Homework #0

RELEASE DATE: 09/02/2024

DUE DATE: 10/07/2024, 13:00, on Gradescope

QUESTIONS ARE WELCOMED ON DISCORD (INFORMALLY) OR NTU COOL (FORMALLY).

Please use Gradescope to upload your choices. For homework 0, you do not need to upload your scanned/printed solutions.

Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.

Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.

Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.

This homework set is of 40 points, which is much smaller than that of a usual homework set. For each problem, there is one correct choice. If you choose the correct answer, you get 2 points; if you choose an incorrect answer, you get 0 points.

Combinatorics and Probability

1. Let $C(N, K) = 1$ for $K = 0$ or $K = N$, and $C(N, K) = C(N-1, K) + C(N-1, K-1)$ for $N \geq 1$. What is the closed-form equation of $C(N, K)$ for $N \geq 1$ and $0 \leq K \leq N$?

[a] $C(N, K) = \frac{N!}{K!(N-K)!}$

[b] $C(N, K) = \sum_{k=0}^{K} \frac{N!}{k!(N-k)!}$

a

[c] $C(N, K) = \frac{K!(N-K)!}{K!}$

[d] $C(N, K) = \sum_{k=0}^{K} \frac{k!(N-k)!}{N!}$

[e] none of the other choices

2. What is the probability of getting exactly 3 tails when flipping 10 fair coins? Choose the closest number.

[a] 0.0

$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10\times\cancel{9}^{3}\times\cancel{8}^{4}}{\cancel{3}\times\cancel{2}} = 120$

[b] 0.1

[c] 0.2

b

$\frac{120}{2^{10}} = \frac{120}{1024} \doteq 0.117$

[d] 0.3

[e] 0.4

3. If your friend flipped a fair coin three times, and then tells you that one of the tosses resulted in head, what is the probability that all three tosses resulted in heads?

Let event A : all tosses are heads i.e. P(A|B)

event B : at least one toss is head

[a] 1/8

[b] 3/8

$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

d

[c] 7/8

[d] 1/7

$P(A) = \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2} = \frac{1}{8}$

$\therefore P(A|B) = \frac{1\cdot\frac{1}{8}}{\frac{7}{8}} = \frac{1}{8}\cdot\frac{8}{7} = \frac{1}{7}$

[e] 1/3

$P(B) = 1 - P(TTT) = 1 - \frac{1}{8} = \frac{7}{8}$

$P(B|A) = 1$

4. A program selects a random integer x like this: a random bit is first generated uniformly. If the bit is 0, x is drawn uniformly from $\{0, 1, \ldots, 3\}$; otherwise, x is drawn uniformly from $\{0, -1, \ldots, -7\}$. If we get an x from the program with $|x| = 1$, what is the probability that x is negative?

[a] 1/3
[b] 1/4
[c] 1/2
[d] 1/12
[e] 2/3

a

random bit: 0 $\longrightarrow$ x drawn uniformly from $[0, 1, 2, 3]$ $\quad \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$

1 $\longrightarrow$ x drawn uniformly from $\{0, -1, -2, -3, -4, -5, -6, -7\}$ $\quad \frac{1}{2} \times \frac{1}{8} = \frac{1}{16}$

$$\frac{\frac{1}{16}}{\frac{1}{8} + \frac{1}{16}} = \frac{1}{16} \times \frac{16}{3} = \frac{1}{3}$$

5. For N random variables $x_1, x_2, \ldots, x_N$, let their mean be $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$ and variance be $\sigma_x^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \bar{x})^2$. Which of the following is provably the same as σ_x^2?

~~[a]~~ $\frac{1}{N}\sum_{n=1}^{N}(x_n^2 - \bar{x}^2)$
[b] $\frac{1}{N-1}\sum_{n=1}^{N}(x_n^2 - \bar{x}^2)$
~~[c]~~ $\frac{1}{N-1}\sum_{n=1}^{N}(\bar{x}^2 - x_n^2)$
[d] $\frac{N}{N-1}(\bar{x}^2)$
[e] none of the other choices

