Question 01

The truncation error of an algorithm is due to:

a. Truncation of an infinite procedure to a finite one;

b. Truncation of the results in arithmetic operation;

c. Truncation of the data.

Feedback

Correct answer: *Truncation of an infinite procedure to a finite one*

$\pi = 3{,}141592\ldots$

$\pi_{trunc} = 3{,}14$

Option: B

Question 2

If $x = 1.2345$ and $y = 0.07894$, which is the result of $z = x+y$ in $\mathbb{F}(10, 5, -4, 4)$?

a. none of the other alternatives

b. $0.131344 \cdot 10^1$

c. 0.13134

Feedback

Correct Answer: *none of the other alternatives*

GIVEN:

$x = 1{,}2345$

$y = 0{,}07894$

$\mathbb{F}(10,5,-4,4)$

GOAL:

Find $z = x+y$ in $\mathbb{F}(10,5,-4,4)$

$\beta = 10$ (base)

$p = 5$ (precision)

$e_{min}, e_{max} = (-4, 4)$ (exponent range)

1) Normalization of x and y:

$x = 1{,}2345$

$x = 0{,}\underbrace{12345}_{=\text{precision } 5} \cdot 10^1$

$y = 0{,}07894$

$y = 0{,}\underbrace{78940}_{=\text{precision } 5} \cdot 10^{-1}$

2) Align the exponents for addition:

(always adjust the exponent by the biggest one)

$x = 0{,}123450 \cdot 10^1$

$y = 0{,}007894 \cdot 10^1$

$x + y = 0{,}131344 \cdot 10^1$

$z = 0{,}131344 \cdot 10^1$

3) Round the result to the precision

$z = 0{,}13134|4 \cdot 10^1 \Rightarrow z = 0{,}13134 \cdot 10^1$

truncate

4) Verify the exponent range:

$e_{min} \leq e \leq e_{max}$

$-4 \leq 1 \leq 4$

5) Write the floating-point approx.

$\bar{z} = 0{,}13134 \cdot 10^1$

Option: A

Question 3

Which of the following statements for $A \in \mathbb{R}^{n\times n}$ is wrong?

a. $\|A\|_2^2 =$ maximum singular value of A;

b. $\|A\|_2^2 = \rho(A^T A)$;

c. $\|A\|_2^2 = \sum_{i=1}^n \sum_{j=1}^n a_{i,j}^2$.

THEORY RECAP:

Matrix Norms: $A \in \mathbb{R}^{m\times n} \rightarrow \mathbb{R}$

- 2-norm: $\|A\|_2^2 = \rho(A^T A)$ - spectral radius (largest eigenvalue of $A^T A$)
- $\|A\|_2$ - LARGEST SINGULAR VALUE of A (denoted $\sigma_{max}(A)$)
- FROBENIUS NORM: $\|A\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 \neq \|A\|_2^2$

1) statement A: $\|A\|_2^2 =$ max singular value of A.
This is INCORRECT because max singular value of A is $\|A\|_2$ and NOT $\|A\|_2^2$
The 2-norm squared relates to the spectral radius $\rho(A^T A)$

2) Statement B: $\|A\|_2^2 = \rho(A^T A)$ - CORRECT

3) statement C: $\|A\|_2^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$ - INCORRECT

This is INCORRECT because this formula defines the FROBENIUS NORM squared, not the 2-norm squared.

Question 4

If $x = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ and $y = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, then:

a. x and y are orthonormal;

b. x and y are orthogonal;

c. none of the other alternatives.

THEORY RECAP:

def.: ORTHOGONAL VECTORS:

$\rfloor\ \forall \vec{v}_1, \vec{v}_2 \in \mathbb{R}^{n \times n}$

$\perp\ \vec{v}_1 \wedge \vec{v}_2$ - ORTHOGONAL iff $\vec{v}_1 \cdot \vec{v}_2 = 0 \Rightarrow \vec{v}_1 \perp \vec{v}_2$

def.: ORTHONORMAL VECTORS:

$\rfloor\ \forall \vec{v}_1, \vec{v}_2 \in \mathbb{R}^{n \times n}$

$\perp\ \vec{v}_1 \wedge \vec{v}_2$ - ORTHONORMAL iff: 1) $\vec{v}_1 \wedge \vec{v}_2$ - ORTHOGONAL

2) $|\vec{v}_1| = |\vec{v}_2| = 1$

$\vec{x} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$; $\vec{y} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

1) $\vec{x} \cdot \vec{y} = \frac{1}{\sqrt{2}} \cdot \left(\frac{-1}{\sqrt{2}}\right) + \left(-\frac{1}{\sqrt{2}}\right) \cdot \left(\frac{1}{\sqrt{2}}\right)$

$\vec{x} \cdot \vec{y} = -\frac{1}{2} - \frac{1}{2} = -1 \neq 0 \Rightarrow \neg(\vec{x} \perp \vec{y})$

2) $\|\vec{x}\| = \sqrt{\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{-1}{\sqrt{2}}\right)^2}$ $\qquad$ $\|\vec{y}\| = \sqrt{\left(\frac{-1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2}$

$x = \sqrt{\frac{1}{2} + \frac{1}{2}}$ $\qquad$ $y = \sqrt{\frac{1}{2} + \frac{1}{2}}$

$x = 1$ $\qquad$ $y = 1$

A x, y - ORTHONORMAL - FALSE

B x, y - ORTHOGONAL - FALSE

C NONE ABOVE - TRUE

Question 5

If A is a matrix $m \times n$, $m > n$, $\sigma_1, \dots, \sigma_n$ are the singular values of A, $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $A^T A$, which of the following statements is correct?

a. $\sigma_i = \lambda_i, \forall i = 1, \dots, n$;

b. $\sigma_i = \lambda_i^2, \forall i = 1, \dots, n$;

c. $\sigma_i^2 = \lambda_i, \forall i = 1, \dots, n$.

THEORY RECAP:

- def.: SINGULAR VALUES of a matrix A
  $\{\sigma_i\}_{i=1}^n$ are the square roots
  of the eigenvalues of $A^T A$.

- def.: EIGENVALUES OF $A^T A$
  $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of $A^T A$.
  $\sigma_i = \sqrt{\lambda_i}$, $\forall i = \overline{1,n}$

Option : c

Question 6

If A is a matrix $m \times n$, $m > n$, and $A = U\Sigma V^T$ is the SVD of A, $A_p = \sum_{i=1}^p u_i \sigma_i v_i^T$, which of the following statements is correct?

a. $\|A - A_p\|_2 = \sigma_1$;

b. none of the other alternatives;

c. $\|A - A_p\|_2 = \sigma_p$.

! THE ANSWER OF THE EXAM SAYS THAT THE CORRECT ANSWER IS LETTER c.

1) RECALL SVD AND APPROXIMATION A_p:

$A = U\Sigma V^T$

- U, V - orthogonal matrices
- Σ - diagonal matrix containing the SINGULAR VALUES $\{\sigma_i\}_{i=1}^r$

$r = \min\{m,n\} = \text{rank}(A)$
$(r = \min(m,n), m > n)$

2) THE rank-p APPROXIMATION A_p is:

$$A_p = \sum_{i=1}^p \sigma_i \cdot u_i \cdot v_i^T$$

3) $\begin{cases} A = U\Sigma V^T = \sum_{i=1}^r u_i \cdot \sigma_i \cdot v_i^T \\ A_p = U_p \Sigma_p V_p^T = \sum_{i=1}^p u_i \sigma_i v_i^T \end{cases}$

Th. ECKART-YOUNG
$\exists A \in \mathbb{R}^{m \times n}$, $r = \text{rank}(A)$
$\forall p \le r \Rightarrow$ rank approximation is: $\|A - A_p\|_2 = \sigma_{p+1}$

Question 7

Let $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^2 x_2 - x_1 x_2 x_3 + 4x_3^2 x_1$, then $\nabla f(1, 2, 0) =$

a. $(4, 1, -2)$;

b. $(4, 1, 2)$;

c. $(2, 1, -2)$.

$$\vec{\nabla} = \frac{\partial}{\partial x_i} \quad (i = \overline{1,n})$$

$$\vec{\nabla} f(\vec{x}) = \frac{\partial f}{\partial x_i} d\vec{x}_i \quad (i=\overline{1,3}) \qquad \vec{x} \in \mathbb{R}^3$$

$$f(x_1, x_2, x_3) = x_1^2 \cdot x_2 - x_1 x_2 x_3 + 4 x_3^2 x_1$$

$$d\vec{f} = \frac{\partial f}{\partial x_1} d\vec{x}_1 + \frac{\partial f}{\partial x_2} d\vec{x}_2 + \frac{\partial f}{\partial x_3} d\vec{f}_3$$

$$\begin{cases} \frac{\partial f}{\partial x_1} = 2x_1 x_2 - x_2 x_3 + 4x_3^2 \\ \frac{\partial f}{\partial x_2} = x_1^2 - x_1 . x_3 \\ \frac{\partial f}{\partial x_3} = -x_1 x_2 + 8 x_3 x_1 \end{cases} \rightarrow \begin{cases} \left.\frac{\partial f}{\partial x_1}\right|_{(1,2,0)} = 2.1.2 - 2.0 + 4.0^2 = 4 \\ \left.\frac{\partial f}{\partial x_2}\right|_{(1,2,0)} = 1^2 - 1.0 = 1 \\ \left.\frac{\partial f}{\partial x_3}\right|_{(1,2,0)} = -1.2 + 8.0.1 = -2 \end{cases} \Rightarrow d\vec{f} = \begin{pmatrix} 4 \\ 1 \\ -2 \end{pmatrix}$$

Question 8

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^2 - 3x_2$, $g : \mathbb{R}^2 \to \mathbb{R}^2$, $g(t_1, t_2) = (2t_1, 2t_2)$, and if $h : \mathbb{R}^2 \to \mathbb{R}$, $h(t_1, t_2) = f(g(t_1, t_2))$, then $\nabla h(2, 0) =$

a. $(4, 0)$;

b. $(8, -6)$;

c. $(4, -6)$.

<u>GIVEN:</u>

$$F(x_1, x_2) = x_1^2 - 3x_2$$

$$g(t_1, t_2) = (2t_1, 2t_2)$$

$$h(t_1, t_2) = F(g(t_1, t_2))$$

$$\vec{\nabla} h(2,0) = ?$$

1) $h(t_1, t_2) = F(g(t_1, t_2)) = F(2t_1, 2t_2)$

$$h(t_1, t_2) = (2t_1)^2 - 3(2t_2)$$

$$h(t_1, t_2) = 4t_1^2 - 6t_2$$

2) $\vec{\nabla} h(t_1, t_2) = \left(\frac{\partial h}{\partial t_1}, \frac{\partial h}{\partial t_2} \right)$

$$\begin{cases} \frac{\partial h}{\partial t_1} = \frac{\partial}{\partial t_1}(4t_1^2 - 6t_2) = 8t_1 \\ \frac{\partial h}{\partial t_2} = \frac{\partial}{\partial t_2}(4t_1^2 - 6t_2) = -6 \end{cases}$$

$$\vec{\nabla} h(t_1, t_2) = (8t_1, -6)$$

3) $\vec{\nabla} h(2,0) = (16, -6)$ NO OPTION !

Question 9

If $f(\theta) = \|\Phi\theta - y\|_2^2$, with $\Phi \in \mathbb{R}^{N\times D}$, $y \in \mathbb{R}^N$, $\theta \in \mathbb{R}^D$, then $\nabla f(\theta) =$

a. none of the other alternatives;

b. $\Phi^T y - \Phi^T\Phi\theta$;

c. $\Phi^T\Phi\theta - \Phi^T y$.

$\varphi \in \mathbb{R}^{N\times D}$
$y \in \mathbb{R}^N$
$\theta \in \mathbb{R}^D$

$$f(\theta) = \|\varphi\theta - y\|_2^2$$

Using the definition of L2-norm

$$f(\theta) = (\varphi\theta - y)^T \cdot (\varphi\theta - y)$$
$$f(\theta) = (\theta^T\varphi^T - y^T)(\varphi\theta - y)$$

THEORY RECAP:

1) if $\theta \in \mathbb{R}^n \wedge A \in \mathbb{R}^{n\times n}$
$\vec{\nabla}_\theta(\theta^T A\theta) = 2A\theta$ (if A symmetric $A = A^T$)

2) if A is not symmetric
$\vec{\nabla}_\theta(\theta^T A\theta) = (A + A^T)\theta$

3) $\exists\, b \in \mathbb{R}^n$:
$\vec{\nabla}_\theta(\theta^T b) = b$

$\square$ $\exists\, \theta \in \mathbb{R}^n \wedge A \in \mathbb{R}^{n\times n}$

$$\vec{\nabla}_\theta(\theta^T A\theta) = \frac{\partial(\theta^T A\theta)}{\partial\theta_k} = \frac{\partial}{\partial\theta_k}\left(\sum_{i=1}^n\sum_{j=1}^n \theta_i A_{ij}\theta_j\right)$$

$$= \left\| \begin{matrix}\text{for } i = k \\ \text{for } j = k\end{matrix} \right\|$$

$$= \frac{\partial}{\partial\theta_k}\left(\sum_{k=1}^n\sum_{j=1}^n \theta_k A_{kj}\theta_j\right) + \frac{\partial}{\partial\theta_k}\left(\sum_{i=1}^n\sum_{k=1}^n \theta_i A_{ik}\theta_k\right)$$

$$= \sum_{j=1}^n A_{kj}\theta_j + \sum_{i=1}^n \theta_i A_{ik}\Big)$$

$$\vec{\nabla}_\theta(\theta^T A\theta) = (A^T + A)\theta$$

if $A^T = A \Rightarrow \vec{\nabla}_\theta(\theta^T A\theta) = 2A\theta$ ■

$f(\theta) = (\varphi\theta - y)^T(\varphi\theta - y)$, $e = \varphi\theta - y$
$f(e) = e^T e$ $\quad e^T = \theta^T\varphi^T - y^T$

$$\vec{\nabla}_\theta f(e) = \frac{\partial f}{\partial e}\frac{\partial e}{\partial\theta}$$

$$\vec{\nabla}_\theta f(e) = 2e^T\varphi$$

$$\vec{\nabla}_\theta f(\theta) = 2(\varphi\theta - y)^T\varphi$$
$$\vec{\nabla}_\theta f(\theta) = 2(\theta^T\varphi^T - y^T)\varphi$$
$$\vec{\nabla}_\theta f(\theta) = 2(\theta^T\varphi^T\varphi - y^T\varphi) \text{-row}$$

$\theta \in \mathbb{R}^D$: $\theta = \begin{pmatrix}\theta_1 \\ \vdots \\ \theta_D\end{pmatrix} \Rightarrow \theta^T \in \mathbb{R}^D$: $\theta^T = (\theta_1 \dots \theta_D)$

$$\vec{\nabla}_\theta f(\theta) = 2(\varphi^T\varphi\theta - \varphi^T y) \text{-column}$$

$= (\varphi\theta - y)^T(\varphi\theta - y)$
$= (\theta^T\varphi^T - y^T)(\varphi\theta - y)$
$= \theta^T\varphi^T\varphi\theta - \theta^T\varphi^T y - y^T\varphi\theta + y^T y$
$= \theta^T\varphi^T\varphi\theta - 2y^T\varphi\theta + y^T y$

- $\vec{\nabla}_\theta(\theta^T\varphi^T\varphi\theta) = 2\varphi^T\varphi\theta$
- $\vec{\nabla}_\theta(-2y^T\varphi\theta) = -2\varphi^T y$
- $\vec{\nabla}_\theta(y^T y) = 0$

$\Rightarrow \vec{\nabla}_\theta f(\theta) = 2(\varphi^T\varphi\theta - \varphi^T y)$

$\theta \in \mathbb{R}^{D\times 1}$, $y \in \mathbb{R}^{N\times 1}$, $\varphi \in \mathbb{R}^{N\times D}$

$\theta^T_{1\times D}\,\varphi^T_{D\times N}\,y_{N\times 1} = \sum_{i=1}^D\sum_{j=1}^N \theta_i\varphi_{ji}y_j$

$y^T_{1\times N}\,\varphi_{N\times D}\,\theta_{D\times 1} = \sum_{i=1}^N\sum_{j=1}^D y_i\varphi_{ij}\theta_j$

$\Rightarrow \left\|\begin{matrix} i = j \\ j = i\end{matrix}\right\| = \sum_{j=1}^N\sum_{i=1}^D y_j\varphi_{ji}\theta_i$

$\Rightarrow \theta^T\varphi^T y = y^T\varphi\theta$

$\varphi \in \mathbb{R}^{N\times D}$, $\theta \in \mathbb{R}^D$, $y \in \mathbb{R}^N$

$-2\,y^T_{1\times N}\,\varphi_{N\times D}\,\theta_{D\times 1} = -2\,x_{1\times 1}$ - SCALAR

$\exists$ $b \in \mathbb{R}^m$ - const. vector
$x \in \mathbb{R}^m$ - variable vector

$\Rightarrow \vec{\nabla}_x(b^T x) = b$

$(b^T_{1\times m}\, x_{m\times 1}) = a_{1\times 1}$

$\exists$ $x = \theta \wedge b = -2\varphi^T y$

$\Rightarrow \vec{\nabla}_\theta(\underbrace{-2y^T\varphi}_{=b^T}\theta) = b = (b^T)^T$

$b = (-2y^T\varphi)^T \Rightarrow b = -2\varphi^T y$

OBS: THIS WAS THE ANSWER THAT I FOUND. SINCE IT IS NOT AMONG THE OPTIONS, I WOULD SAY THAT THE ANSWER IS A: NONE OF THE OTHER ALTERNATIVES.
HOWEVER, IF WE REWRITE IN TERMS OF COLUMNS WE HAVE ONE ALTERNATIVE SIMILAR TO C

Question 10

If $f : \mathbb{R}^n \to \mathbb{R}$, x^* is a STRICTLY global minimum for f, if:

a. none of the other alternatives;

b. $f(x^*) < f(x), \forall x : \|x - x^*\| < \epsilon, \epsilon > 0$;

c. $f(x^*) \leq f(x), \forall x \in \mathbb{R}^n$.

def.: STRICT LOCAL MINIMUM:

$x_0 \in \mathbb{R}^N$ - STRICT LOCAL MINIMUM OF f iff:

$\exists \varepsilon > 0 : f(x_0) < f(x) \quad \forall x \in \mathbb{R}^N : |x - x_0| < \varepsilon$

def.: STRICT GLOBAL MINIMUM

$x_0 \in \mathbb{R}^N$ - SGM OF f iff:

$f(x_0) < f(x) \quad \forall x \in \mathbb{R}^N$

B definition of STRICT LOCAL MINIMUM

C definition of GLOBAL MINIMUM

Question 11

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^3 + 2x_1x_2^2 - x_2$, which of the following is a stationary point for f?

a. none of the other alternatives;

b. $(1, 0)$;

c. $(0, 0)$.

