} \Rightarrow & \frac{\partial}{\partial y}(uv) = v\frac{\partial u}{\partial y} + u\frac{\partial v}{\partial y} \\ &= sin(u-v)\frac{\partial}{\partial v}(cos(u+v)) + cos(u+v)\frac{\partial}{\partial v}(sin(u-v)) \\ &= sin(u-v)(-sin(u+v)) + cos(u+v)cos(u-v) \\ &= -sin(u+v)sin(u-v) - cos(u+v)cos(u-v) \end{aligned}$$

As a result, we should choose $[a]$ as our solution.

P16

$$
\begin{aligned}
\text{Let } E(u,v) &= (ue^v - 2ve^{-u})^2 \text{, then we have} \\
\frac{\partial E}{\partial u} &= \frac{\partial}{\partial u}(ue^v - 2ve^{-u})^2 \\
\text{Since chain rule} &\Rightarrow \frac{\partial}{\partial u}((ue^v - 2ve^{-u})^2) = \frac{\partial x^2}{\partial x}\frac{\partial x}{\partial u}\text{, where} \\
&\Rightarrow x = ue^v - 2ve^{-u}\text{, and } \frac{\partial}{\partial x}(x^2) = 2x \\
&= 2(ue^v - 2ve^{-u})(\frac{\partial}{\partial u}(ue^v - 2ve^{-u})) \\
&= 2(ue^v - 2ve^{-u})(e^v - 2v(\frac{\partial}{\partial u}(e^{-u}))) \\
&= 2(ue^v - 2ve^{-u})(e^v + 2ve^{-u}) \\
&= 2(ue^v - 2ve^{-u})(e^v + 2ve^{-u})(e^{2u}e^{-2u}) \\
&= 2(ue^{u+v} - 2v)(e^{u+v} + 2v)e^{-2u} \\
\frac{\partial E}{\partial v} &= \frac{\partial}{\partial v}((ue^v - 2ve^{-u})^2) \\
\text{Since chain rule} &\Rightarrow x = ue^v - 2ve^{-u}\text{, and } \frac{\partial}{\partial x}(x^2) = 2x \\
&= 2(ue^v - 2ve^{-u})(\frac{\partial}{\partial v}(ue^v - 2ve^{-u})) \\
&= 2(ue^v - 2ve^{-u})(ue^v - 2e^{-u}) \\
&= 2(ue^v - 2ve^{-u})(ue^v - 2e^{-u})(e^{2u}e^{-2u}) \\
&= 2(ue^{u+v} - 2v)(ue^{u+v} - 2)e^{-2u} \\
\therefore \nabla E(u,v) &= \begin{pmatrix} \frac{\partial E}{\partial u} \\ \frac{\partial E}{\partial v} \end{pmatrix}\Bigg|_{(u,v)=(1,1)} = \begin{pmatrix} 2(ue^{u+v} - 2v)(e^{u+v} + 2v)e^{-2u} \\ 2(ue^{u+v} - 2v)(ue^{u+v} - 2)e^{-2u} \end{pmatrix}\Bigg|_{(u,v)=(1,1)} \\
&= \begin{pmatrix} 2(e^2 - 2)(e^2 + 2)e^{-2} \\ 2(e^2 - 2)(e^2 - 2)e^{-2} \end{pmatrix} = (\frac{2(e^2 - 2)(e^2 + 2)}{e^2}, \frac{2(e^2 - 2)(e^2 - 2)}{e^2}) \\
&\approx \begin{pmatrix} 13.70 \\ 7.86 \end{pmatrix}
\end{aligned}
$$

As a result, we should choose $[d]$ as our solution.