b

$$\sigma_x^2 = \left(\frac{1}{N-1}\right)\sum_{n=1}^{N}(x_n - \bar{x})^2$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}(x_n^2 - 2x_n\bar{x} + \bar{x}^2)\right]$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}x_n^2 - 2\bar{x}\sum_{n=1}^{N}x_n + N\bar{x}^2\right]$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}x_n^2 - \bar{x}\left(2\sum_{n=1}^{N}x_n - N\bar{x}\right)\right]$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}x_n^2 - \frac{1}{N}\sum_{n=1}^{N}x_n\left(2\sum_{n=1}^{N}x_n - N\cdot\frac{1}{N}\sum_{n=1}^{N}x_n\right)\right]$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}x_n^2 - \left(\frac{1}{N}\sum_{n=1}^{N}x_n \cdot \sum_{n=1}^{N}x_n\right)\right]$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}x_n^2 - \frac{\left(\sum_{n=1}^{N}x_n\right)^2}{N}\right]$$
$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}x_n^2 - N\bar{x}^2\right] = \frac{1}{N-1}\sum_{n=1}^{N}(x_n^2 - \bar{x}^2)$$

6. For two events A and B, if their probability $P(A) = 0.2$ and $P(B) = 0.5$, what is the tightest possible range of $P(A \cup B)$?

[a] $[0.3, 0.4]$
[b] $[0, 0.4]$
[c] $[0.5, 0.7]$
[d] $[0.3, 1]$
[e] $[0.2, 0.7]$

c

$\to P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B) = 0.2 + 0.5 = 0.7$

$\to A \subset B \Rightarrow P(A \cup B) = 0.5$

Linear Algebra

7. Consider a line $w_0 + w_1x_1 + w_2x_2 = 0$ on the (x_1, x_2) plane with a non-zero w_1. Which of the following point is on the line?

[a] $(\frac{w_0}{w_1}, 0)$
√ [b] $(-\frac{w_0}{w_1}, 0)$
[c] (w_2, w_1)
[d] $(\frac{1}{2}w_0w_2, \frac{1}{2}w_0w_1)$
[e] none of the other choices

b

× (a) $w_0 + w_1 \cdot \frac{w_0}{w_1} + w_2 \cdot 0 = w_0 + w_0$

(b) $w_0 - w_1 \cdot \frac{w_0}{w_1} + w_2 \cdot 0 = w_0 - w_0 = 0$

× (c) $w_0 + w_1 \cdot w_2 + w_2 \cdot w_1 = w_0 + 2w_1w_2$

× (d) $w_0 + w_1 \cdot \frac{1}{2}w_0w_2 + w_2 \cdot \frac{1}{2}w_0w_1 = w_0 + w_0w_1w_2$

8. What is the diagonal on the inverse of $\begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{pmatrix}$?

[a] $[3/4, 1/4, 1/8]$
[b] $[1/4, 1/8, 3/4]$
[c] $[1/4, 3/4, 1/8]$
[d] $[1/8, 3/4, 1/4]$
[e] none of the other choices

d

$$\left[\begin{array}{ccc|ccc} 0 & 2 & 4 & 1 & 0 & 0 \\ 2 & 4 & 2 & 0 & 1 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \times\frac{1}{2} \\ \times\frac{1}{2} \\ \\ \end{array} \to \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 1 & 2 & 1 & 0 & \frac{1}{2} & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} -2 \\ -3 \end{array} \to \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & -3 & -1 & \frac{1}{2} & 0 \\ 3 & 0 & -5 & -\frac{3}{2} & 0 & 1 \end{array}\right] -3 \to \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & -3 & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 4 & \frac{3}{2} & -\frac{3}{2} & 1 \end{array}\right] \times\frac{1}{4}$$

$$\left[\begin{array}{ccc|ccc} 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & -3 & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{array}\right] \begin{array}{l} -2 \\ 3 \end{array} \to \left[\begin{array}{ccc|ccc} 0 & 1 & 0 & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ 1 & 0 & 0 & \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ 0 & 0 & 1 & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ 0 & 1 & 0 & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{array}\right]$$