TO DETERMINE WHICH POINT IS STATIONARY FOR $F(x_1, x_2)$ WE NEED TO COMPUTE THE GRADIENT OF F, SET IT EQUAL TO ZERO, AND SOLVE IT FOR x_1 AND x_2.

1) $\vec{\nabla} f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) \Rightarrow \begin{cases} \frac{\partial f}{\partial x_1} = 3x_1^2 + 2x_2^2 = 0 \\ \frac{\partial f}{\partial x_2} = 4x_1x_2 - 1 = 0 \end{cases} \Rightarrow \begin{matrix} \frac{3}{2}x_1^2 + x_2^2 = 0 \Rightarrow x_1 = x_2 = 0 \\ x_1 \cdot x_2 = \frac{1}{4} \Rightarrow 0 \cdot 0 \neq \frac{1}{4} \end{matrix}$

$x_1^2 \geq 0 \wedge x_2^2 \geq 0$

$\Rightarrow$ both equal ϕ

THUS $(x_1, x_2) = (0, 0)$ - NOT A STATIONARY POINT.

Question 12

If $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$, then the Gradient Descent method $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$ converges to a stationary point of f if:

a. $\alpha_k > 0 \, \forall k \in \mathbb{N}$;

b. $\alpha_k > 0$ is chosen with a backtracking procedure;

c. $\alpha_k \to 0$ as $k \to \infty$.

THEORY RECAP:

- GRADIENT DESCENT METHOD: $X_{k+1} = X_k - \alpha_k \vec{\nabla} f(x_k)$ converges to a stationary point of f, $f : \mathbb{R}^n \to \mathbb{R} \wedge \underbrace{f \in C^1(\mathbb{R}^n)}$

meaning: f is continuosly differentiable

* KEY PROPRIERTIES FOR CONVERGENCE OF GD

1) STATIONARY POINT: $\vec{\nabla} f(x_0) = \vec{0}$
2) STEP SIZE $\alpha_k > 0$:
   - $\alpha_k > 0$ ENSURES THAT DESCENT DIRECTION $-\vec{\nabla} f(x_k)$ DECREASE THE VALUE OF $f(x_k)$
   - IMPROPER CHOISE OF α_k (E.G. TO LARGE OR TO SMALL) CAN CAUSE THE METHOD TO DIVERGE OR CONVERGE TOO SLOWLY.
3) BACKTRACKING LINE SEARCH:
   - PRATICAL STRATEGY TO CHOOSE α_k DINAMICALLY TO ENSURE CONVERGENCE. IT ADJUST α_k AT EACH ITERATION BASED ON THE LOCAL BEHAVIOR OF f AND ENSURES SUFFICIENT DECREASE.
4) DIMINISHING STEP SIZE ($\alpha_k \to 0$ AS $k \to \infty$):
   - IN SOME GD METHODS $\alpha_k \to 0$ AS $k \to \infty$ IS IMPOSED TO ENSURE CONVERGENCE. HOWEVER, IF α_k SHRINKS TOO QUICKY, CONVERGENCE CAN STALL ("STOP")
5) CONVERGENGE CRITERIA:
   - G.D. CONVERGES IF:
     i) α_k - CHOOSEN PROPERLY (const. or using backtracking)
     ii) $\nabla f(x_k) \to 0$ AS $k \to \infty$

A : $\alpha_k > 0 \; \forall k \in \mathbb{N}$ (INSUFFICIENT)

THIS ALONE DOES NOT GARANTEE CONVERVENGE UNLESS α_k IS CHOSEN PROPERLY.
FOR INSTANCE $\perp$ $\alpha_k = \text{const} >> 1$, $\perp$ G.D. MIGHT DIVERGE.

C : $\alpha_k \to 0$ AS $k \to \infty$ (INSUFFICIENT)

WHILE $\alpha_k \to 0$ ENSURES THAT STEP SIZES DIMINISH OVER TIME, IT IS NOT ALWAYS NECESSARY FOR CONVERGENCE.
IF α_k DECREASE TO QUICKLY, CONVERGENCE CAN STALL.
IN PRACTICE: $\begin{cases} 1. \text{ const STEP sizes} \\ 2. \text{ backtracking} \end{cases}$

Question 13

If X is a discrete random variable with target space T, which of the following sentences is TRUE?

a. T is a countable set;

b. none of the other alternatives;

c. T is an interval.

THEORY RECAP:

1. DISCRETE RANDOM VARIABLE:

A random variable X is discrete if it can take on only a COUNTABLE set of values.

- rolling a die ($X \in \{1, 2, 3, 4, 5, 6\}$)
- flipping a coin ($X \in \{0, 1\}$)
- number of customers entering a store ($X \in \mathbb{N}$)

The values that X can take form is TARGET SPACE T

2. TARGET SPACE T:

The target space T is the set of all possible outcomes X can take.

For DISCRETE RANDOM VARIABLES, T is a COUNTABLE SET.

- T can be finite (rolling a die)
- T can be infinite but countable (number of costumers $\in \mathbb{N}$)

c) T is an interval

FALSE - because interval implies an uncontable set.

Question 14

If $X \in \mathbb{R}^m, Y \in \mathbb{R}^n$ are univariate random variables and $\text{cov}(X, Y)$ is the covariance between X and Y, then:

a. $\text{cov}(X, Y) \in \mathbb{R}$;

b. $\text{cov}(X, Y) \in \mathbb{R}^{m \times n}$;

c. $\text{cov}(X, Y) \in \mathbb{R}^{n \times m}$.

TO SOLVE THIS WE NEED TO DETERMINE THE SIZE AND TYPE OF THE COVARIANCE MATRIX $\text{cov}(X, y)$ WHEN $X \in \mathbb{R}^m$ AND $y \in \mathbb{R}^n$.

THEORY RECAP:

1. COVARIANCE DEFINITION:

Covariance between 2 random variables X and y measures the LINEAR RELATIONSHIP between their components.

$$\text{cov}(X, y) = \mathbb{E}\left[(X - \mathbb{E}[X])(y - \mathbb{E}[y])^T\right]$$

- X : random vector $\in \mathbb{R}^m$
- y : random vector $\in \mathbb{R}^n$
- $\mathbb{E}[X]$ and $\mathbb{E}[y]$: expectations (means) of X and y.
- The RESUL is a MATRIX

2. SIZE OF THE COVARIACE MATRIX

- X has m components
- y has n components
- $\text{cov}(X, y)$ captures all pairwise covariance between components of X and y.
  - The entry $(\text{cov}(x, y))_{ij}$ is the covariance between the i-th comp. of X and j-th comp. y.

Thus, $\text{cov}(X, y) \in \mathbb{R}^{m \times n}$

(A) $\text{cov}(x, y) \in \mathbb{R}$ (INCORRECT)

Because covariance is <u>NOT A ESCALAR</u>.

It's a matrix when X and y are multivariate.

(C) $\text{cov}(x, y) \in \mathbb{R}^{n \times m}$ (INCORRECT)

Because covariance has rows corresponding to components of the 1st element and columns corresponding to components of the 2nd element.

Since $X \in \mathbb{R}^m \wedge y \in \mathbb{R}^n \Rightarrow \text{cov}(x, y) \in \mathbb{R}^{m \times n}$

Question 15

If $X : \Omega \to T$ is a continuous random variable and p is the PDF of X, then the distribution mean (or expectation) is defined as:

a. $\mathbb{E}(x) = \int_T p(x)dx$;

b. $\mathbb{E}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$;

c. none of the other alternatives.

THEORY RECAP:

EXPECTATION FOR CONTINUOS RANDOM VARIABLES.

1. EXPECTED VALUE (MEAN):

The expected value (or mean) of a CONTINUOS RANDOM VARIABLE X is a measure of the "central tendency" of its distribution. It is defined as:

$$\mathbb{E}[X] = \int_T x\, p(x)\, dx$$

- T : domain of X ($\mathbb{D}m(X)$, set of all possible value X can take).
- p(x): probability density function (PDF) of X.
- x : variable of integration.

2. KEY PROPERTIES OF THE EXPECTATION:

- FOR CONTINUOS RANDOM VARIABLES: The expectation is calculate as an integral using the PDF p(x).
- FOR DISCRETE RANDOM VARIABLES: The expectation is a weighted sum of the possible outcomes:

$$\mathbb{E}[X] = \sum_i x_i . P(X = x_i)$$

- FOR SAMPLE DATA: The mean of N observed sample $x_1, x_2, \dots, x_N$ is:

$$\mathbb{E}[X] \approx \frac{1}{N}\sum_{i=1}^{N} x_i$$

OBS: This is a empirical approximation, not the exact expectation

A) $\mathbb{E}(x) = \int_T p(x)\, dx$ (INCORRECT)

This integral represent the TOTAL PROBABILITY of PDF, which must be equal to 1.

B) $\mathbb{E}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$ (INCORRECT)

This formula computes the SAMPLE MEAN from N obsrved data points While the sample mean approximates the expectations, it is not the formal definition for the expectation of a cont. rand. var.

Question 16

Given two random variables X and Y such that $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ and $p(y|x) = ce^{-|y-ax|}$, then the MLE reads:

a. $x^* = \arg\min_x |y - ax| + x^2$;

b. $x^* = \arg\min_x |y - ax|$;

c. $x^* = \arg\min_x \frac{1}{2}(y - ax)^2$.

THEORY RECAP:

MAXIMUM LIKELIHOOD ESTIMATION (MLE):

1. MAXIMUM LIKELIHOOD ESTIMATION:
   It is the value of parameters x_0 that maximize the likelihood function $p(y|x)$ given the observed data y.
2. LOG-LIKELIHOOD:
   To simplify the calculation, we often take the logarithm of the likelihood function because it transforms products into sums and simplifies differentiation.
3. PROCEDURE:
   i) starts with the likelihood $p(y|x)$;
   ii) Take the negative log of $p(y|x)$ to form the objective function;
   iii) Minimize the negative log-likelihood to find x_0.

Given: 1. $p(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$ (Standard Gaussian prior for x)

2. $p(y|x) = c \cdot \exp(-|y-ax|)$ ($c = \text{const}$)

Goal: To find MLE, which depends only on $p(y|x)$, because $p(x)$ is INDEPENDENT of y and doesn't affect the maximization with respect to x.

$$p(y|x) = c \cdot \exp(-|y-ax|) \quad | \ln(\cdot)$$

$$\ln(p(y|x)) = \ln(c \cdot \exp(-|y-ax|))$$
$$\ln(p(y|x)) = \ln c + \ln(\exp(-|y-ax|))$$
$$\ln(p(y|x)) = \ln c - |y-ax| \cdot \underbrace{\ln e}_{=1}$$
$$\ln(p(y|x)) = \ln c - |y-ax| \quad | \times(-1)$$
$$-\ln(p(y|x)) = |y-ax| + c' \quad (c' = \ln c = \text{const})$$

To MAXIMIZE the likelihood, MINIMIZE the negative log-likelihood:

$$\Rightarrow x_0 = \arg\min_x |y-ax|$$

Question 17

In linear regression, we suppose that the measured data are:

a. Not affected by noise;

b. Affected by noise with normal distribution with mean equal to 0;

c. Affected by noise with normal distribution with mean equal to 1.

THEORY RECAP:

NOISE IN LINEAR REGRESSION

Linear regression assumes that the relationship between the dependent variable y and the independent variable X can be modeled as a linear function:

$$y = X\beta + \varepsilon$$

- X : matrix of independent variable (PREDICTORS)
- β : vector of regression coefficients.
- ε : error term of noise.

ASSUMPTIONS ABOUT THE NOISE (ε):

1. MEAN OF THE NOISE:
   - The noise is assumed to have a mean of 0: $\mathbb{E}[\varepsilon] = 0$
   - This ensures that, on average, the prediction are UNBIASED.
2. DISTRIBUTION OF THE NOISE:
   - The noise is typically assumed to follow a NORMAL DISTRIBUTION: $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

     σ^2 is the variance of the noise.
3. INDEPENDENCE:
   - The noise terms are INDEPENDENT of each other and of the predictors.

(A) NOT AFFECT BY NOISE (INCORRECT)

In linear regression the data is explicitly assumed to be affected by noise. The noise accounts for the randomness or variability not captured by the predictors.

(C) AFFECTED BY NOISE WITH NORMAL DISTRIBUTION WITH MEAN EQUAL TO 1. (INCORRECT)

The assumption in LR is that mean of noise is ZERO.
A non-zero mean woul bias the prediction.

(B) AFFECTED BY NOISE WITH NORMAL DISTRIBUTION WITH MEAN EQUAL TO 0. (CORRECT)

- Mean $\mathbb{E}(\varepsilon) = 0$
- Variance $Var(\varepsilon) = \sigma^2$

Question 18

Let A be the matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

What is the output of Python code: `B = A[:, 1:2]`?

a. None of the others;

b.

$$B = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}$$

c.

$$B = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix}$$

```python
import numpy as np

A = np.array([
    [1, 2, 3],
    [4, 5, 6],
    [7, 8, 9],
])

B = A[:, 1:2]
print(B)

```

✓ 0.0s

```
[[2]
 [5]
 [8]]
```

Question 19

Importing numpy with the instruction: `import numpy as np` indicate which of the following instructions creates the array:

$$x = [0\ 0.1\ 0.2\ 0.3]$$

a. `x=np.linspace(0,0.3,4);`

b. `x=np.linspace(0,0.4,4);`

c. `x=np.linspace(0,0.4).`

```python
import numpy as np

x = [0, 0.1, 0.2, 0.3]

x_a = np.linspace(start=0, stop=0.3, num=4)
x_b = np.linspace(start=0, stop=0.4, num=4)
x_c = np.linspace(start=0, stop=0.4)

try:
    a = np.allclose(x_a, x)
except ValueError as e:
    a = False

try:
    b = np.allclose(x_b, x)
except ValueError as e:
    b = False

try:
    c = np.allclose(x_c, x)
except ValueError as e:
    c = False

print(f"a: {a}")
print(f"b: {b}")
print(f"c: {c}")

```

✓ 0.0s

```
a: True
b: False
c: False
```

Question 20

Which of the following Python instructions returns the matrix multiplication between two arrays A and B?

a. `np.dot(A,B);`

b. `A**B;`

c. `A*B.`

```python
import numpy as np

A = np.array([
    [1, 2],
    [3, 4],
])
B = np.eye(2)
print(f"A:\n {A}"); print(f"B:\n {B}"); print("-" * 20)
a = A @ B; # Matrix multiplication
b = A ** B; # Element-wise exponentiation
c = A * B # Element-wise multiplication
print(f"A @ B:\n {a}"); print("-" * 20)
print(f"A ** B:\n {b}"); print("-" * 20)
print(f"A * B:\n {c}"); print("-" * 20)
```

[26] ✓ 0.0s

```
A:
 [[1 2]
 [3 4]]
B:
 [[1. 0.]
 [0. 1.]]
--------------------
A @ B:
 [[1. 2.]
 [3. 4.]]
--------------------
A ** B:
 [[1. 1.]
 [1. 4.]]
--------------------
A * B:
 [[1. 0.]
 [0. 4.]]
--------------------
```

Question 21

The normalized scientific representation of 45.6 is:

a. $4.56 \cdot 10^1$;

b. $0.0456 \cdot 10^3$;

c. $0.456 \cdot 10^2$. (EXAM'S ANSWER)

THEORY RECAP:

NORMALIZED SCIENTIFIC NOTATION

Normalized scientific notation express a number as:

$$a \cdot 10^n$$

- $1 \leq |a| < 10$
- $n \in \mathbb{Z}$
- The value of the number remains unchanged

$$45,6 = 4,56 \cdot 10^1$$

OBS: The answer in the exam was c) $0,456 \cdot 10^2$. I don't understand why. If there is any sugestion, please, get in contact.