P17

Given $A > 0, B > 0$, we can calculate $\min_{\alpha} Ae^{\alpha} + Be^{-2\alpha}$ by finding the first derivative, where $\frac{\partial}{\partial \alpha}(Ae^{\alpha} + Be^{-2\alpha}) = 0$

$$\begin{aligned}
\text{Let } f(\alpha) &= Ae^{\alpha} + Be^{-2\alpha} \\
\therefore \frac{\partial f(\alpha)}{\partial \alpha} &= 0 = \frac{\partial}{\partial \alpha}(Ae^{\alpha} + Be^{-2\alpha}) = Ae^{\alpha} - 2Be^{-2\alpha} \\
&\Rightarrow Ae^{\alpha} = 2Be^{-2\alpha} \Rightarrow Ae^{3\alpha} = 2B \\
&\Rightarrow e^{3\alpha} = \frac{2B}{A} \Rightarrow 3\alpha = \ln(\frac{2B}{A}) \\
&\Rightarrow \alpha = \frac{1}{3}\ln(\frac{2B}{A})
\end{aligned}$$

As a result, we should choose $[a]$ as our solution.

P18

$$\begin{aligned}
\text{Let } E(\mathbf{w}) &= \frac{1}{2}\mathbf{w}^T A\mathbf{w} + b^T\mathbf{w} \\
E(\mathbf{w}) &= \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} w_i A_{ij} w_j + \sum_{i=1}^{d} b_i w_i \\
A = A^T &\Leftrightarrow \because \text{A is a symmetric matrix} \\
\therefore \nabla E(\mathbf{w}) &= \frac{\partial E(w)}{\partial w_k} = \frac{1}{2}\sum_{j=1}^{d} A_{kj} w_j + \frac{1}{2}\sum_{i=1}^{d} w_i A_{ik} + b_k \\
&= \sum_{j=1}^{d} A_{kj} w_j + b_k \\
&= A\mathbf{w} + b
\end{aligned}$$

As a result, we should choose $[c]$ as our solution.

P19

$$\begin{aligned}
\text{Let } E(\mathbf{w}) &= \frac{1}{2}\mathbf{w}^T A\mathbf{w} + b^T\mathbf{w} \\
\nabla E(\mathbf{w}) &= A\mathbf{w} + b = 0 \\
&\Rightarrow A\mathbf{w} = -b \\
&\Rightarrow \mathbf{w} = -A^{-1}b
\end{aligned}$$

As a result, we should choose $[b]$ as our solution.

P20

$$\underset{w1,w2,w3}{min}\frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) \text{ subject to } w_1 + w_2 + w_3 = 11$$

$$
\begin{aligned}
\text{Let } f(w_1, w_2, w_3) &= \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) \\
\text{Let } g(w_1, w_2, w_3) &= w_1 + w_2 + w_3 - 11 \\
\because \mathcal{L}(w_1, w_2, w_3, \lambda) &= f(w_1, w_2, w_3) - \lambda g(w_1, w_2, w_3) = 0 \\
&= \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) - \lambda(w_1 + w_2 + w_3 - 11) \\
\frac{\partial}{\partial w_1}(\mathcal{L}(w_1, w_2, w_3, \lambda)) &= w_1 - \lambda = 0 \Rightarrow w_1 = \lambda \\
\frac{\partial}{\partial w_2}(\mathcal{L}(w_1, w_2, w_3, \lambda)) &= 2w_2 - \lambda = 0 \Rightarrow w_2 = \frac{\lambda}{2} \\
\frac{\partial}{\partial w_3}(\mathcal{L}(w_1, w_2, w_3, \lambda)) &= 3w_3^2 - \lambda = 0 \Rightarrow w_3 = \frac{\lambda}{3} \\
\therefore w_1 + w_2 + w_3 = 11 &\Rightarrow \lambda + \frac{\lambda}{2} + \frac{\lambda}{3} = 11 \Rightarrow \lambda = 6 \\
&\Rightarrow w_1 = 6, w_2 = 3, w_3 = 2
\end{aligned}
$$

As a result, we should choose $[e]$ as our solution.

Ref: 14.8: Lagrange Multipliers

(https://math.libretexts.org/Bookshelves/Calculus/Book%3A_Calculus_(OpenStax)/14%3A_Differentiation_of_Functions_of_Several_Variables/14.08%3A_Lagrange_Multipliers)