$\therefore$ diag $= [\frac{1}{8} \ \frac{3}{4} \ \frac{1}{4}]$

$$\begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} -\frac{2}{4}+\frac{12}{8} & \frac{6}{4}-\frac{12}{8} & -1+1 \\ \frac{2}{8}-1+\frac{6}{8} & -\frac{10}{8}+3-\frac{6}{4} & \frac{6}{4}-2+\frac{2}{4} \\ \frac{3}{8}-\frac{3}{4}+\frac{3}{8} & -\frac{15}{8}+\frac{9}{4}-\frac{3}{8} & \frac{9}{4}-\frac{3}{2}+\frac{1}{4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

9. What is the largest eigenvalue of $\begin{pmatrix} 2023 & 1 & 1 \\ 2 & 2024 & 2 \\ -1 & -1 & 2021 \end{pmatrix}$?

[a] 2020
[b] 2021
[c] 2022
[d] 2023
[e] 2024

e

$$\begin{vmatrix} 2023-\lambda & 1 & 1 \\ 2 & 2024-\lambda & 2 \\ 1 & 1 & 2021-\lambda \end{vmatrix} \Big]_2 \rightarrow \begin{vmatrix} 2023-\lambda & 1 & 1 \\ 0 & 2022-\lambda & 4044-2\lambda \\ 1 & 1 & 2021-\lambda \end{vmatrix} \rightarrow \begin{vmatrix} 2023-\lambda & 1 & 0 \\ 0 & 2022-\lambda & 2022-\lambda \\ 1 & 1 & 2022-\lambda \end{vmatrix} \Big]_1$$

$$\rightarrow \begin{vmatrix} 2023-\lambda & 1 & 0 \\ 0 & 2022-\lambda & 2022-\lambda \\ 2022-\lambda & 0 & 2022-\lambda \end{vmatrix} = (2023-\lambda)\begin{vmatrix} 2022-\lambda & 2022-\lambda \\ 0 & 2022-\lambda \end{vmatrix} - \begin{vmatrix} 0 & 2022-\lambda \\ 2022-\lambda & 2022-\lambda \end{vmatrix} = (2023-\lambda)(2022-\lambda)^2 + (2022-\lambda)^2$$
$$= (2024-\lambda)(2022-\lambda)^2$$

10. For a real matrix M, let $M = U\Sigma V^T$ be its singular value decomposition, with U and V being unitary matrices. Define $M^\dagger = V\Sigma^\dagger U^T$, where $\Sigma^\dagger[j][i] = \frac{1}{\Sigma[i][j]}$ when $\Sigma[i][j]$ is nonzero, and 0 otherwise. Which of the following is always the same as $MM^\dagger M$?

[a] MM^TM
[b] MV^T
[c] U^TM
[d] U^TMV^T
[e] M

e

$$MM^+M = U\Sigma \cancel{V^T}\cancel{V}\Sigma^+\cancel{U^T}\cancel{U}\Sigma V^T = U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = M$$

(a) $MM^TM = U\Sigma V^T(U\Sigma V^T)^T U\Sigma V^T = U\Sigma\cancel{V^T}\cancel{V}\Sigma^T\cancel{U^T}\cancel{U}\Sigma V^T = U\Sigma\Sigma^T\Sigma V^T$

(b) $MV^T = V\Sigma V^TV^T$

(c) $U^TM = \cancel{U^T}\cancel{U}\Sigma V^T = \Sigma V^T$

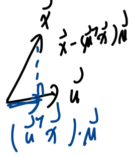
(d) $U^TMV^T = \cancel{U^T}\cancel{U}\Sigma V^TV^T = \Sigma V^TV^T$

11. Which of the following matrix is <u>not</u> guaranteed to be positive semi-definite?

√ [a] Z^TZ for any real matrix Z
√ [b] a real symmetric matrix S whose eigenvalues are all non-negative
√ [c] an all-zero square matrix $\vec{x}^T 0\, \vec{x} = 0 \geq 0 \quad \forall \vec{x} \in F^{n\times 1}$ 0 : n×n zero matrix
[d] a real symmetric matrix whose entries are all positive $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \rightarrow \det\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = 1-4 = -3 < 0$
[e] none of the other choices

d

12. Consider a fixed $\mathbf{x} \in \mathbb{R}^d$ and some varying $\mathbf{u} \in \mathbb{R}^d$ with $\|\mathbf{u}\| = 1$. Which of the following is the smallest value of $\mathbf{u}^T\mathbf{x}$?