Question 22

If $x = 1.47 \cdot 10^{-4}$ and $y = 1$, which is the result of $z = x + y$ in $\mathbb{F}(10, 4, -5, 5)$?

a. none of the other alternatives;

b. 1;

c. 1.0001.

$F(10, 4, -5, 5)$

$$\begin{cases} \beta = 10 & \text{(base)} \\ p = 4 & \text{(precision)} \\ e_{min} = -5 & \text{(exponent must lie between emin and emax)} \\ e_{max} = 5 \end{cases}$$

$$\begin{cases} x = 1.47 \cdot 10^{-4} \\ y = 1 \cdot 10^0 \end{cases} \Longrightarrow \begin{cases} x = 0,000147 \cdot 10^0 \\ y = 1 \cdot 10^0 \end{cases}$$

align with the highest exponent

$z = x + y$

$z = \underbrace{1,000}_{\text{4 digits (precision)}}147 \cdot 10^0 \Rightarrow z \approx 1,000 \cdot 10^0 \Rightarrow z = 1$

$-4 \leq 0 \leq 4$

Question 23

If you generate with NUMPY the following array: $x =$ `numpy.arange(5)`, then $\|x\|_1$ is:

a. 13;

b. $\sqrt{55}$;

c. 10.

$x = \text{np.arange}(\text{stop} = 5)$

OBS: "np.arange(n)" starts from zero and stop in $(n-1)$

$x = [0, 1, 2, 3, 4]$

L_p-norm: $\|\vec{x}\|_p = \sum_{i=1}^{n} (|x_i|^p)^{1/p}$

$$\|x\|_1 = \sum_{i=1}^{n} (|x_i|^1)^{1/1}$$

$$\|x\|_1 = (|0| + |1| + |2| + |3| + |4|)^1$$

$$\|x\|_1 = 10$$

Question 24

If $U \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then:

a. $U^T = U$;

b. U is symmetric;

c. $U^T = U^{-1}$.

THEORY RECAP:

ORTHOGONAL MATRIX

1. DEFINITION: $\exists U \in \mathbb{R}^{n \times n}$ is ORTHOGONAL if: $U^T U = U U^T = E$ $\begin{cases} \bullet\ U^T: \text{transpose of } U. \\ \bullet\ E: \text{identity matrix.} \end{cases}$

2. KEY PROPRIETIES OF ORTHOGONAL MATRICES (OM):
   - $U^T = U^{-1}$
   - OM preserves norms, meaning $\|U\vec{x}\| = \|\vec{x}\| \quad \forall \vec{x} \in \mathbb{R}^n$
   - OM preserves angles between vectors.
   - OM are not necessarily symmetric.

(A) $U^T = U$ (INCORRECT)
- This property implies that U is symmetric $(U^T = U)$
- OM are not necessarily symmetric. For example: $U = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ $\begin{cases} \bullet\ \text{ORTHOGONAL} \\ \bullet\ \text{NOT SYMMETRIC} \end{cases}$

(B) U is symmetric (INCORRECT)
- This is not true in general for OM.
- A matrix can be orthogonal without being symmetric
- Symmetry is a stricter condition than orthogonality.

(C) $U^T = U^{-1}$ (CORRECT)
- This is the defining property of OM
- by definition: $U^T U = E \mid \times U^{-1} \Rightarrow U^T U U^{-1} = E U^{-1} = U^T E = E U^{-1} \Rightarrow U^T = U^{-1}$

Question 25

If A is a matrix $m \times n$, $m > n$, $\text{rg}(A) = k$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are the singular values of A, which of the following sentences is correct ($k_2(A)$ is the condition number of A in norm 2):

a. $k_2(A) = \frac{\sigma_1}{\sigma_2}$;

b. none of the other alternatives;

c. $k_2(A) = \frac{\sigma_1}{\sigma_n}$.

Theory Recap:

Condition Number in the 2-Norm

1. **Definition of Condition Number (CN):**
   - CN of a matrix A in the 2-norm is defined as: $K_2(A) = \frac{\sigma_1}{\sigma_{min}}$
     - σ_1: largest singular value of A
     - σ_{min}: smallest nonzero singular value of A (if A has full rank)
2. **Proprieties of Singular Values:**
   - Singular Values: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{min} > 0$
   - If A has rank K, σ_k is the smallest nonzero SV, and $\sigma_{k+1}, \dots, \sigma_n = 0$
3. **Importance of CN:**
   - The CN measures the sensitivity of solution to changes in the input. A large $K_2(A)$ indicates numerical instability.

☐

i) Condition Number of $A \in \mathbb{R}^{m \times n}$ in the 2-norm is defined as: $K_2(A) = \|A\|_2 \cdot \|A^{-1}\|_2$

- $\|A\|_2$: operator 2-norm. $\|A\|_2 = \sup_{\|x\|_2 = 1} \|A\vec{x}\|_2 = \sigma_1$
- $\|A^{-1}\|_2$: operator 2-norm of inverse of A. $\|A^{-1}\|_2 = \sup_{\|x\|_2 = 1} \|A^{-1}\vec{x}\|_2 = \frac{1}{\sigma_{min}}$

ii) Proprieties of SV:

1. Singular Values: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{min} > 0$ are the square roots of the eigenvalues of $A^T A$
2. The largest SV σ_1 represents the max stretching of a vector by A.
3. The smallest SV σ_{min} represents the min stretching of a vector by A.

iii) Substituting the norms

$$\begin{cases} \|A\|_2 = \sigma_1 \\ \|A^{-1}\| = \frac{1}{\sigma_{min}} \end{cases} \Rightarrow K_2(A) = \|A\|_2 \|A^{-1}\|_2 \qquad K_2(A) = \frac{\sigma_1}{\sigma_{min}}$$

(A) $K_2(A) = \frac{\sigma_1}{\sigma_2}$ (INCORRECT) because σ_2 is not necessarily the smallest singular value.

(C) $K_2(A) = \frac{\sigma_1}{\sigma_2}$ (CORRECT)
- σ_n - smallest SV ($\sigma_n = \sigma_{min}$)
- For full-rank matrices, $\sigma_{min} = \sigma_n$

Question 26

If the matrix A has singular values $\sigma_1 = 7$, $\sigma_2 = 4$, $\sigma_3 = 1$. Then:

a. $\|A\|_2 = 1$;

b. $k_2(A) = 4$;

c. $\|A\|_2 = 7$.

i) $K_2(A) = \frac{\sigma_1}{\sigma_{min}} \Rightarrow k_2(A) = \frac{7}{1} \Rightarrow k_2(A) = 7$

ii) $\|A\|_2 = \sigma_1 \Rightarrow \|A\|_2 = 7$

Question 27

Let $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^2 x_2 - x_1 x_2 x_3 + 4x_3^2 x_1$, then $\nabla f(1, 2, 0) =$:

a. $(4, 1, 0)$;

b. none of the other alternatives;

c. $(4, 1, -2)$.

$f(x_1, x_2, x_3) = x_1^2 . x_2 - x_1 . x_2 . x_3 + 4 . x_3^2 . x_1$

$\vec{\nabla} f(1, 2, 0) = ?$

i) $\vec{\nabla} f(x_1, x_2, x_3) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \right)$

$$\begin{cases} \frac{\partial f}{\partial x_1} = 2 \cdot x_1 \cdot x_2 - x_2 \cdot x_3 + 4 x_3^2 \\ \frac{\partial f}{\partial x_2} = x_1^2 - x_1 \cdot x_3 \\ \frac{\partial f}{\partial x_3} = -x_1 . x_2 + 8 x_1 x_3^2 \end{cases}$$

ii) $\left.\frac{\partial f}{\partial x_1}\right|_{(1,2,0)} = 2.1.2 - 2.0 + 4.0^2 = 4 - 0 + 0 = 4$

$\left.\frac{\partial f}{\partial x_2}\right|_{(1,2,0)} = 1^2 - 1.0 = 1 - 0 = 1$ $\Rightarrow (4, 1, -2)$

$\left.\frac{\partial f}{\partial x_3}\right|_{(1,2,0)} = -1.2 + 8.1.0 = -2 + 0 = -2$

Question 28

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^2 - 3x_2$, $g : \mathbb{R}^2 \to \mathbb{R}^2$, $g(t_1, t_2) = (2t_1, 2t_2)$, and if $h : \mathbb{R}^2 \to \mathbb{R}$, $h(t_1, t_2) = f(g(t_1, t_2))$, then $\nabla h(1, 1) =$:

a. $(8, -6)$;

b. $(4, -6)$;

c. none of the other alternatives.

$$F(x_1, x_2) = x_1^2 - 3x_2$$
$$g(t_1, t_2) = (2t_1, 2t_2)$$

$$h(t_1, t_2) = F(g(t_1, t_2)) = F(t_1, t_2) = (2t_1)^2 - 3(2t_2)$$

$$h(t_1, t_2) = 4t_1^2 - 6t_2$$

$$\vec{\nabla} h(t_1, t_2) = \left(\frac{\partial h}{\partial t_1}, \frac{\partial h}{\partial t_2}\right) = (8t_1, -6)$$

$$\vec{\nabla} h(1, 1) = (8, -6)$$

AGAIN NO ANSWER ?!

Question 29

If $f(\theta) = \|\Phi\theta - y\|_2^2$, with $\Phi \in \mathbb{R}^{N \times D}, y \in \mathbb{R}^N, \theta \in \mathbb{R}^D$, then $\nabla f(\theta) =$:

a. $\Phi^T\Phi\theta - \Phi^T y$;

b. $2\Phi^T\Phi\theta - 2\Phi^T y$;

c. $\Phi\theta - y$.

LOOK UP TO THE QUESTION NUMBER 09

$$F(\theta) = \|\varphi\theta - y\|_2^2$$

$$\|\varphi\theta - y\|_2^2 = (\varphi\theta - y)^T \cdot (\varphi\theta - y) \qquad \varphi\theta - y = e$$

$$F(\theta) = (\theta^T\varphi^T - y^T)(\varphi\theta - y)$$
$$F(\theta) = \theta^T\varphi^T\varphi\theta \underbrace{- \theta^T\varphi^T y - y^T\varphi\theta}_{\theta^T\varphi^T y = y^T\varphi\theta} + y^T y$$

$$F(\theta) = \theta^T\varphi^T\varphi\theta - 2y^T\varphi\theta + y^T y$$

$$\vec{\nabla}_\theta F(\theta) = \underbrace{\vec{\nabla}_\theta(\theta^T\varphi^T\varphi\theta)}_{= 2\varphi^T\varphi\theta} - 2\underbrace{\vec{\nabla}_\theta(y^T\varphi\theta)}_{= \varphi^T y} + \underbrace{\vec{\nabla}_\theta(y^T y)}_{= 0}$$

$$\vec{\nabla}_\theta F(\theta) = 2\varphi^T\varphi\theta - 2\varphi^T y$$

Question 30

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, which of the following sentences is FALSE:

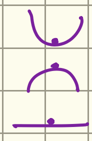
a. if x^* is a local minimum for f, then $\nabla f(x^*) = 0$;

b. $\nabla f(x^*) = 0$, if and only if x^* is a stationary point for f;

c. $\nabla f(x^*) = 0$, if and only if x^* is a local minimum for f.

KEY THEORY:

1. LOCAL MINIMUM AND GRADIENT:
   - @ LOCAL MINIMUM x_0 of a differentiable f, the $\vec{\nabla} f(x_0) = 0$. This is a NECESSARY CONDITION, but is not sufficient for a local minimum.
2. STATIONARY POINT:
   - A point x_0 is called a STATIONARY POINT if $\vec{\nabla} f(x_0) = 0$. Could be { Local minimum / local maximum / Saddle point



3. SUFFICIENCY OF $\vec{\nabla} f(x_0) = 0$:
   - $\vec{\nabla} f(x_0) = 0$ DOES NOT GARANTEE that x_0 is a local minimum.

(A) if x_0 - LOCAL MINIMUM for f, then $\vec{\nabla} f(x_0) = 0$

(TRUE) because the gradient $= 0$ @ any local minimum of a differentiable f.

(B) $\vec{\nabla} f(x_0) = 0$ iff x_0 is a STATIONARY POINT of y.

(TRUE) $\forall x_0 : \vec{\nabla} f(x_0) = 0 \Rightarrow x_0$ - SP, regardless SP is min, max or saddle.

(C) $\vec{\nabla} f(x_0) = 0$ iff x_0 is a LOCAL MINIMUM for f.

(FALSE) $\vec{\nabla} f(x_0) = 0$ is a NECESSARY condition it is NOT A SUFFICIENT.

Example: @ x_0 - SADDLE POINT $\Rightarrow \vec{\nabla} f(x_0) = 0$

Question 31

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 2\sin x_1 + \sin x_1 \cos x_2$, which of the following is **NOT** a stationary point for f?

a. $(0, \pi)$

b. $(\pi/2, 0)$

c. $(\pi/2, \pi)$

$f(x_1, x_2) = 2\sin x_1 + \sin x_1 \cdot \cos x_2$

$\vec{\nabla} f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)$

$$\begin{cases} \frac{\partial f}{\partial x_1} = 2 \cdot \cos x_1 + \cos x_1 \cdot \cos x_2 \\ \frac{\partial f}{\partial x_2} = -\sin x_1 \cdot \sin x_2 \end{cases} \Rightarrow \begin{cases} \cos x_1 \cdot (2 + \cos x_2) = 0 & (i) \\ -\sin x_1 \cdot \sin x_2 = 0 & (ii) \end{cases}$$

i) $\cos x_1 (2 + \cos x_2) = 0 \iff \cos x_1 = 0 \lor (2 + \cos x_2) = 0$

1) $\cos x_1 = 0$

$\Rightarrow x_1 = \pi/2 + k_1\pi \quad (k_1 \in \mathbb{Z}) : C_1$

2) $2 + \cos x_2 = 0$

$\cos x_2 = -2 \quad \nexists\, x_2 \in \mathbb{R}$. T.k. $\forall x \in \mathbb{R} : |\cos x| \leq 1 : C_3$

ii) $\sin x_1 \cdot \sin x_2 = 0 \iff \sin x_1 = 0 \lor \sin x_2 = 0$

1) $\sin x_1 = 0$

$x_1 = k_3\pi \quad (k_3 \in \mathbb{Z}) : C_3$

2) $\sin x_2 = 0$

$x_2 = k_4\pi \quad (k_4 \in \mathbb{Z}) : C_4$

iii) $C_1 : \{ x_1 = \frac{\pi}{2} + k_1\pi, k_1 \in \mathbb{Z} \}$

iv) $C_2 \cap C_4 : \{ x_2 = \pi + 2k_2\pi, k_2 \in \mathbb{Z} \} \cap \{ x_2 = 2k_4\pi, k_4 \in \mathbb{Z} \} \Rightarrow k \in \mathbb{Z} : x_2 = \pi + 2k\pi$

$$\begin{cases} x_1 = \frac{\pi}{2} + k_1\pi \\ x_2 = \pi + 2k_2\pi \end{cases} \quad (k_1, k_2 \in \mathbb{Z})$$

A) $(0, \pi)$ (FALSE)

$\frac{\partial f}{\partial x_1} = \cos x_1 (2 + \cos x_2) \xRightarrow{(0,\pi)} \cos 0 \cdot (2 + \cos\pi)$

$= 1 \cdot (2 + (-1)) = 1 \neq 0$ NOT STATIONARY POINT

B) $(\pi/2, 0)$ (TRUE)

$\frac{\partial f}{\partial x_1} = \cos x_1 (2 + \cos x_2) \xRightarrow{(\frac{\pi}{2},0)} \cos\frac{\pi}{2} \cdot (2 + \cos 0)$

$= 0 \cdot (2 + 1) = 0$ STATIONARY POINT

C) $(\pi/2, \pi)$ (TRUE)

$\left.\frac{\partial f}{\partial x_1}\right|_{(\frac{\pi}{2},\pi)} = \cos\frac{\pi}{2} \cdot (2 + \cos\pi) = 0 \cdot (2 - 1) = 0$ STATIONARY POINT

Question 32

If $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$, then x^* is a minimum point if and only if:

a. $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite.

b. $\nabla f(x^*) = 0$.

c. $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semi-definite.