[a] 0
~~[b]~~ $-\infty$
[c] $-\|\mathbf{x}\|$
~~[d]~~ $-\|\mathbf{u}\|$
[e] none of the other choices

c

Schwarz inequality

$|\vec{a}^T\vec{b}| \leq \|\vec{a}\|\|\vec{b}\|$

$\rightarrow |\vec{u}^T\vec{x}| \leq \|\vec{u}\|\|\vec{x}\|$

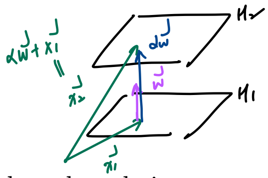
$\rightarrow |\vec{u}^T\vec{x}| \leq \|\vec{x}\| \rightarrow \vec{u}^T\vec{x} \in [-\|\vec{x}\|, \|\vec{x}\|]$

13. Consider two parallel hyperplanes in R^d:

$$H_1 : \mathbf{w}^T\mathbf{x} = +3,$$
$$H_2 : \mathbf{w}^T\mathbf{x} = -2,$$

$= \|\alpha\vec{w}\|$

What is the distance between H_1 and H_2?



[a] 5
[b] $5/\|\mathbf{w}\|$
[c] $5/\|\mathbf{w}\|^2$
[d] $5 \cdot \|\mathbf{w}\|$
[e] none of the other choices

b

Suppose $\vec{x}_1 \in H_1$

line L passes through $\vec{x}_1$ in the direction of $\vec{w}$

$\rightarrow L : \alpha\vec{w} + \vec{x}_1 \quad \forall \alpha \in \mathbb{R}$

$\vec{w}^T(\alpha\vec{w} + \vec{x}_1) = -2$

$\rightarrow \alpha\vec{w}^T\vec{w} + \vec{w}^T\vec{x}_1 = -2$

$\rightarrow \alpha\vec{w}^T\vec{w} + 3 = -2$

$\rightarrow \alpha\vec{w}^T\vec{w} = -5$

$\rightarrow \alpha = -\frac{5}{\|\vec{w}\|^2}$

$\|\alpha\vec{w}\| = \|-\frac{5}{\|\vec{w}\|^2}\cdot\vec{w}\| = \frac{5}{\|\vec{w}\|^2}\cdot\|\vec{w}\| = \frac{5}{\|\vec{w}\|}$

Calculus

14. Let $f(x, y) = xy$, $x(u, v) = \cos(u + v)$, $y(u, v) = \sin(u - v)$. What is $\dfrac{\partial f}{\partial v}$?

a

[a] $-\sin(u + v)\sin(u - v) - \cos(u + v)\cos(u - v)$
[b] $+\sin(u + v)\sin(u - v) - \cos(u + v)\cos(u - v)$
[c] $-\sin(u + v)\sin(u - v) + \cos(u + v)\cos(u - v)$
[d] $+\sin(u + v)\sin(u - v) + \cos(u + v)\cos(u - v)$
[e] none of the other choices

$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}$$
$$= \sin(u-v)(-\sin(u+v)) + \cos(u+v)(-\cos(u-v))$$
$$= -\sin(u-v)\sin(u+v) - \cos(u+v)\cos(u-v)$$

15. Let $E(u, v) = (ue^v - 2ve^{-u})^2$. Calculate the gradient $\nabla E(u, v) = \begin{pmatrix} \frac{\partial E}{\partial u} \\ \frac{\partial E}{\partial v} \end{pmatrix}$ at $[u, v] = [1, 1]$. Choose the closest vector.

d

~~[a]~~ $[-13.70, -7.86]$
[b] $[-13.70, +7.86]$
~~[c]~~ $[+13.70, -7.86]$
[d] $[+13.70, +7.86]$
~~[e]~~ $[1, 1]$

$$\frac{\partial E}{\partial u}\Big|_{(1,1)} = 2(ue^v - 2ve^{-u})(e^v + 2ve^{-u})\Big|_{(1,1)} = 2(e - 2\cdot\tfrac{1}{e})(e + 2\tfrac{1}{e}) = 2(\underset{7.39}{e^2} - \underset{0.54}{\tfrac{4}{e^2}}) \fallingdotseq 13.7$$

$$\frac{\partial E}{\partial v}\Big|_{(1,1)} = 2(ue^v - 2ve^{-u})(ue^v - 2e^{-u})\Big|_{(1,1)} = 2(e - 2\tfrac{1}{e})(e - 2\tfrac{1}{e}) = 2(\underset{2.718}{e} - \underset{0.735}{\tfrac{2}{e}})^2 \fallingdotseq 7.86$$