THEORY RECAP:

1. FIRST-ORDER NECESSARY CONDITION:

- x_0 is a stationary point if $\vec{\nabla} f(x_0) = 0$
- This condition is necessary for x_0 to be: $\begin{cases} \text{min} \\ \text{max} \\ \text{saddle} \end{cases}$

2. SECOND-ORDER NECESSARY CONDITION:

- If x_0 is LOCAL MINIMUM the Hessian matrix $\vec{\nabla}^2 f(x_0)$ must be posite semi-definite:

$$\vec{v}^T \vec{\nabla}^2 f(x_0) \vec{v} \geq 0, \quad \forall \vec{v} \in \mathbb{R}^n$$

$$(\vec{v}^T \cdot \Delta f(x_0) \cdot \vec{v} \geq 0)$$

$$\begin{cases} v^T \Delta f \, v \geq 0 \\ v^T \Delta f \, v \leq 0 \end{cases}$$

$$\begin{cases} v^T \Delta f \, v > 0 \\ v^T \Delta f \, v < 0 \end{cases}$$

3. SECOND-ORDER SUFFICIENT CONDITION:

- To ensure that x_0 is a STRICT LOCAL MINIMUM the HESSIAN matrix $\vec{\nabla}^2 f(x_0)$ must be positive definite:

$$\vec{v}^T \cdot \vec{\nabla}^2 f(x_0) \cdot \vec{v} > 0, \quad \forall \vec{v} \neq \vec{0}$$

(A) $\vec{\nabla} f(x_0) = \vec{0} \wedge \vec{\nabla}^2 f(x_0) > 0$ (positive definite) (CORRECT)

(B) $\vec{\nabla} f(x_0) = \vec{0}$ (INCORRECT)

$\vec{\nabla} f(x_0) = \vec{0}$ - NECESSARY but NOT SUFFICIENT for x_0-min.
It only garantees that x_0 is a stationary point.

(C) $\vec{\nabla} f(x_0) = \vec{0} \wedge \vec{\nabla}^2 f(x_0) \geq 0$ (positive semi-definite) (INCORRECT)

NECESSARY but NOT SUFFICIENT for min.
POSITIVE SEMI-DEFINITE $\Rightarrow \begin{cases} \text{min} \\ \text{saddle} \end{cases}$

Question 33

If X is a continuous random variable, which of the following is a possible target space for X?

a. none of the other alternatives

b. $T = \mathbb{N}$ (the natural numbers)

c. $T = [1, 2] \cup [3, 4]$

THEORY RECAP:

CONTINUOS RANDOM VARIABLES

1. DEFINITION: A random variable X is CONTINUOS if its set of possible values (target spaces) forms a CONTINUOS SUBSET of $\mathbb{R}$
   - X can take any value in a interval or union of intervals.
   - The probability of X taking any single value is zero $P(X=x) = 0$

2. EXAMPLES OF TARGET SPACES:
   - $[a;b]$, $(a;b)$, $[a;\infty)$
   - $[a;b] \cup [c;d]$
   - Uncountable set within $\mathbb{R}$

3. CONTRAST WITH DISCRETE RANDOM VARIABLES:
   A discrete random variable has a COUNTABLE TARGET SPACE ($\mathbb{N}$, $\mathbb{Z}$ or a finite set)

B $T = \mathbb{N}$ (INCORRECT)

$\mathbb{N}$ IS COUNTABLE

Question 34

The expected value of a continuous random variable X with target space $T = [a, b]$ is defined as (with $p(x)$ the PDF function):

a. $\mathbb{E}(x) = \int_a^b xp(x)dx$

b. $\mathbb{E}(x) = \int_a^b \frac{p(x)}{x}dx$

c. $\mathbb{E}(x) = \int_a^b p(x)dx$

THEORY RECAP:

EXPECTED VALUE

1. DEFINITION: Expected Value - For a CONTINUOUS RANDOM VALUE VARIABLE X with probability density function (PDF) $p(x)$ and target space $T = [a;b]$, the EV $\mathbb{E}(x)$ is defined as:

$$\mathbb{E}(x) = \int_a^b x . p(x) . dx$$

- x : random variable;
- $p(x)$: PDF, satisfying $\int_a^b p(x) . dx = 1$

2. KEY POINTS:
   - The factor x in the integral comes from the weighted average interpretation of the expected value.
   - The term $p(x)$ ensures that probabilities are properly accounted for in the computation.

(B) $\mathbb{E}(x) = \int_a^b \frac{p(x)}{x} dx$ (INCORRECT)

(C) $\mathbb{E}(x) = \int_a^b p(x) dx = 1$ (INCORRECT)
↳ represents the total probability (not the expected value)

Question 35

F_X is the probability mass function (PMF) of a random variable X. Which of the following sentences is TRUE (with Ω sample space and T target space)?

a. $F_X : \Omega \to T$

b. $F_X : T \to \mathbb{R}$

c. $F_X : \Omega \to \mathbb{R}$

Theory Recap:

Probability Mass Function:

1. What is PMF?
   - PMF $F_X(x)$ of a discrete random variable X maps the values in the TARGET SPACE T of X to their respective probabilities

     $F_X : T \to [0;1]$

     - T is the target space (possible values of X)

2. PMF Properties:
   - For each $t \in T$, $F_X(t) \geq 0$
   - The sum of the probabilities over T equals 1:

     $\sum_{t \in T} F_X(t) = 1$

3. Mapping in Probability Theory:
   - The random variable X itself is defined as $X : \Omega \to T$ (Ω sample space)
   - The PMF F_X operates on T and output probabilities, not on Ω

(A) $F_X : \Omega \to T$ (INCORRECT)
- F_X (PMF) maps values in the TARGET SPACE T to probabilities in $\mathbb{R}$ (specifically to $[0;1]$)
- The mapping $\Omega \to T$ describes the random variable X, not the PMF.

(C) $F_X : \Omega \to \mathbb{R}$ (INCORRECT)
- Incorrect because F_X operates on T (target space), not directly on the Ω (sample space)

(B) $F_X : T \to \mathbb{R}$ (CORRECT)
- Correct because F_X takes an input from T (target space) and output a probability in $[0;1]$ which is a subset of $\mathbb{R}$

1. Probability Mass Function (PMF)

2. Probability Density Function (PDF)

3. Cumulative Distribution Function (CDF)

4. Differences Between PMF, PDF, and CDF

Property	PMF	PDF	CDF
Type of Variable	Discrete	Continuous	Both
Definition	$P(X = x)$	Probability density	$P(X \leq x)$
Domain	$T \subset \mathbb{Z}$	$\mathbb{R}$	$\mathbb{R}$
Integration/Summation	$\sum_{x \in T} F_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

Question 36

Given two random variables X and Y such that $p(x) = ce^{-|x|}$ and $p(y|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-ax)^2}$, then the MAP reads:

a. $x^* = \arg\min_x \frac{1}{2}(y-ax)^2 + \frac{1}{2}x^2$

b. $x^* = \arg\min_x \frac{1}{2}(y-ax)^2 + |x|$

c. $x^* = \arg\min_x \frac{1}{2}(y-ax)^2$

THEORY RECAP:

MAP ESTIMATION

1. DEFINITION:

The MAP estimate maximazes the posteriori probability:

$$x_0 = \arg\max_x p(x|y) = \arg\min_x -p(x|y)$$

2. BAYES' RULE:

$$\underbrace{p(x|y)}_{=\text{posterior}} = \frac{\underbrace{p(y|x)}_{=\text{likelihood}} \cdot \underbrace{p(x)}_{=\text{prior}}}{\underbrace{p(y)}_{=\text{evidence}}}$$

$$\Rightarrow \arg\max_x p(x|y) = \arg\max_x \underbrace{\frac{p(y|x)\cdot p(x)}{p(y)}} = \arg\max_x p(y|x)\cdot p(x)$$

since $p(y)$ is const with respect to x.

3. LOGARITHMIC TRANSFORMATION:

Since log or ln are monotonic function we can take the log or ln of the posterior:

$$\arg\max_x \{\ln p(x|y)\} = \arg\max_x \{\ln p(y|x) + \ln p(x)\}$$

$$\arg\max_x \{f(x)\} = \arg\min_x \{-f(x)\}$$

$$\arg\min_x \{-\ln p(x|y)\} = \arg\min_x \{-\ln p(y|x) - \ln p(x)\}$$

1. $p(x) = c \cdot \exp(-|x|)$

$\ln p(x) = \ln c - |x|$

$-\ln p(x) = |x| + \text{const}$

2. $p(y|x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(y-ax)^2}{2}\right)$

$\ln p(y|x) = -\ln(\sqrt{2\pi}) - \frac{(y-ax)^2}{2}$

$-\ln p(y|x) = \frac{(y-ax)^2}{2} + \text{const}$

$$\Rightarrow x_0 = \arg\min_x \left\{ \frac{(y-ax)^2}{2} + |x| \right\}$$

Question 37

Suppose a set of data $(x_i, y_i), i = 1, \ldots, N, y_i = f(x_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. In linear regression, the likelihood function is:

a. $p(y|x,\theta) = \prod_{i=1}^{N} \mathcal{N}(y|x_i^T\theta x_i, \sigma^2)$

b. $p(y|x,\theta) = \prod_{i=1}^{N} \mathcal{N}(y|x_i^T\theta, \sigma^2)$

c. None of the above

Theory Recap:

Linear Regression and Likelihood Function

1. Model Assumption:

- The model (x_i, y_i) satisfies the model:

$$y_i = f(x_i) + \varepsilon_i$$

- $f(x_i) = x_i^T\theta$: is linear prediction and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is independent Gaussian noise.

2. Likelihood Function:

- The likelihood function is the joint probability of the observed data $y = [y_1, y_2, \ldots, y_N]$ conditioned on $x = [x_1, x_2, \ldots, x_N]$ and parameter θ and σ^2.
- Since ε_i are independent:

$$p(y|x,\theta) = \prod_{i=1}^{N} p(y_i|x_i,\theta) \qquad p(y_i|x_i,\theta) = \mathcal{N}(y_i|x_i^T\theta, \sigma^2)$$

- $\mathcal{N}$ normal distribution.

3. Gaussian Distribution for y:

- $y_i | x_i, \theta \sim \mathcal{N}(x_i^T\theta, \sigma^2)$, so: $p(y|x_i,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i^T\theta)^2}{2\sigma^2}\right)$

4. Joint Likelihood: $p(y|x,\theta) = \prod_{i=1}^{N} \mathcal{N}(y_i|x_i^T\theta, \sigma^2)$

(A) $p(y|x,\theta) = \prod_{i=1}^{N} \mathcal{N}(y|x_i^T\theta x_i, \sigma^2)$ (INCORRECT)

- Incorrect because the argument $x_i^T\theta x_i$

Question 38

Let A be the matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

What is the output of the Python code

$$B = A[:, 1:2]?$$

a. None of the others

b.

$$B = \begin{bmatrix} 2 & 3 \\ 5 & 6 \\ 8 & 9 \end{bmatrix}$$

c.

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

```python
import numpy as np

A = np.array([
    [1, 2, 3],
    [4, 5, 6],
    [7, 8, 9],
])

B = A[:, 1:2]
print(B)

```

✓ 0.0s

```
[[2]
 [5]
 [8]]
```

Question 39

Importing numpy with the instruction: `import numpy as np`, indicate which of the following instructions creates the array:

$$x = \begin{bmatrix} 0 & 0.1 & 0.2 & 0.3 \end{bmatrix}$$

a. None of the others

b. `x = np.logspace(0, 0.3)`

c. `x = np.linspace(0, 0.3)`

```python
import numpy as np


x_b = np.logspace(start=0, stop=0.3)
x_c = np.linspace(start=0, stop=0.3)

print(f"x_b: {x_b[:3]}")
print(f"x_c: {x_c[:3]}")
```

✓ 0.0s

```
x_b: [1.         1.0141973  1.02859616]
x_c: [0.         0.00612245 0.0122449 ]
```

Question 40

Importing numpy with the instruction `import numpy as np`, how to define the temporary variable A such that `A = np.reshape(A, (2,2))` returns the matrix:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

a. None of the others

b. `A = np.linspace(1, 4, 4)`

c. `A = np.arange(4)`

```python
import numpy as np


A_b = np.linspace(start=1, stop=4, num=4)
A_c = np.arange(stop=4)

print(f"A_b: {A_b}")
print(f"A_c: {A_c}")

A_b_reshaped = np.reshape(A_b, (2, 2))
A_c_reshaped = np.reshape(A_c, (2, 2))

print(f"A_b_reshaped:\n {A_b_reshaped}")
print(f"A_c_reshaped:\n {A_c_reshaped}")
```

0.0s

```
A_b: [1. 2. 3. 4.]
A_c: [0 1 2 3]
A_b_reshaped:
 [[1. 2.]
 [3. 4.]]
A_c_reshaped:
 [[0 1]
 [2 3]]
```

Question 41

In $\mathcal{F}(10, 5, -10, 10)$, the number $x = 10.\bar{2}$ is approximated by:

a. $\tilde{x} = 0.10222 \cdot 10^1$

b. $\tilde{x} = 0.10222 \cdot 10^2$

c. $\tilde{x} = 0.01022 \cdot 10^3$

$\bar{x} = m \cdot 10^e$

- m : mantissa $\quad 1 \leq |m| < 10$
- e : exponent bound by emin ∧ emax

$x = 10,\bar{2}$

$x = 10,2222 \dots \cdot 10^0$

$x = 1,0222|2 \dots \cdot 10^1$

$p = 5$

$\tilde{x} = 1,0222 \cdot 10^1$

$\tilde{x} = 0,10222 \cdot 10^2$

$F(10, 5, -10, 10)$

$\beta = 10$ (base)

$p = 5$ (precision)

$e_{min} = -10$

$e_{max} = 10$

Question 42

The 2-norm of a matrix $A = (a_{i,j})$ with shape $m \times n$ is defined as $\rho(A)$ (where $\rho(A)$ is the spectral radius of A):

a. $\|A\|_2 = \sqrt{\rho(A)}$

b. $\|A\|_2 = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j}^2}$

c. $\|A\|_2 = \sqrt{\rho(A^T A)}$

THEORY RECAP:

def.: 2-norm of A : $\|A\|_2 = \sqrt{\rho(A^TA)}$ • $\rho(A^TA)$: spectral radius of A^TA which is the largest eigenvalue of A^TA.

1. SPECTRAL RADIUS: $\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$

2. DERIVATION OF 2-norm : $\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2$

   - This measures the max stretching effect of A on any vector x with unit norm.

3. CONNECTION TO A^TA :

   - $\|A\|_2^2$ corresponds to the largest eigenvalue of A^TA
   - Thus : $\|A\|_2 = \sqrt{\lambda_{max}(A^TA)}$

(A) $\|A\|_2 = \sqrt{\rho(A)}$ (INCORRECT)

(B) $\|A\|_2 = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$ (INCORRECT) - This is Frobenius norm

Question 43

If $A \in \mathbb{R}^{n \times n}$ is symmetric and positive definite, then $\ker(A)$ represents the kernel of A:

a. $\ker(A) = A^T$

b. None of the above

c. $\ker(A) = \emptyset$

THEORY RECAP:

DEFINITE MATRICES

1. DEFINITION:
   - $A \in \mathbb{R}^{n \times n}$ is POSITIVE DEFINITE if: $x^T A x > 0 \quad \forall x \neq 0, \ x \in \mathbb{R}^n$
   - POSITIVE DEFINITE MATRICES ARE ALWAYS SYMMETRIC

2. KEY PROPERTIES:
   - All eigenvalue of A are positives.
   - Since $\forall \lambda_i \in \mathbb{R} : |\lambda_i| > 0 \hookrightarrow \det(A) \neq 0 \Rightarrow \exists A^{-1}$

     (Since all eigenvalue are nonzero, the determinant A is nonzero, making A invertible)
   - The kernel (null space) of A contains ONLY THE ZERO VECTOR: $\text{Ker}(A) = \{0\}$

3. KERNEL ($\text{Ker}(A)$):
   - The kernel of A is the set of all vector x such that: $Ax = 0$
   - If A invertible the only solution to $Ax = 0$ is $x = 0$, meaning: $\text{Ker}(A) = \{0\}$

(A) $\text{Ker}(A) = A^T$ (INCORRECT)

- A is not related to A^T.

  In fact, since A is symmetric

  ($A^T = A$) this statement does not work.

(C) $\text{Ker}(A) = \emptyset$ (INCORRECT)

The kernel of A is not empty

It contains the ZERO VECTOR: $\text{Ker}(A) = \{0\}$

Question 44

If $x = (3, -1)$, $y = (\frac{1}{3}, 1)$, then:

a. x and y are parallel.

b. x and y are orthonormal.

c. x and y are orthogonal.

1. CHECK FOR PARALELISM:

$$\forall x, y \in \mathbb{R}^n \wedge k \in \mathbb{R} \hookrightarrow x \parallel y \Leftrightarrow x = ky$$

$$\begin{pmatrix} 3 \\ -1 \end{pmatrix} = k \begin{pmatrix} 1/3 \\ 1 \end{pmatrix} \Rightarrow \begin{cases} 3 = \frac{k}{3} \\ -1 = k \end{cases} \Rightarrow \begin{cases} k = 9 \\ k = -1 \end{cases} \rightarrow \nexists\, k \in \mathbb{R}$$ (NOT PARALEL)

2. CHECK FOR ORTHOGONALITY:

$$\forall \vec{x}, \vec{y} \in \mathbb{R}^n \hookrightarrow \vec{x} \perp \vec{y} \Leftrightarrow \vec{x} \cdot \vec{y} = 0$$

$$\vec{x} \cdot \vec{y} = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1 \end{pmatrix} = 3 \cdot \frac{1}{3} + 1(-1) = 0$$ (ORTHOGONAL)

3. CHECK FOR ORTHONORMALITY:

i) $\vec{x} \perp \vec{y}$ - ORTHOGONAL

ii) $\|\vec{x}\| = 1 \wedge \|\vec{y}\| = 1$ - unit lenght

$\square$ $\|\vec{x}\| = \sqrt{3^2 + (-1)^2} = \sqrt{10} \neq 1$

$\|\vec{y}\| = \sqrt{\left(\frac{1}{3}\right)^2 + 1^2} = \frac{\sqrt{10}}{3} \neq 1$ (NOT ORTHONORMAL)

Question 45

If V is a vector space with $\dim(V) = n$, $U \subseteq V$ is a subspace with $\dim(U) =$ k, and $\Pi_U : V \to U$ is a projection. Then:

$\lrcorner\, k = m$

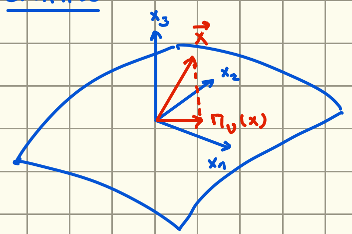
a. $\Pi_U(x) \in \mathbb{R}^n$. ← answer given by the professor in the exam answer sheet.

b. $\Pi_U(x) \in \mathbb{R}^m$. ← answer that I believe be "correct"

c. $\Pi_U(x) \in \mathbb{R}^{n-m}$.

1. V is vector space with dimension $\dim(V) = n$
2. $U \subseteq V$ is a sub-space with dimension $\dim(U) = k$
3. $\Pi_U : V \to U$ is the projection of $x \in V$ onto the subspace U

EXAMPLE:



$V : \{x_1, x_2, x_3\} \Rightarrow \dim(V) = n = 3$

$U : \{x_1, x_2\} \Rightarrow \dim(U) = m = 2$

$\vec{x}_V = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longrightarrow \vec{x}_U = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$

The projection does not change the numeric dimension of x (still in $\mathbb{R}^n$) but the result $\Pi_U(x)$ is conceptually confined to the subspace U, which has a lower dimension k.

- The NUMERIC REPRESENTATION the answer is $\Pi_U(x) \in \mathbb{R}^n$
- The GEOMETRIC REPRESENTATION the answer is $\Pi_U(x) \in \mathbb{R}^m$

Question 46

If

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

then:

a. $K_2(A) = \frac{1}{2}$.

b. $K_2(A) = 1$.

c. $K_2(A) = 4$.

1- $K_2(A) = \frac{\sigma_{max}}{\sigma_{min}}$ $\quad \sigma : \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$

$\sigma : \{1, 1, 1, 1\}$

$\sigma_{max} = 1 \wedge \sigma_{min} = 1 \implies K_2(A) = 1$

Question 47

If $A = U\Sigma V^T$ is the SVD decomposition of $A \in \mathbb{R}^{m \times n}$, then the rank-$k$ approximation of A is:

a. $\hat{A}(k) = \sum_{i=1}^{k} A_i$, where $A_i = u_i v_i^T$ is a dyade.

b. $\hat{A}(k) = \sum_{i=1}^{k} \sigma_i A_i$, where $A_i = u_i v_i^T$ is a dyade.

c. $\hat{A}(k) = \sum_{i=1}^{n} \sigma_i A_i$, where $A_i = u_i v_i^T$ is a dyade.

rank-k approximation uses the FIRST K SINGULAR VALUES and the associated singular vectors, as described by:

$$\hat{A}(k) = \sum_{i=1}^{k} \sigma_i (u_i v_i^T)$$

- $u_i v_i^T$: dyad

each dyad is scaled by corresponding singular value σ_i.

(A) $\hat{A}(k) = \sum_{i=1}^{k} A_i$, $A_i = u_i v_i^T$ - dyad (INCORRECT)

because the singular values σ_i are missing in the Σ.

(C) $\hat{A}(k) = \sum_{i=1}^{n} \sigma_i A_i$, $A_i = u_i v_i^T$ - dyad (INCORRECT)

because it includes all n singular values, not just the first k.

Question 48

A solution of $\min_x \|Ax-b\|_2^2$, where A is an $m \times n$ matrix, $m \geq n$, rank$(A) = k < n$, can be computed as:

a. $A^+x = b$.

b. $AA^Tx = A^Tb$.

c. $x = A^+b$.

Theory Recap:

1. Moore-Penrose Pseudoinverse (A^+):
   - If A is not of full rank ($k < n$) the solution to $Ax = b$ may not exist or might not be unique
   - The least-squares solution is given by: $x = A^+b$ • A^+: pseudoinverse of A

2. Normal Equation:
   - The minimization problem can also be solved by forming the normal equation: $A^TAx = A^Tb$

$$\square \quad Ax = b$$
$$Ax - b = 0$$
$$\|Ax-b\|_2^2 = 0$$
$$\vec{\nabla}_x \|Ax-b\|_2^2 = 2A^T(Ax-b) = 0 \Rightarrow A^TAx - A^Tb = 0 \Rightarrow A^TAx = A^Tb \quad \blacksquare$$

3. Relationship between A^+ and the Normal Equation:
   - A^+ satisfies: $A^+ = (A^TA)^{-1}A^T$

   assuming A^TA is invertible (which it is when A has full column rank).

(A) $A^+x = b$ (INCORRECT)

The equation should be $x = A^+b$

(B) $AA^Tx = A^Tb$ (INCORRECT)

Question 49

If $f:\mathbb{R}^3\to\mathbb{R}$, $f(x_1,x_2,x_3)=\sin x_1-\sin x_2\cos x_3+x_3^2$, then $\nabla f(\frac{\pi}{2},\frac{\pi}{2},\pi)$ equals to:

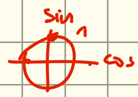
a. $(0,0,\pi)$.

b. None of the above.

c. $(0,0,0)$.

$f(x_1,x_2,x_3)=\sin x_1-\sin x_2\cdot\cos x_3+x_3^2$, $\nabla f(\frac{\pi}{2},\frac{\pi}{2},\pi)=?$

$\vec{\nabla}_x f(x_1,x_2,x_3)=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\frac{\partial f}{\partial x_3}\right)$



$$\begin{cases}\frac{\partial f}{\partial x_1}=\cos x_1\\ \frac{\partial f}{\partial x_2}=-\cos x_2\cdot\cos x_3\\ \frac{\partial f}{\partial x_3}=\sin x_2\cdot\sin x_3+2x_3\end{cases}\Rightarrow\begin{array}{l}\left.\frac{\partial f}{\partial x_1}\right|_{(\frac{\pi}{2},\frac{\pi}{2},\pi)}=\cos\frac{\pi}{2}=0\\ \left.\frac{\partial f}{\partial x_2}\right|_{(\frac{\pi}{2},\frac{\pi}{2},\pi)}=-\cancel{\cos\frac{\pi}{2}}^{=0}\cdot\cos\pi=0\\ \left.\frac{\partial f}{\partial x_3}\right|_{(\frac{\pi}{2},\frac{\pi}{2},\pi)}=\sin\frac{\pi}{2}\cdot\cancel{\sin\pi}^{=0}+2\pi=2\pi\end{array}\Rightarrow(0,0,2\pi)$$

Question 50

If $f:\mathbb{R}^2\to\mathbb{R}, f(x_1,x_2)=x_1+x_1x_2, g:\mathbb{R}^2\to\mathbb{R}^2, g(t_1,t_2)=(t_1,t_2)$, then, if $h(t_1,t_2)=f(g(t_1,t_2))$,

a. $\nabla h(x_1,x_2)=(2+2x_1,2x_1)$.

b. $\nabla h(x_1,x_2)=(1+2x_1^2,2x_1)$.

c. None of the above.

$f:\mathbb{R}^2\to\mathbb{R}$, $f(x_1,x_2)=x_1+x_1\cdot x_2$

$g:\mathbb{R}^2\to\mathbb{R}^2$, $g(t_1,t_2)=(t_1,t_2)$

if $h(t_1,t_2)=f(g(t_1,t_2))$

$h(t_1,t_2)=f(t_1,t_2)=t_1+t_1\cdot t_2$

$\nabla h(x_1,x_2)=\left(\frac{\partial h}{\partial x_1},\frac{\partial h}{\partial x_2}\right)$

$$\begin{cases}\frac{\partial h}{\partial x_1}=\frac{\partial(x_1+x_1\cdot x_2)}{\partial x_1}=1+x_2\\ \frac{\partial h}{\partial x_2}=\frac{\partial(x_1+x_1\cdot x_2)}{\partial x_2}=x_1\end{cases}\Rightarrow\vec{\nabla}h(x_1,x_2)=(1+x_2,x_1)$$

Question 51

$f : \mathbb{R}^n \to \mathbb{R}$ is differentiable if:

a. $\frac{\partial f}{\partial x_i}$ exists and is continuous for any $i = 1, \dots, n$.

b. $\nabla f(x)$ exists.

c. $\frac{\partial f}{\partial x_i}$ exists for any $i = 1, \dots, n$.

DIFFERENTIABILITY:

def.: $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $x_0 \in \mathbb{R}^n$ if $\exists$ a linear map (the gradient) L:

$$\lim_{h \to 0} \frac{|f(x_0+h) - f(x_0) - L(h)|}{\|h\|} = 0$$

- For $f : \mathbb{R}^n \to \mathbb{R}$ the linear map $L(h)$ can be expressed as: $L(h) = \vec{\nabla} f(x_0) \cdot h$
- Where $\vec{\nabla} f(x_0)$ is gradient of f at x_0

This means that the function can be well-approximated locally by a linear map (gradient)

CONTINUITY

def.: $f : \mathbb{R}^n \to \mathbb{R}$ is CONTINUOS AT A POINT $x_0 \in \mathbb{R}^n$ if:

$$\forall \varepsilon > 0 \ \exists \delta = \delta(\varepsilon) > 0 : \forall x \in \mathbb{R} \hookrightarrow x \in \mathring{U}_\delta(x_0) \Rightarrow f(x) \in U_\varepsilon(f(x_0))$$

B $\exists \vec{\nabla} f(x)$ (INCORRECT)

This is a NECESSARY condition, but NOT SUFFICIENT.
The existence of the gradient $(\vec{\nabla} f(x))$ does not garantee differentiability unless the linear approximation condition holds.

C $\forall i = \overline{1,n} \ \exists \frac{\partial f}{\partial x_i}$ (INCORRECT)

The existence of partial derivates is also NECESSARY but NOT SUFFICIENT for differentiability. For example, a function can have partial derivates that exists everywhere but is not differentiable (e.g. $f(x,y) = |x| + |y|$)

A $\forall i = \overline{1,n} \ \exists \frac{\partial f}{\partial x_i} \wedge \forall \varepsilon > 0 \ \exists \delta = \delta(\varepsilon) > 0 : \forall x \in \mathbb{R} \ \exists x \in \mathring{U}_\delta(x_0) \hookrightarrow f(x) \in U_\varepsilon(f(x_0))$ (CORRECT)

- This condition is SUFFICIENT for differentiability but NOT NECESSARY
- If the partial derivates are continuos, then f is differentiable. However, differentiability can still hold even if the partial derivates are not continuous.

Question 52

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 9x_2^2 - x_1$, $\nabla f(x_1, x_2) = (9x_2^2 - 1, 18x_1x_2)$. Then which of the following is a stationary point for f?

a. $(0, 0)$.

b. None of the above.

c. $(0, 3)$.

$f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 9x_2^2 - x_1$

$\vec{\nabla} f(x_1, x_2) = (9x_2^2 - 1, 18x_1x_2)$

Stationary points of f?

1) $\vec{\nabla} F(x_1, x_2) = \begin{pmatrix} 9x_2^2 - 1 \\ 18x_1x_2 \end{pmatrix}$

stationary point $\Rightarrow \vec{\nabla} f = \vec{0} \Rightarrow \begin{cases} 9x_2^2 - 1 = 0 \\ 18x_1 . x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_2 = \frac{1}{3} \vee x_2 = \frac{-1}{3} \\ 18x_1x_2 = 0 \Rightarrow x_1 = 0 \vee x_2 = 0 \end{cases}$

stationarie points : $\begin{cases} (0, 1/3) \\ (0, -1/3) \end{cases}$

Question 53

If $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$ and x^* is a minimum point of f, then:

a. $\nabla f(x^*) = 0$.

b. $\nabla f(x^*) \geq 0$.

c. $\nabla f(x^*)$ is positive definite.

THEORY RECAP:

1. GRADIENT AT A MINIMUM POINT:
   - If $F(x)$ is differentiable $\wedge$ x_0 – min point : $\vec{\nabla} F(x_0) = \vec{0}$
   - This is a necessary condition for : local and global minimum point.

2. SECOND-ORDER CONDITION (HESSIAN):
   - To determine x_0 is a minimum $\vec{\nabla}^2 F(x_0) > 0$

Ⓐ $\vec{\nabla} F(x_0) = 0$: NECESSARY condition for a minimum

Ⓑ $\vec{\nabla} F(x_0) \geqslant 0$: The gradient must be equal zero, not just be non-negative

Ⓒ $\vec{\nabla} F(x_0)$ is positive definite : This refers to the Hessian, not the gradient.

Question 54

Given two random variables X and Y, then the marginal probability on Y is defined as:

a. $P(Y = y)$.

b. $P(X = x, Y = y)$.

c. $P(X = x)$.

THEORY RECAP:

1. JOINT PROBABILITY:

- The probability of two events X and Y happening together is given by: $P(X=x, Y=y)$

2. MARGINAL PROBABILITY:

- The marginal probability of $Y=y$ is obtained by summing (or integrating, in the case of continuous variables) the joint probability over all possible value of X:

$$P(Y=y) = \sum_x P(X=x, Y=y) \quad \text{(discrete case)}$$

or

$$P(Y=y) = \int P(X=x, Y=y)\, dx \quad \text{(continuous case)}$$

- This gives the probability of $Y=y$ irrespective of the value of X.

(A) $P(Y=y)$ definition of the marginal probability of Y (CORRECT)

(B) $P(X=x, Y=y)$ this is the joint probability, not the marginal probability (INCORRECT)

(C) $P(X=x)$ this is the marginal probability of X, not Y. (INCORRECT)

Question 55

The mean and standard deviation of a standard normal distribution are, respectively:

a. $(0, 1)$.

b. $(1, 1)$.

c. $(1, 0)$.

THEORY RECAP:

The STANDARD NORMAL DISTRIBUTION is a specific type of normal distribution that has the following properties:

1. MEAN: $\mu = 0$ (central value around which the data is distributed)
2. STANDARD DEVIATION: $\sigma = 1$ (measures the spread of the distribution. For ND the spread is fixed $= 1$).

The standard normal distribution has the probability density function: $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \ x \in \mathbb{R}$

☐ Mean (μ)

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} x\, p(x)\, dx = \int_{-\infty}^{\infty} \frac{x \cdot e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\, dx$$

→ odd function: $\begin{cases} x \text{ (odd)} \\ e^{-\frac{x^2}{2}} \text{ (even)} \end{cases} \Rightarrow \text{(odd)}$

- integral of an odd function over a symmetric interval is 0.

$\mu = 0$ ▨

Variance (σ^2) and Standard Deviation (σ)

$$\sigma^2 = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 . p(x) . dx \qquad (\text{i.e. } \mu = 0)$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-0)^2 p(x) . dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 . p(x) . dx \quad , \quad p(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 . \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} . dx$$

$$\rightarrow I = \int_{-\infty}^{\infty} x^2 . e^{-\frac{x^2}{2}} . dx$$ using property of Gaussian Integrals

$$\int_{-\infty}^{\infty} e^{-ax^2} . dx = \sqrt{\frac{\pi}{a}} \quad (a > 0)$$

after normalization:

$$\sigma^2 = 1 \rightarrow \sigma = 1$$

$$\Rightarrow I = \sqrt{2\pi} . 1 = \sqrt{2\pi}$$

Question 56

Given a continuous random variable $X : \Omega \to T$, with $T = [0, 1]$, and $p(x) = 3x^2$ its PDF, then:

a. $\mathbb{E}[X] = 2$.

b. $\mathbb{E}[X] = 3$.

c. $\mathbb{E}[X] = \frac{3}{4}$.

$$\mathbb{E}[X] = \int_T x . p(x) . dx \qquad \begin{cases} T = [0;1] \\ p(x) = 3x^2 \end{cases}$$

1. Verify that $p(x)$ is a valid PDF:
   - Must satisfy two conditions:
     - i) $p(x) \geq 0 \quad \forall x \in T$.
     - ii) $\int_T p(x) . dx = 1$

i) $3x^2 \geq 0 \quad \forall x \in T$ (TRUE)

ii) $\int_0^1 3x^2 . dx = x^3 \Big|_0^1 = 1$ (TRUE)

$\Rightarrow$ Valid PDF

2. Compute $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \int_0^1 x \cdot 3x^2 . dx = \frac{3}{4} x^4 \Big|_0^1 = \frac{3}{4}$$

Question 57

If X, Y are random variables, the correlation between X and Y is computed as:

a. $\text{Corr}(X, Y) = \frac{\text{Cov}(X,Y)}{V_x(X)V_y(Y)}$.

b. $\text{Corr}(X, Y) = \frac{V_x(X)V_y(Y)}{\text{Cov}(X,Y)}$.

c. $\text{Corr}(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{V_x(X)V_y(Y)}}$.