16. For some given $A > 0, B > 0$, what is the optimal α that solves

$$\min_{\alpha} Ae^{\alpha} + Be^{-2\alpha}?$$

a

[a] $\frac{1}{3}\ln(\frac{2B}{A})$
[b] $\frac{1}{3}\ln(\frac{A}{2B})$
[c] $\ln(\frac{2B}{A})$
[d] $\ln(\frac{A}{2B})$
[e] none of the other choices

$$\frac{d}{d\alpha} Ae^{\alpha} + Be^{-2\alpha} = 0$$
$\rightarrow Ae^{\alpha} - 2Be^{-2\alpha} = 0$
$\rightarrow Ae^{\alpha} = 2Be^{-2\alpha}$
$\rightarrow e^{3\alpha} = \frac{2B}{A}$
$\rightarrow \ln(e^{3\alpha}) = \ln(\frac{2B}{A})$
$\rightarrow 3\alpha = \ln(\frac{2B}{A})$
$\rightarrow \alpha = \frac{1}{3}\ln(\frac{2B}{A})$

17. Let $\mathbf{w}$ be a vector in $\mathbb{R}^d$ and $E(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \mathrm{A}\mathbf{w} + \mathbf{b}^T\mathbf{w}$ for some symmetric matrix A and vector $\mathbf{b}$. What is the gradient $\nabla E(\mathbf{w})$?

c

[a] $\mathbf{w}^T\mathrm{A}\mathbf{w} + \mathbf{w}^T\mathbf{b}$
[b] $\mathbf{w}^T\mathrm{A}\mathbf{w} - \mathbf{w}^T\mathbf{b}$
[c] $\mathrm{A}\mathbf{w} + \mathbf{b}$
[d] $\mathrm{A}\mathbf{w} - \mathbf{b}$
[e] none of the other choices

$A \in \mathbb{R}^{d\times d}$

$$\rightarrow \vec{w}^T A \vec{w} = [w_1 \cdots w_d]_{1\times d}\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{d1} & \cdots & & a_{dd}\end{bmatrix}_{d\times d}\begin{bmatrix} w_1 \\ \vdots \\ w_d\end{bmatrix}_{d\times 1} = \left[\sum_{i=1}^d a_{i1}w_i \ \ \sum_{i=1}^d a_{i2}w_i \ \cdots \ \sum_{i=1}^d a_{id}w_i\right]_{1\times d}\begin{bmatrix} w_1 \\ \vdots \\ w_d\end{bmatrix}_{d\times 1}$$

$$= w_1\sum_{i=1}^d a_{i1}w_i + \cdots + w_d\sum_{i=1}^d a_{id}w_i = \sum_{j=1}^d w_j\sum_{i=1}^d a_{ij}w_i = \sum_{j=1}^d\sum_{i=1}^d w_j a_{ij} w_i$$

Let $f(\vec{w}) = \vec{w}^T A\vec{w}$

then $\nabla_{\vec{w}} f(\vec{w}) = \left[\frac{\partial f}{\partial w_1} \cdots \frac{\partial f}{\partial w_d}\right]^T$

$$\rightarrow \frac{\partial f}{\partial w_k} = \sum_{j=1}^d w_j a_{kj} + \sum_{i=1}^d w_i a_{ik}$$
$$= \sum_{j=1}^d w_j a_{jk} + \sum_{i=1}^d w_i a_{ik} \qquad \because A\text{: symmetric}$$
$$= 2\sum_{i=1}^d w_i a_{ik}$$

$$\therefore \nabla_{\vec{w}} f(\vec{w}) = 2\left[\sum_{i=1}^d w_i a_{i1} \cdots \sum_{i=1}^d w_i a_{id}\right]^T$$
$$= 2\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & \ddots & \vdots \\ a_{d1} & \cdots & & a_{dd}\end{bmatrix}\begin{bmatrix} w_1 \\ \vdots \\ w_d\end{bmatrix}$$
$$= 2A\vec{w}$$