1. COVARIANCE ($Cov(X,Y)$):
   - Measures the linear relationship between X and Y
   - $Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$

2. VARIANCE ($Var(X,Y)$):
   - Measures the spread of X and Y around their respective means.
   - $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

3. CORRELATION ($Corr(X,Y)$):
   - Normalizes the covariance by dividing it by the product of the standard deviation $\sqrt{Var(X)}$ and $\sqrt{Var(Y)}$
   - $Corr(X,Y) \in [-1; 1]$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

Question 58

Given two random variables X and Y, Bayes Theorem implies that

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

where:

a. $p(x)$ is called likelihood on x.

b. $p(x)$ is called prior distribution on x.

c. $p(x)$ is called posterior distribution on x.

$$\overbrace{p(x|y)}^{=\text{posterior}} = \frac{\overbrace{p(y|x)}^{=\text{likelihood}} \cdot \overbrace{p(x)}^{=\text{prior}}}{\underbrace{p(y)}_{=\text{evidence}}}$$

Question 59

In Maximum Likelihood Estimation, the prediction function is:

a. A probabilistic function.

b. None of the above.

c. A deterministic function.

DETERMINISTIC OR PROBABILISTIC:

- The MLE the process itself involves maximizing the likelihood of observing the data under a probabilistic model.
- Once the parameters are estimated using MLE, the prediction function is derivated from the model. The nature of the prediction function (whether deterministic or probabilistic) depends on the type of model being used.

1. PROBABILISTIC MODEL (logistic regression, Bayesian models)
   - The prediction function often outputs probabilities (e.g. the probability of a class in classification)
   - Thus, the PREDICTION FUNCTION IS PROBABILISTIC, as it provides a distribution or likelihood rather than a single, fixed output.

2. DETERMINISTIC MODEL (e.g. linear regression)
   - The prediction function gives a single, specific output for a given input after estimating the parameters.
   - In this case, the PREDICTION FUNCTION IS DETERMINISTIC.

CONCLUSION:

- MLE does not enforces wheter the PF is deterministic or probabilistic (depends on the type of model being used)
- In general sense MLE, the PF is PROBABILISTIC because MLE is commonly associated with probabilistic models.

Question 60

Suppose a set of data $(x_i, y_i), i = 1, \dots, N, y_i = f(x_i) + \epsilon_i$ where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. In linear regression, the predictor function f is chosen as:

a. $f(x, \theta) = \theta^T x$.

b. None of the above.

c. $f(x, \theta) = \frac{\theta^T}{x}$.

THEORY RECAP:

Linear Regression:

- $y_i = f(x_i) + \varepsilon_i \quad , \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$
- The predictor function $f(x, \theta)$ represents the relationship between $(\underset{\text{(input)}}{x}, \underset{\text{(output)}}{y})$

$$f(x, \theta) = \theta^T x$$

- x : input vector
- θ : vector of model parameters

$$\begin{cases} \text{Model definition (LR):} & y = \theta^T x + \varepsilon \quad , \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \\ \text{Prediction } (\overset{\text{noise}}{\varepsilon} = 0): & \hat{y} = \theta^T x \end{cases}$$

Question 61

Given a matrix $A \in \mathbb{R}^{m \times n}, m > n$, with $r = \text{rank}(A)$, then:

a. None of the above.

b. It is always possible to write A as $U \Sigma V^T$, where $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal, $U \in \mathbb{R}^{m \times n}, V \in \mathbb{R}^{n \times n}$ are orthogonals.

c. It is always possible to write A as $U \Sigma V^T$, where $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal, $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$ are orthogonals.

$$A \in \mathbb{R}^{m \times n}, \quad m > n \quad , \quad r = \text{rank}(A)$$

$$A = U \Sigma V^T$$

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V^T_{n \times n} \qquad \begin{cases} U \in \mathbb{R}^{m \times m} \\ \Sigma \in \mathbb{R}^{m \times n} \\ V^T \in \mathbb{R}^{n \times n} \end{cases}$$

$$\underbrace{\begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix}}_{m \times n} = \underbrace{\begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{pmatrix}}_{n \times m} \underbrace{\begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{pmatrix}}_{m \times n} \underbrace{\begin{pmatrix} - & v_1 & - \\ - & v_2 & - \\ & \vdots & \\ - & v_n & - \end{pmatrix}}_{n \times m}$$

Question 62

If A be a matrix $m \times n$ $(m > n)$ $\text{rg}(A) = k$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are the singular values of A, which of the following sentences is correct ($k_2(A)$ is the condition number of A in norm 2):

a. None of the other alternatives

b. $k_2(A) = \frac{\sigma_n}{\sigma_1}$

c. $k_2(A) = \frac{\sigma_1}{\sigma_n}$

$A \in \mathbb{R}^{m \times n}$ $(m > n)$

$\text{rank}(A) = k$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$

$K_2(A) = \frac{\|A\|_2}{\sigma_{min}} = \frac{\sigma_{max}}{\sigma_{min}} = \frac{\sigma_1}{\sigma_k}$

rank(A) = k

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{k-1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma_k & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Question 63

If A is a matrix $m \times n$ $(m > n)$ and $A = U\Sigma V^T$ is the SVD of A, $A_p = \sum_{i=1}^{p} u_i \sigma_i v_i^T$, which of the following statements is correct?

a. $\|A_p\|_2 = \sigma_p$

b. A_p has rank p

c. None of the other alternatives

$A = U \Sigma V^T$ - Singular Value Decomposition

$A_p = \sum_{i=1}^{p} u_i \sigma_i v_i^T$ - rank-p approximation of A

(A) $\|A_p\|_2 = \sigma_p$ (INCORRECT)

$\|A_p\|_2 = \sigma_1$

(B) A_p has rank p (CORRECT)

$\text{rank}(A_p) = p$

$\Rightarrow K_2(A_p) = \frac{\|A_p\|_2}{\sigma_{min}} = \frac{\sigma_1}{\sigma_p}$

Question 64

Let $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = \exp(x_1 x_2) - \sin(x_3)\cos(x_2)$, then $\nabla f(0,0,0) =$

a. None of the other alternatives

b. $(0, 0, 1)$

c. $(0, 0, 0)$

$f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = e^{x_1 x_2} - \sin x_3 . \cos x_2$

$\vec{\nabla} f(0,0,0) = ?$

$\vec{\nabla} f(x_1, x_2, x_3) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \right)$

$$\begin{cases} \frac{\partial f}{\partial x_1} = x_2 e^{x_1 x_2} \\ \frac{\partial f}{\partial x_2} = x_1 . e^{x_1 x_2} + \sin x_3 . \sin x_2 \\ \frac{\partial f}{\partial x_3} = -\cos x_3 . \cos x_2 \end{cases} \Rightarrow \begin{cases} \left.\frac{\partial f}{\partial x_1}\right|_{=(0,0,0)} = 0.e^0 = 0 \\ \left.\frac{\partial f}{\partial x_2}\right|_{=(0,0,0)} = 0.e^0 + \sin 0 . \sin 0 = 0 \\ \left.\frac{\partial f}{\partial x_3}\right|_{=(0,0,0)} = -\cos 0 . \cos 0 = -1 \end{cases} \Rightarrow (0, 0, -1)$$

Question 65

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 + 2x_1^2$, $g : \mathbb{R} \to \mathbb{R}^2$, $g(t) = (t, t^2)$, and if $h(t) = f(g(t))$, then

a. None of the other alternatives

b. $h'(t) = 8t^2 + t$

c. $h'(t) = 4t^3 + 1$

$f : \mathbb{R}^2 \to \mathbb{R}$

$f(x_1, x_2) = x_1 + 2x_1^2$

$g : \mathbb{R} \to \mathbb{R}^2$

$g(t) = (t, t^2)$

$h(t) = f(g(t))$

$h(t) = t + 2t^2$

$h'(t) = 1 + 4t$

Question 66

If $f(x) = \|Ax - b\|_2^2$ with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, x \in \mathbb{R}^n$, then $\nabla f(0) =$

a. None of the other alternatives

b. 0

c. $-2A^Tb$

THEORY RECAP:

$$\begin{cases} A \in \mathbb{R}^{m\times n} \\ b \in \mathbb{R}^{m\times 1} \\ x \in \mathbb{R}^{n\times 1} \end{cases}$$

1. $\|A\|_2^2 = A^TA$
2. $\nabla_\theta(\theta^TA\theta) = 2A\theta$ (if A-symmetric: $A^T = A$)
3. $\nabla_\theta(\theta^Tb) = b$
4. $\nabla_b(\theta^Tb) = \theta$
5. $\nabla_\theta(A) = 0$ (A-independent of θ)

$$f(x) = \|Ax-b\|_2^2 = (Ax-b)^T(Ax-b)$$
$$= (x^TA^T - b^T)(Ax-b)$$
$$= x^TA^TAx - \underbrace{x^TA^Tb}_{=(*)} - \underbrace{b^TAx}_{=(**)} - b^Tb$$

$$\begin{cases} (*): x^TA^Tb = x^T_{1\times n} A^T_{n\times m} b_{m\times 1} = \tilde{x}_{1\times 1} \\ (**): b^TAx = b^T_{1\times m} A_{m\times n} x_{n\times 1} = \tilde{x}_{1\times 1} \end{cases} \Rightarrow x^TA^Tb = b^TAx$$

$$f(x) = x^TA^TAx - 2b^TAx - b^Tb$$

$$\nabla f = \underbrace{\nabla(x^TA^TAx)}_{= 2A^TAx} - 2\underbrace{\nabla(b^TAx)}_{= A^Tb} - \underbrace{\nabla(b^Tb)}_{= 0}$$

$$\nabla f = 2A^TAx - 2A^Tb$$
$$\nabla f = 2A^T(Ax-b)$$
$$\nabla f(0) = -2A^Tb$$

Question 67

If $f : \mathbb{R}^n \to \mathbb{R}$, x^* is a global minimum for f, if:

a. $f(x) \le f(x^*), \forall x \in \mathbb{R}^n$

b. $f(x^*) \le f(x), \forall x \in \mathbb{R}^n$

c. $f(x) \le f(x^*), \forall x : \|x - x^*\| < \epsilon, \epsilon > 0$

THEORY RECAP:

$\perp$ 1. $f : \mathbb{R}^n \to \mathbb{R}$ continuos $\wedge$ differentiable in $\mathbb{R}^n$

2. $x_0 \in \mathbb{R}^n$

$\perp$ 1. Stationary Point : $\vec{\nabla} f(x_0) = \vec{0}$

2. Local Minimum : $\exists \varepsilon \in \mathbb{R} : f(x_0) \le f(x), \forall x \in \mathbb{R}^n : |x - x_0| < \varepsilon$

3. Strict Local Minimum : $\exists \varepsilon \in \mathbb{R} : f(x_0) < f(x), \forall x \in \mathbb{R}^n : |x - x_0| < \varepsilon$

4. Global Minimum : $f(x_0) \le f(x) \ \forall x \in \mathbb{R}^n$

5. Strict Global Minimum : $f(x_0) < f(x) \ \forall x \in \mathbb{R}^n$

Question 68

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 2 \sin x_1 + \sin x_1 \cos x_2$, which of the following is a stationary point for f?

a. $(\pi/2, 0)$

b. None of the other alternatives

c. $(\pi/2, \pi/2)$

$f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 2 \cdot \sin x_1 + \sin x_1 \cdot \cos x_2$

$$\vec{\nabla} f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = \vec{0}$$

$$\begin{cases} \frac{\partial f}{\partial x_1} = 2 \cos x_1 + \cos x_1 \cdot \cos x_2 = 0 \\ \frac{\partial f}{\partial x_2} = - \sin x_1 \cdot \sin x_2 = 0 \end{cases} \Rightarrow \begin{cases} \cos x_1 \cdot (2 + \cos x_2) = 0 & (i) \\ \sin x_1 \cdot \sin x_2 = 0 & (ii) \end{cases}$$

i) $\cos x_1 (2 + \cos x_2) = 0$

$\cos x_1 = 0 \quad \vee \quad 2 + \cos x_2 = 0$

$\Rightarrow x_1 = \frac{\pi}{2} + k_1 \pi \qquad \cos x_2 = -2$

$(k_1 \in \mathbb{Z}) \qquad \Rightarrow \nexists x_2 \in \mathbb{R}$

ii) $\sin x_1 \cdot \sin x_2 = 0$

$\sin x_1 = 0 \quad \vee \quad \sin x_2 = 0$

$x_1 = k_2 \pi \qquad x_2 = k_3 \pi$

$(k_2, k_3 \in \mathbb{Z})$

$(\pi/2, 0)$ - solution

Question 69

Gradient descent methods:

a. Always converges to a minimum of $f(x)$.

b. If α is suitably chosen, $f \in C^1$, for any x_0, always converges to a minimum of $f(x)$.

c. If α is suitably chosen, $f \in C^1$, for any x_0, always converges to a stationary point of $f(x)$.

THEORY RECAP:

GRADIENT DESCENT

- Is an optimization used to minimize $f(x)$. It iteratively updates the parameters x by moving in the direction opposite to the $\vec{\nabla} f(x)$.

$$x_{k+1} = x_k - \alpha_k \cdot \vec{\nabla} f(x_k)$$

- α_k : step size (learning rate).
- $\vec{\nabla} f(x)$: gradient of $f(x)$ at x_k.

KEY POINTS:

1. CONVERGENCE TO A STATIONARY POINT:
   - if x_0 S.P. $\Rightarrow \vec{\nabla} f(x_0) = \vec{0}$. G.D. garantees to converge to a S.P. of $f(x)$ given a suitable α_k assuming $f(x)$ is differentiable ($f \in C^1$).
2. S.P. and MINIMUM:
   - S.P. could be local minimum, local maximum or a saddle point.
   - G.D. does NOT garantee convergence to a local minimum unless additional assumptions (e.g. convexity) are made.
3. CHOICE OF α_k:
   - α_k = const and works under specific conditions, but might fail to converge in some cases.
   - Techniques like LINE SEARCH or BACKTRACKING ensures a suitable α_k for convergence.

A (INCORRECT) because G.D. can converge to a saddle point or a local maximum, depending on the starting point and the nature of $f(x)$.

B (INCORRECT) because, even with a suitable α_k, the method only garantees convergence to a stationary point, not necessarily a minimum.

Question 70

What is the output of the following Python code?

```
import numpy as np
x = np.linspace(1,10,4)
for i in range(0,4):
    print(x[i])
```

a. 1 3 5 7 (1,7,4)

b. 0 1 2 3 (0,3,4)

c. None of the others

```
import numpy as np

x = np.linspace(start=1, stop=10, num=4)

for i in range(0, 4):
    print(f"x[{i}]: {x[i]}")
```

```
x[0]: 1.0
x[1]: 4.0
x[2]: 7.0
x[3]: 10.0
```

x = np.linspace(start, stop, num)

$step = \frac{stop - start}{num - 1}$

$(1, 10, 4) \Rightarrow step = \frac{10-1}{4-1} \Rightarrow step = \frac{9}{3} \Rightarrow step = 3$

$= \{1, 4, 7, 10\}$

Question 71

Importing numpy with the instruction `import numpy as np`, how to define the temporary variable A such that `A=np.reshape(A, (2,2))` returns the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

a. `A=np.arange(4) + 1`

b. None of the others

c. `A=np.linspace(1,5,4)`

```
import numpy as np

A_a = np.arange(4) + 1
A_c = np.linspace(1, 5, 4)

print(f"A_a: \n{np.reshape(A_a, (2, 2))}")
print(f"A_c: \n{np.reshape(A_c, (2, 2))}")
```

```
A_a:
[[1 2]
 [3 4]]
A_c:
[[1.         2.33333333]
 [3.66666667 5.        ]]
```

Question 72

If Ω is the sample space, A the event space, and $T = [0, 1]$, a random variable X is:

a. A function $X : \Omega \to T$

b. A function $X : A \to T$

c. A function $X : T \to \Omega$

THEORY RECAP:

A RANDOM VARIABLE X is a measurable function that maps outcomes from the sample space Ω to a target space T (usually subsets of $\mathbb{R}$).

1. SAMPLE SPACE Ω : set of all possible outcomes of an experiment
2. TARGET SPACE T : set where the random variable maps the outcomes, often a subset $\mathbb{R}$ (e.g. $[0,1]$))
3. EVENT SPACE A : collection of subsets of Ω that are events (measurable).

$X : \Omega \to T$

($\mathbb{R}$ or intervals)

The random variable X assign a value in T to each outcome in Ω. The function must satisfy the property of MEASURABILITY, meaning that the preimage of any set in T must belong to A.

Question 73

The quantity of interest in Bayes' theorem is:

a. The posterior

b. The likelihood

c. The marginal

$$\overbrace{p(x|y)}^{=\text{posterior}} = \frac{\overbrace{p(y|x)}^{=\text{likelihood}} \cdot \overbrace{p(x)}^{=\text{prior}}}{\underbrace{p(y)}_{=\text{evidence}}}$$

Question 74

If $X : \Omega \to T$ is a discrete random variable and F_X is the PMF of X, then the distribution mean (or expectation) is defined as:

a. $\mathbb{E}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i F_X(x_i), x_i \in T$

b. $\mathbb{E}(x) = \frac{1}{N}\sum_{i=1}^{N} F_X(x_i), x_i \in T$

c. $\mathbb{E}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i, x_i \in T$

THEORY RECAP:

The EXPECTED VALUE (mean) of a discrete random variable X with values $x_1, x_2, \dots, x_N$ in the target space T is defined as:

$$\mathbb{E}(X) = \sum_{i=1}^{N} x_i \cdot P(X = x_i) \Rightarrow \mathbb{E}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \cdot F_X(x_i)$$

- $P(X = x_i)$: probability mass function (PMF) $F_X(x_i)$

Question 75

Given two random variables X and Y such that $p(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ and $p(y|x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-ax)^2}$, then the MLE reads:

a. $x^* = \arg\min_x \frac{1}{2}(y-ax)^2 + \frac{1}{2}x^2$

b. $x^* = \arg\min_x \frac{1}{2}(y-ax)^2 + x^2$

c. $x^* = \arg\min_x \frac{1}{2}(y-ax)^2$

MLE maximize the likelihood $L(x) = p(y|x)$

$$L(x) = \frac{e^{-\frac{(y-ax)^2}{2}}}{\sqrt{2\pi}} \Rightarrow \ln L(x) = -\frac{(y-ax)^2}{2} + \text{const}$$

$\arg\max_x f(x) = \arg\min_x \{-f(x)\}$

$$\Rightarrow \arg\min_x \{-\ln L(x)\} = \arg\min_x \left\{\frac{(y-ax)^2}{2}\right\}$$

Question 76

The regression matrix A of the data $(x_i, y_i), i = 1, \dots, N, x_i, y_i \in \mathbb{R}$ for a polynomial model of degree $k-1$ has elements:

a. $a_{ij} = x_i^j, i = 1, \dots, N, j = 1, \dots, k$

b. $a_{ij} = x_i^{j+1}, i = 1, \dots, N, j = 1, \dots, k$

c. $a_{ij} = x_i^{j-1}, i = 1, \dots, N, j = 1, \dots, k$

THEORY RECAP:

Polynomial Regression Model

└ degree = k−1

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_{k-1} x_i^{k-1} + \varepsilon_i$$

$$\vec{y} = A\vec{\beta} + \vec{\varepsilon} \Rightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^{k-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{k-1} \end{pmatrix}$$

- $\vec{y} \in \mathbb{R}^N$: vector of observed outputs
- $\vec{\beta} = [\beta_0\ \beta_1 \cdots \beta_{k-1}]^T$: vector of coefficients
- $A \in \mathbb{R}^{N \times k}$: regression matrix
- $A_{ij} = x_i^{j-1}$ j-column of A represents the powers of x_i starting from x_i^0

Question 77

Let A be the matrix:

$$A = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \\ 0.5 & 0.6 \end{bmatrix}$$

What is the output of the Python code:

```
B = A[-1:, :]
```

a. None of the others

b. $B = \begin{bmatrix} 0.2 \end{bmatrix}$

c. $B = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$

```python
import numpy as np

A = np.array([
    [0.1, 0.2],
    [0.3, 0.4],
    [0.5, 0.6],
])
B = A[-1:, :]
print(B)
```

✓ 0.0s

```
[[0.5 0.6]]
```

Question 78

The normalized scientific representation of 3.7856 is:

a. $3.7856 \cdot 10^0$

b. $3.7856 \cdot 10^1$

c. $0.37856 \cdot 10^1$ ← However the answer in the exam is this.

What is Scientific Normalization?

In scientific normalization:

1. A number is expressed as $a \cdot 10^b$, where:
   - a (the significand or mantissa) satisfies $1 \leq |a| < 10$ (for base 10).
   - b (the exponent) is an integer.

This ensures that:

- The first digit of a (to the left of the decimal point) is **nonzero**.
- The number is scaled by 10^b to maintain equality with the original value.

Scientific notation

48 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

This article is about a numeric notation. For the musical notation, see Scientific pitch notation.

"E notation" redirects here. For the series of preferred numbers, see E series. For the food additive codes, see E number.

Scientific notation is a way of expressing numbers that are too large or too small to be conveniently written in decimal form, since to do so would require writing out an inconveniently long string of digits. It may be referred to as **scientific form** or **standard index form**, or **standard form** in the United Kingdom. This base ten notation is commonly used by scientists, mathematicians, and engineers, in part because it can simplify certain arithmetic operations. On scientific calculators, it is usually known as "SCI" display mode.

In scientific notation, nonzero numbers are written in the form

$m \times 10^n$

or m times ten raised to the power of n, where n is an integer, and the coefficient m is a nonzero real number (usually between 1 and 10 in absolute value, and nearly always written as a terminating decimal). The integer n is called the exponent and the real number m is called the significand or mantissa.[1] The term "mantissa" can be ambiguous where logarithms are involved, because it is also the traditional name of the fractional part of the common logarithm. If the number is negative then a minus sign precedes m, as in ordinary decimal notation. In normalized notation, the exponent is chosen so that the absolute value (modulus) of the significand m is at least 1 but less than 10.

Decimal notation	Scientific notation
2	2×10^0
300	3×10^2
4 321.768	$4.321\,768 \times 10^3$
−53 000	-5.3×10^4
6 720 000 000	6.72×10^9
0.2	2×10^{-1}
987	9.87×10^2
0.000 000 007 51	7.51×10^{-9}

Decimal floating point is a computer arithmetic system closely related to scientific notation.

Question 79

In the floating point system $\mathcal{F}(10, 5, -4, 4)$, which is the preceding number of $x = 10^{-4}$:

a. $y = 9 \cdot 10^{-5}$

b. None of the other alternatives

c. $y = 10^{-5}$

1. Base: $b = 10$
   Precision: $p = 5$
   Exponent Range: $[e_{min}, e_{max}] = [-4, 4]$
   General form of a floating-point number
   $$x = m \cdot 10^{e}$$
   - m (mantissa) : $1 \leq |m| < 10$
   - e (exponent) : within $[-4, 4]$
   - m is rounded to $p = 5$ digits

2. DEFINING THE GIVEN NUMBER

   $x = 10^{-4}$

   Expressing x in normalized form:
   - mantissa $m = 1,0000$ (5 significant digits)
   - exponent $e = -4$

   $\Rightarrow x = 1,0000 \cdot 10^{-4}$

3. Finding the precending Number:
   - To find the precending number, decrement the mantissa by the smallest possible step in base 10 with precision $p = 5$

   3.1. $\Delta m = 10^{-(p-1)}$
   $\Delta m = 10^{-(5-1)}$
   $\Delta m = 10^{-4}$

   3.2. $m_{new} = m - \Delta m$
   $m_{new} = 1,0000 - 0,0001$
   $m_{new} = 0,9999$

$\Rightarrow y = m_{new} \cdot 10^{e}$

$y = 0,9999 \cdot 10^{-4}$

Question 80

Given the matrix:

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -8 \end{bmatrix}$$

Then:

a. $\|A\|_2 = -8$

b. None of the other alternatives

c. $\|A\|_2 = \sqrt{105}$

$\|A\|_2 = \sqrt{\rho(A^TA)}$

$\rho(A^TA)$: spectral radius of A^TA
(which is the largest eigenvalue of A^TA

OBS : For diagonal matrices, the 2-norm simplifies directly to the largest absolute value of its diagonal entries

$\|A\|_2 = \max\{|4|, |5|, |-8|\}$
$\|A\|_2 = 8$

Question 81

The matrix:

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$$

is:

a. Symmetric and positive definite

b. None of the other alternatives

c. Symmetric and positive semidefinite

$A = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix} = A^T \Rightarrow$ A-symmetric

A matrix is positive $\begin{cases} \text{definite if } \lambda_i > 0 \ \forall i \in N \\ \text{semidefinite if } \lambda_i \geq 0 \ \forall i \in N \end{cases}$

$\det(A - \lambda E) = 0$

$\begin{vmatrix} 2-\lambda & 4 \\ 4 & 8-\lambda \end{vmatrix} = 0$

$(2-\lambda)(8-\lambda) - 16 = 0$

$16 - 2\lambda - 8\lambda + \lambda^2 - 16 = 0$

$\lambda^2 - 10\lambda = 0$

$\lambda(\lambda - 10) = 0$

$\lambda_1 = 0 \lor \lambda_2 = 10 \Rightarrow$ SEMIDEFINITE

$\lambda_1 \geq 0 \quad \lambda_2 \geq 0$

Question 82

The accuracy is:

a. The number of correct significant digits in approximating some quantity.

b. The number of digits with which a number is expressed.

c. None of the above.

THEORY RECAP:

def.: Accuracy - in numerical analysis, accuracy refers to how close a computed or measured value is to the true value. It tipically accounts for the number of correct significant digits in approximating some quantity, meaning it reflects the quality of the approximation.

Question 83

Given two random variables X and Y, Bayes Theorem implies that

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

where:

a. $p(x|y)$ is called prior distribution on x.

b. $p(x|y)$ is called posterior distribution on y.

c. $p(x|y)$ is called likelihood on y.

$$\underbrace{p(x|y)}_{=\text{posterior}} = \frac{\underbrace{p(y|x)}_{=\text{likelihood}} \cdot \underbrace{p(x)}_{=\text{prior}}}{\underbrace{p(y)}_{=\text{evidence}}}$$

Question 85

If $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1 e^{x_2}, g : \mathbb{R} \to \mathbb{R}^2, g(t) = (e^t, t)$, then, if $h(t) = f(g(t))$:

$$h'(t) = ?$$

a. $h'(t) = te^t$.

b. $h'(t) = 2e^{2t}$.

c. $h'(t) = e^{2t}(t+1)$. ← However the exam's answer shows this one.

$f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 . e^{x_2}$

$g : \mathbb{R} \to \mathbb{R}^2$, $g(t) = (e^t, t)$

$h(t) = f(g(t))$, $h'(t) = ?$

$h(t) = e^t . e^t \Rightarrow h(t) = e^{2t} \Rightarrow h'(t) = 2e^{2t}$

Question 86

If $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1^2 + x_1 x_2, g : \mathbb{R} \to \mathbb{R}^2, g(t) = (\sin(t), \cos(t))$, then, if $h(t) = f(g(t))$:

$$h'(t) = ?$$

a. $h'(t) = \sin(2t) - \sin^2(t)$.

b. $h'(t) = \sin(t) - \sin^2(2t)$.

c. $h'(t) = \sin(t)\cos(t) - \sin^2(t)$.

NO ANSWER AVAIABLE !

$f(x_1, x_2) = x_1^2 + x_1 . x_2$

$g(t) = (\sin t, \cos t)$

$h(t) = f(g(t))$

$h(t) = \sin^2 t + \sin t . \cos t$

$h'(t) = \underbrace{2 \sin t \cos t}_{= \sin 2t} + \underbrace{\cos^2 t - \sin^2 t}_{= \cos 2t}$

$h'(t) = \sin 2t + \cos 2t$

Question 87

If $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1^2 + x_2^2, g : \mathbb{R}^2 \to \mathbb{R}^2, g(x_1, x_2) = (x_2, x_1)$, then, if $h(x_1, x_2) = f(g(x_1, x_2))$:

$$\nabla h(x_1, x_2) = ?$$

a. $\nabla h(x_1, x_2) = (2x_1, 2x_2)$.

b. $\nabla h(x_1, x_2) = (2x_2, 2x_1)$.

c. $\nabla h(x_1, x_2) = (1, 1)$.

$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x_1, x_2) = x_1^2 + x_2^2$

$g : \mathbb{R}^2 \to \mathbb{R}, \quad g(x_1, x_2) = (x_2, x_1)$

$h(x_1, x_2) = f(g(x_1, x_2))$

$h(x_1, x_2) = x_2^2 + x_1^2$

$\vec{\nabla} h(x_1, x_2) = \left(\frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_2}\right) = (2x_1, 2x_2)$

$$\begin{cases} \frac{\partial h}{\partial x_1} = 2x_1 \\ \frac{\partial h}{\partial x_2} = 2x_2 \end{cases}$$

Question 88

If $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 e^{x_2}$,, where $g : \mathbb{R} \to \mathbb{R}^2$, $g(t) = (t, \log(t))$ then, if $h(t) = f(g(t))$:

$$h'(t) = ?$$

a. $h'(t) = t + 1$.

b. $h'(t) = t^2 + 1$.

c. $h'(t) = 2t$.

$F : \mathbb{R}^2 \to \mathbb{R}$, $F(x_1, x_2) = x_1 . e^{x_2}$

$g : \mathbb{R} \to \mathbb{R}^2$, $g(t) = (t, \log t)$

$h(t) = F(g(t)) \Rightarrow h(t) = t . e^{\log t} \Rightarrow \frac{dh(t)}{dt} = \frac{d(t . e^{\log t})}{dt} \Rightarrow$

$\Rightarrow \frac{dh}{dt} = \frac{dt}{dt} . e^{\log t} + t . \frac{de^{\log t}}{dt}$

$\frac{dh}{dt} = e^{\log t} + t . \frac{1}{t \ln 10}$ | $\frac{d \log_a x}{dx} = \frac{1}{x . \ln a}$

$\frac{dh}{dt} = e^{\log t} + \frac{1}{\ln 10}$ ← NO ANSWER

Reminder: $\begin{cases} \log x = \log_{10} x \\ \ln x = \log_e x \end{cases}$

Apparently in Italy $\log x = \ln x$

$\rightarrow$ $\log x = \ln x$: $h(t) = F(g(t)) = t . e^{\ln t} = t^2 \Rightarrow h'(t) = 2t$

Question 89

If $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1^2 + x_1x_2, g : \mathbb{R} \to \mathbb{R}^2, g(t) = (t^2, t)$, then, if $h(t) = f(g(t))$:

a. $h'(t) = t(2t-1)^2 + t$.

b. $h'(t) = 4t^2 + 2t + 1$.

c. $h'(t) = t(2t+1)^2 - 2t^2$.

$f(x_1, x_2) = x_1^2 + x_1 . x_2$

$g(t) = (t^2, t)$

$h(t) = t^4 + t^2 \Rightarrow h'(t) = 4t^3 + 2t$ $\quad \nexists$ ANSWER (AGAIN ...)

Question 90

If $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1^2 + x_2^2, g : \mathbb{R}^2 \to \mathbb{R}^2, g(x_1, x_2) = (x_1e^{x_2}, x_2)$, then, if $h(x_1, x_2) = f(g(x_1, x_2))$:

a. $\nabla h(x_1, x_2) = (2x_1e^{x_2}(e^{x_2} + x_1), 2e^{x_2})$.

b. $\nabla h(x_1, x_2) = (2x_1e^{2x_2}(e^{x_1} + x_1), 2x_2)$.

c. $\nabla h(x_1, x_2) = (2x_1e^{x_2}(e^{x_2} + x_1), 2x_2)$.

$f(x_1, x_2) = x_1^2 + x_2^2$

$g(x_1, x_2) = (x_1 e^{x_2}, x_2)$

$h(x_1, x_2) = f(g(x_1, x_2))$

$h(x_1, x_2) = x_1^2 . e^{2x_2} + x_2^2$

$\frac{\partial h}{\partial x_1} = 2x_1 . e^{2x_2}$

$\frac{\partial h}{\partial x_2} = 2x_1^2 . e^{2x_2} + 2x_2$

GUESS what?!

NO ANSWER ...

Question 91

If $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1x_2, g : \mathbb{R} \to \mathbb{R}^2, g(t) = (t, t^2)$, then, if $h(t) = f(g(t))$:

a. $h'(t) = 3t^2$.

b. $h'(t) = 3t^3$.

c. $h'(t) = t^2$.

$f(x_1, x_2) = x_1 . x_2$

$g(t) = (t, t^2)$

$h(t) = f(g(t))$

$h(t) = t . t^2 = t^3$

$h'(t) = 3t^2$

Question 92

If

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

then:

a. $K_2(A) = 4$.

b. $K_2(A) = 2$.

c. $K_2(A) = \frac{4}{3}$.