$\rightarrow \nabla_{\vec{w}}(\vec{b}^T\vec{w}) = \vec{b}$

Thus, $\nabla E(\vec{w}) = \frac{1}{2}\cdot 2A\vec{w} + \vec{b} = A\vec{w} + \vec{b}$

$\exists A^{-1}$

18. Let $\mathbf{w}$ be a vector in $\mathbb{R}^d$ and $E(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \mathrm{A}\mathbf{w} + \mathbf{b}^T\mathbf{w}$ for some symmetric *and strictly positive definite* matrix A and vector $\mathbf{b}$. What is the optimal $\mathbf{w}$ that minimizes $E(\mathbf{w})$?

b

[a] $+\mathrm{A}^{-1}\mathbf{b}$

[b] $-\mathrm{A}^{-1}\mathbf{b}$

[c] $-\mathrm{A}^{-1}\mathbf{1} + \mathbf{b}$, where $\mathbf{1}$ is a vector of all 1's

[d] $+\mathrm{A}^{-1}\mathbf{1} - \mathbf{b}$

[e] none of the other choices

$\nabla \vec{x}^T A \vec{x} = 2A\vec{x}$

$\nabla E(\vec{w}) = A\vec{w} + \vec{b}$

$A\vec{w} + \vec{b} = \vec{0}$

$\Rightarrow A\vec{w} = -\vec{b}$

$\Rightarrow \vec{w} = -A^{-1}\vec{b}$

19. Solve

$$\min_{w_1,w_2,w_3} \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) \text{ subject to } w_1 + w_2 + w_3 = 11.$$

What is the optimal w_1? (Hint: refresh your memory on "Lagrange multipliers")

e

[a] 0

[b] 1

[c] 2

[d] 3

[e] 6

Let $f(w_1, w_2, w_3) = \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2)$

$g(w_1, w_2, w_3) = w_1 + w_2 + w_3 - 11$

$\nabla f = \lambda \nabla g$

$w_1 \vec{i} + 2w_2 \vec{j} + 3w_3 \vec{k} = \lambda(\vec{i} + \vec{j} + \vec{k})$

$w_1 = 2w_2 = 3w_3 \Rightarrow w_2 = \frac{1}{2}w_1, \quad w_3 = \frac{1}{3}w_1$

$w_1 + \frac{1}{2}w_1 + \frac{1}{3}w_1 = \frac{11}{6}w_1 = 11 \Rightarrow w_1 = 11 \times \frac{6}{11} = 6$

20. Solve

$$\begin{aligned} &\min_{w_1,w_2,w_3} && \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) \\ &\text{subject to} && w_1 + w_2 + w_3 \geq 11, \\ & && w_2 + 2w_3 \geq -11. \end{aligned}$$

What is the optimal (w_1, w_2, w_3)? (Hint: you can also consider using "Lagrange multipliers" to solve this.)

e

[a] $(3, 6, 2)$

[b] $(3, 2, 6)$

[c] $(6, 2, 3)$

[d] $(3, 6, 2)$

[e] $(6, 3, 2)$

Let $f(w_1, w_2, w_3) = \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2)$

$g_1(w_1, w_2, w_3) = w_1 + w_2 + w_3 - 11 - s^2 = 0$

$g_2(w_1, w_2, w_3) = w_2 + 2w_3 + 11 - t^2 = 0$

$L(w_1, w_2, w_3, \lambda, \theta) = \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) - \lambda(w_1 + w_2 + w_3 - 11 - s^2) - \theta(w_2 + 2w_3 + 11 - t^2)$