$\sigma = \{ |2|, |3|, |2|, |4| \}$

$\sigma_{max} = \max\{\sigma\} = 4$

$\sigma_{min} = \min\{\sigma\} = 2$

$K_2(A) = \frac{\sigma_{max}}{\sigma_{min}} \Rightarrow K_2(A) = \frac{4}{2} \Rightarrow K_2(A) = 2$

Question 93

If vector

$$v = \begin{bmatrix} 10^6 \\ 0 \end{bmatrix}$$

is approximated by vector

$$\tilde{v} = \begin{bmatrix} 999996 \\ 1 \end{bmatrix},$$

then in $\|\cdot\|_2$ the relative error between v and $\tilde{v}$ is:

a. $\sqrt{17} \cdot 10^{-6}$.

b. None of the above.

c. $4 \cdot 10^{-6}$.

1. Relative error between v and $\tilde{v}$ in the L2-norm (Euclidean Space)

$$\text{Relative error} = \frac{\|v - \tilde{v}\|_2}{\|v\|_2}$$

i) $v - \tilde{v} = \begin{pmatrix} 10^6 - 999996 \\ 0 \quad\quad 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \end{pmatrix}$

ii) $\|v - \tilde{v}\|_2 = \sqrt{4^2 + (-1)^2} = \sqrt{17}$

iii) $\|v\|_2 = \sqrt{(10^6)^2 + 0^2} = 10^6$

iv) $\frac{\|v - \tilde{v}\|_2}{\|v\|_2} = \frac{\sqrt{17}}{10^6} = \sqrt{17} \cdot 10^{-6}$

Question 94

If

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

then:

a. $K_2(A) = \frac{1}{2}$.

b. $K_2(A) = 4$.

c. $K_2(A) = 2$.

$\sigma = \{|2|, |1|, |2|, |4|\}$

$\sigma_{max} = max\{\sigma\} = 4$

$\sigma_{min} = min\{\sigma\} = 1$

$K_2(A) = \dfrac{\sigma_{max}}{\sigma_{min}} = \dfrac{4}{1} \Rightarrow K_2(A) = 4$

Question 95

A random variable $X : \Omega \to \mathcal{T}$ is continuous when:

a. $\mathcal{T}$ is countable.

b. $\mathcal{T} = \mathbb{R}$.

c. Ω is continuous.

THEORY RECAP:

1. Continuos Random Variables (X):

A RV X is continuous if it can take any value in an interval of $\mathbb{R}$, and its probability distribution can be described by a probability density function (PDF)

2. Target Space (T) :

The range of target space T of X must support a continuous range of values for X to be continuous.

(A) T - countable (INCORRECT)

countable → discrete values. This describe a discrete random variable, not a continuous.

(C) Ω - continuous (INCORRECT)

Ω represents the set of all possibles outcome of an experiment. While Ω can be continuous, it does not directly define whether X is continuos.

Question 96

A random variable $X : \Omega \to \mathcal{T}$ is discrete when:

a. $\mathcal{T} = \mathbb{R}$.

b. Ω is countable.

c. $\mathcal{T}$ is countable.

THEORY RECAP:

1. DISCRETE RANDOM VARIABLE:

A random variable X is discrete if it takes values from a COUNTABLE set. Each value has a positive probability mass, and the probability distribution is described by the PROBABILITY MASS FUNCTION.

Question 97

If

$$A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

then:

a. $x = (1, 2)^T$ is an eigenvector of A.

b. $x = (2, 1)^T$ is an eigenvector of A.

c. $x = (0, 0)^T$ is an eigenvector of A.

$A\vec{x} = \lambda \vec{x}$

$(A - \lambda E)\vec{x} = \vec{0}$

$\det(A - \lambda E) = 0$

$\det\begin{pmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix} = 0$

$(4-\lambda)(3-\lambda) - 2 = 0$

$12 - 4\lambda - 3\lambda + \lambda^2 - 2 = 0$

$\lambda^2 - 7\lambda + 10 = 0$

$\Delta = 49 - 4.1.10$

$\Delta = 49 - 40$

$\Delta = 9 \Rightarrow \sqrt{\Delta} = 3$

$\lambda : \frac{7 \pm 3}{2} = \begin{cases} \lambda_1 = 5 \\ \lambda_2 = 2 \end{cases}$

@ $\lambda = 5$:

$\begin{pmatrix} 4-5 & 2 \\ 1 & 3-5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$-x_1 + 2x_2 = 0$

$x_1 = 2x_2$

$\vec{x} = t\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}^*$

@ $\lambda = 2$:

$\begin{pmatrix} 4-2 & 2 \\ 1 & 3-2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$2x_1 + 2x_2 = 0$

$x_1 = -x_2$

$\vec{x} = t\begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}^*$

Question 98

If

$$A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

then:

a. $\lambda = 5$ is the eigenvalue associated with the eigenvector $x = (2,1)^T$.

b. $\lambda = 2$ is the eigenvalue associated with the eigenvector $x = (2,1)^T$.

c. $\lambda = 2$ is the eigenvalue associated with the eigenvector $x = (1,2)^T$.

$A\vec{x} = \lambda \vec{x}$

$(A - \lambda E)\vec{x} = \vec{0}$

$\det(A - \lambda E) = 0$

$\det\begin{pmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix} = 0$

$(4-\lambda)(3-\lambda) - 2 = 0$

$12 - 4\lambda - 3\lambda + \lambda^2 - 2 = 0$

$\lambda^2 - 7\lambda + 10 = 0$

$\Delta = 49 - 4.1.10$

$\Delta = 49 - 40$

$\Delta = 9 \Rightarrow \sqrt{\Delta} = 3$

$\lambda : \frac{7 \pm 3}{2} = \begin{cases} \lambda_1 = 5 \\ \lambda_2 = 2 \end{cases}$

@ $\lambda = 5$:

$\begin{pmatrix} 4-5 & 2 \\ 1 & 3-5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$-x_1 + 2x_2 = 0$

$x_1 = 2x_2$

$\vec{x} = t\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}^*$

@ $\lambda = 2$:

$\begin{pmatrix} 4-2 & 2 \\ 1 & 3-2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$2x_1 + 2x_2 = 0$

$x_1 = -x_2$

$\vec{x} = t\begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}^*$

Question 99

If

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

then:

a. $x = (1,1,0)^T$ is an eigenvector of A.

b. $x = (0,1,0)^T$ is an eigenvector of A.

c. $x = (0,-1,1)^T$ is an eigenvector of A.

$\det(A - \lambda E) = 0$

$A = \begin{pmatrix} 4-\lambda & 0 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & -1-\lambda \end{pmatrix}$

$(4-\lambda)(2-\lambda)(\lambda+1) = 0$

$\{\lambda_1 = 4 \,;\, \lambda_2 = 2 \,;\, \lambda_3 = -1\}$

@ $\lambda = 4$

$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$0.x_1 + 0.x_2 + 0.x_3 = 0 \qquad \forall x_1, x_2, x_3 \in \mathbb{R}$

$0.x_1 - 2x_2 + 0.x_3 = 0 \qquad x_2 = 0 \wedge \forall x_1, x_3 \in \mathbb{R}$

$0.x_1 + 0.x_2 - 5x_3 = 0 \qquad x_3 = 0 \wedge \forall x_1, x_2 \in \mathbb{R}$

$\vec{x} = t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \forall t \in \mathbb{R}$

@ $\lambda = 2$

$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$2x_1 = 0 \qquad x_1 = 0 \quad \forall x_2, x_3 \in \mathbb{R}$

$0 = 0 \qquad \forall x_1, x_2, x_3 \in \mathbb{R}$

$-3x_3 = 0 \qquad x_3 = 0 \quad \forall x_1, x_2 \in \mathbb{R}$

$\Rightarrow \vec{x} = t\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \forall t \in \mathbb{R}$

@ $\lambda = -1$

$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$\begin{cases} 5x_1 = 0 \\ 3x_2 = 0 \\ 0 = 0 \end{cases}$ $\quad \begin{matrix} x_1 = 0 \quad \forall x_2, x_3 \in \mathbb{R} \\ x_2 = 0 \quad \forall x_1, x_3 \in \mathbb{R} \\ \forall x_1, x_2, x_3 \in \mathbb{R} \end{matrix}$

$\vec{x} = t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}$

Question 100

If

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

then:

a. $x = (0,0)^T$ is an eigenvector of A.

b. $x = (1,0)^T$ is an eigenvector of A.

c. $x = (1,1)^T$ is an eigenvector of A.

$\det\begin{pmatrix} 4-\lambda & 0 \\ 0 & 2-\lambda \end{pmatrix} = 0$

$(4-\lambda)(2-\lambda) = 0$

$\begin{cases} \lambda_1 = 4 \\ \lambda_2 = 2 \end{cases}$

@ $\lambda = 4$

$\begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \vec{0}$

$\begin{cases} 0 \cdot x_1 + 0 \cdot x_2 = 0 & \forall x_1, x_2 \in \mathbb{R} \\ 0 \cdot x_1 - 2x_2 = 0 & x_2 = 0 \ \forall x_1 \in \mathbb{R} \end{cases}$

$\vec{x} = t\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \forall t \in \mathbb{R}$

@ $\lambda = 2$

$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{cases} 2 \cdot x_1 + 0 \cdot x_2 = 0 & x_1 = 0 \ \forall x_2 \in \mathbb{R} \\ 0 \cdot x_1 + 0 \cdot x_2 = 0 & \forall x_1, x_2 \in \mathbb{R} \end{cases}$

$\vec{x} = t\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}$

Question 101

If

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

then:

a. $\lambda = 2$ is the eigenvalue associated with the eigenvector $x = (1,0)^T$.

b. $\lambda = 2$ is the eigenvalue associated with the eigenvector $x = (0,1)^T$.

c. $\lambda = 1$ is the eigenvalue associated with the eigenvector $x = (1,0)^T$.

$\begin{pmatrix} 2-\lambda & 0 \\ 0 & 1-\lambda \end{pmatrix} = 0$

$(2-\lambda)(1-\lambda)$

$\lambda_1 = 2 \ \wedge \ \lambda_2 = 1$

@ $\lambda = 2$

$\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$0x_1 + 0x_2 = 0 \quad \forall x_1, x_2 \in \mathbb{R}$

$0x_1 - x_2 = 0 \quad x_2 = 0 \ \forall x_1 \in \mathbb{R}$

$\vec{x} = t\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

@ $\lambda = 1$

$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$x_1 + 0x_2 = 0 \quad x_1 = 0 \ \forall x_2 \in \mathbb{R}$

$0x_1 + 0x_2 = 0 \quad \forall x_1, x_2 \in \mathbb{R}$

$\vec{x} = t\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}$

Question 102

If $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$ and

$$Ax = \lambda x$$

for $\lambda \in \mathbb{R}$, then:

a. For any $c \in \mathbb{R}, c \neq 0$, cx is an eigenvector of A.

b. cx is an eigenvector of A if and only if $c = 1$.

c. None of the above.

if $A \in \mathbb{R}^{n \times n}$, $\vec{x} \in \mathbb{R}^n$

for $\lambda \in \mathbb{R}$ $\perp$ $A\vec{x} = \lambda\vec{x}$ $\quad$ λ - eigenvalue

$\vec{x}$ - eigenvector

(A) if $\vec{x}$ - eigenvector, $\perp$ $\forall c \in \mathbb{R} \setminus \{0\} \Rightarrow c\vec{x}$ - eigenvector

□ $Ax' = \lambda x'$, $x' = cx$, x - eigenvector

$A(cx) = \lambda(cx)$

$c(Ax) = c(\lambda x)$ ■

Question 103

If

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

then:

a. $\lambda = 2$ is the eigenvalue associated with the eigenvector $x = (0, 1, 0)^T$.

b. $\lambda = -1$ is the eigenvalue associated with the eigenvector $x = (0, 0, 1)^T$.

c. $\lambda = 1$ is the eigenvalue associated with the eigenvector $x = (1, 0, 0)^T$.

$$\det \begin{pmatrix} 2-\lambda & 0 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & -1-\lambda \end{pmatrix} = 0$$

$$(2-\lambda)(1-\lambda)(1+\lambda) = 0$$

@ $\lambda = 2$:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} -x_2 = 0 & x_2 = 0 \quad \forall x_1, x_3 \in \mathbb{R} \\ -3x_3 = 0 & x_3 = 0 \quad \forall x_1, x_2 \in \mathbb{R} \end{cases}$$

$$\vec{x} = t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \forall t \in \mathbb{R}$$

@ $\lambda = 1$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} x_1 = 0 & x_1 = 0 \quad \forall x_2, x_3 \in \mathbb{R} \\ -2x_3 = 0 & x_3 = 0 \quad \forall x_1, x_2 \in \mathbb{R} \end{cases}$$

$$\vec{x} = t \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \forall t \in \mathbb{R}$$

@ $\lambda = -1$

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} 3x_1 = 0 & x_1 = 0 \quad \forall x_2, x_3 \in \mathbb{R} \\ 2x_2 = 0 & x_2 = 0 \quad \forall x_1, x_3 \in \mathbb{R} \end{cases}$$

$$\vec{x} = t \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \forall t \in \mathbb{R}$$

Question 104

In $\mathcal{F}(10, 2, -2, 2)$, if $x = \pi$, $y = e$, and $z = fl(x) - fl(y)$, then:

a. $fl(z) = 0.43 \times 10^0$.

b. $fl(z) = 0.44 \times 10^0$.

c. $fl(z) = 0.40 \times 10^0$.

Base $b = 10$

Precision $p = 2$

$[e_{min}; e_{max}] = [-2; 2]$

$\pi \approx 3{,}14159 \qquad fl(x) = fl(\pi) = 3{,}1 \cdot 10^0$

$e \approx 2{,}71828 \qquad fl(x) = fl(e) = 2{,}7 \cdot 10^0$

$z = \pi - e$

$z = 3{,}14159 - 2{,}71828$

$z = 0{,}42331$

$fl(z) = fl(fl(\pi) - fl(e))$

$fl(z) = fl(3{,}1 - 2{,}7)$

$fl(z) = fl(0{,}4) = 0{,}4 \cdot 10^0$ (NO ANSWER!)

Question 105

In $\mathcal{F}(10,2,-2,2)$, if $x = \pi, y = e$, and $z = fl(x) \cdot fl(y)$, then:

a. $fl(z) = 0.84 \times 10^1$.

b. $fl(z) = 0.0837 \times 10^2$.

c. $fl(z) = 0.837 \times 10^1$.

$\pi \approx 3,14 \qquad fl(\pi) = 3,1 \cdot 10^0$

$e \approx 2,71 \qquad fl(e) = 2,7 \cdot 10^0$

$z = fl(\pi) \cdot fl(e)$

$z = 3,1 \cdot 2,7 \cdot 10^0 = 8,37 \cdot 10^0$

$fl(z) = 8,4 \cdot 10^0 = 0,84 \cdot 10^1$

Question 106

In $\mathcal{F}(10,6,-3,3)$, if $x = 192.403$, $y = 0.635782$, and $z = fl(x) + fl(y)$, then:

a. $fl(z) = 0.193039 \times 10^3$.

b. $fl(z) = 0.193038 \times 10^3$.

c. $fl(z) = 0.193038782 \times 10^3$.

$x = 192,403 \qquad fl(x) = 192,403 \cdot 10^0$

$y = 0,635782 \qquad fl(y) = 0,63578 \cdot 10^0$

$fl(x) + fl(y) = 193,03878$

$fl(z) = 193,038$

$= 0,193038 \cdot 10^3$

Question 107

In $\mathcal{F}(10,2,-2,2)$, if $x = \pi, y = e$, and $z = fl(x) + fl(y)$, then:

a. $fl(z) = 0.585 \times 10^1$.

b. $fl(z) = 0.58 \times 10^1$.

c. $fl(z) = 0.59 \times 10^1$.

$x = \pi \approx 3,14 \qquad fl(x) = 3,1 \cdot 10^0$

$y = e \approx 2,71 \qquad fl(y) = 2,7 \cdot 10^0$

$z = fl(x) + fl(y)$

$z = 5,8 \cdot 10^0$

$fl(z) = 0,58 \cdot 10^1$

Question 108

If $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$, then x^* is a minimum point if and only if:

a. $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semi-definite.

b. $\nabla f(x^*) = 0$.

c. $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite.

- FIRST-ORDER CONDITION:
  1. if $\vec{\nabla} f(x_0) = \vec{0}$
- SECOND-ORDER CONDITION:
  2. if $\vec{\nabla}^2 f(x_0) > 0$ - strict local minimum
     $\vec{\nabla}^2 f(x_0) \geq 0$ - local minimum

Question 109

Gradient descent methods solve the optimization problem

$$\min_x f(x)$$

by:

a. Generating a sequence $\{x_k\}_k$ such that, given x_0, computes $x_{k+1} = x_k - \alpha \nabla f(x_k)$ for $\alpha > 0$ step-size.

b. Generating a sequence $\{x_k\}_k$ such that, given x_0, computes $x_{k+1} = x_k + \alpha \nabla f(x_k)$ for $\alpha > 0$ step-size.

c. Generating a sequence $\{x_k\}_k$ such that, given x_0, computes $x_{k+1} = x_k - \alpha \nabla f(x_k)$ for $\alpha \neq 0$ step-size.