$\frac{\partial L}{\partial w_1} = w_1 - \lambda$

$\frac{\partial L}{\partial w_2} = 2w_2 - \lambda - \theta$

$\frac{\partial L}{\partial w_3} = 3w_3 - \lambda + 2\theta$

$\frac{\partial L}{\partial \lambda} = -(w_1 + w_2 + w_3 - 11 - s^2)$

$\frac{\partial L}{\partial \theta} = -(w_2 + 2w_3 + 11 - t^2)$

$$\boxed{\begin{aligned} &\frac{\partial L}{\partial w_1} = 0 \Rightarrow w_1 - \lambda = 0 \Rightarrow w_1 = \lambda \\ &\frac{\partial L}{\partial w_2} = 0 \Rightarrow 2w_2 - \lambda - \theta = 0 \Rightarrow w_2 = \frac{1}{2}\lambda + \frac{1}{2}\theta \\ &\frac{\partial L}{\partial w_3} = 0 \Rightarrow 3w_3 - \lambda + 2\theta = 0 \Rightarrow w_3 = \frac{1}{3}\lambda - \frac{2}{3}\theta \end{aligned}}$$

Case1: $\lambda = \theta = 0$, $s^2 > 0$, $t^2 > 0$

Case2: $\lambda = 0$ $\theta \neq 0$, $s^2 > 0$, $t^2 = 0$

Case3: $\lambda \neq 0$, $\theta = 0$, $s^2 = 0$, $t^2 > 0$

Case4: $\lambda \neq 0$, $\theta \neq 0$, $s^2 = t^2 = 0$

$\hookrightarrow$ Case1: $w_1 = 0$, $w_2 = 0$, $w_3 = 0$, $s^2 = -11$ (X)

Case2: $w_1 = 0$, $w_2 = \frac{1}{2}\theta$, $w_3 = -\frac{2}{3}\theta$

$-(w_2 + 2w_3 + 11) = 0 \Rightarrow -\frac{1}{2}\theta + \frac{4}{3}\theta + 11 = 0$

$\Rightarrow \frac{1}{3}\theta = -11 \Rightarrow \theta = -33$

$\therefore w_1 = 0, w_2 = \frac{-33}{2}, w_3 = 22$

Case3: $w_1 = \lambda$, $w_2 = \frac{1}{2}\lambda$, $w_3 = \frac{1}{3}\lambda$

$-(w_1 + w_2 + w_3 - 11) = 0 \Rightarrow \frac{11}{6}\lambda - 11 = 0 \Rightarrow \frac{11}{6}\lambda = 11 \Rightarrow \lambda = 6$

$(w_1, w_2, w_3) = (6, 3, 2)$

$-(3 + 4 + 11 - t^2) = 0 \Rightarrow t^2 = 18$ $\square$

Case4: $\begin{cases} w_1 + w_2 + w_3 - 11 = 0 \\ w_2 + 2w_3 + 11 = 0 \end{cases}$

$$\begin{array}{rr} & w_1 + w_2 + w_3 = 11 \\ -) & w_2 + 2w_3 = -11 \\ \hline & w_1 \quad -w_3 = 22 \end{array}$$

$w_1 = 22 + w_3$

$w_2 = -11 - 2w_3$

$\lambda = 22 + \frac{1}{3}\lambda - \frac{2}{3}\theta$

$\frac{1}{2}\lambda + \frac{1}{2}\theta = -11 - \frac{2}{3}\lambda + \frac{4}{3}\theta$

续2)

$\lambda = 22 + \frac{1}{3}\lambda - \frac{2}{3}\theta$

$\frac{1}{2}\lambda + \frac{1}{2}\theta = -11 - \frac{2}{3}\lambda + \frac{4}{3}\theta$

$$\begin{cases} \frac{2}{3}\lambda + \frac{2}{3}\theta = 22 \\ \frac{7}{6}\lambda - \frac{5}{6}\theta = -11 \end{cases} \Rightarrow \begin{cases} \lambda + \theta = 33 \\ 7\lambda - 5\theta = -66 \end{cases} \Rightarrow \begin{array}{rl} & 5\lambda + 5\theta = 165 \\ +) & 7\lambda - 5\theta = -66 \\ \hline & 12\lambda = 99 \end{array}$$

$\lambda = \frac{99}{12} = \frac{33}{4}$

$\theta = 33 - \frac{33}{4} = \frac{132}{4} - \frac{33}{4} = \frac{99}{4}$

$$(w_1, w_2, w_3) = \left(\lambda, \frac{1}{2}\lambda + \frac{1}{2}\theta, \frac{1}{3}\lambda - \frac{2}{3}\theta\right)$$

$$= \left(\frac{33}{4} \ \ldots\right.$$