Necessary Minimum Background Test

① Multivariable Calculus

$y = 2\sin(x)e^{-x}$

$\frac{\partial}{\partial x} = -2\sin(x)e^{-x} + 2\cos(x)e^{-x}$

$\boxed{\frac{\partial}{\partial x} = 2e^{-x}(\cos(x) - \sin(x))}$

② Linear Algebra

$X = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad z = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

A) $\begin{pmatrix} 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

$(1 \times 2 + 3 \times 3)$

$(2 + 9)$

$\boxed{11}$

B) $Xy = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \cdot 1 + 4 \cdot 3 \\ 1 \cdot 1 + 2 \cdot 3 \end{pmatrix} = \begin{pmatrix} 2 + 12 \\ 1 + 6 \end{pmatrix}$

$= \boxed{\begin{pmatrix} 14 \\ 7 \end{pmatrix}}$

c) $\begin{vmatrix} 2 & 4 \\ 1 & 2 \end{vmatrix} = 4 - 4 = 0 \Rightarrow$ $\boxed{\text{not invertible}}$

D) $\begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \quad R_1 = 2R_2$

1 linearly independent row/column

$\Rightarrow \boxed{\text{rank} = 1}$

(3) Probability and Statistics

$$S = (x_1, x_2, x_3, x_4, x_5) = (1, 1, 0, 1, 0)$$

A) $\text{mean} = \dfrac{1+1+0+1+0}{5} = \boxed{0.6}$

B) $\frac{1}{5}\left((1-0.6)^2 + (1-0.6)^2 + (0-0.6)^2 + (1-0.6)^2 + (0-0.6)^2\right)$

$= \frac{1}{5}(1.2)$

$= \boxed{0.24}$

c) $0.5^5 = \boxed{0.03125}$

D)

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

$x_1 = 1$, $x_2 = 1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 0$

$L(p) = P(x_1, x_2, x_3, x_4, x_5 \mid p)$

$= P(x_1 \mid p)\, P(x_2 \mid p)\, P(x_3 \mid p)\, P(x_4 \mid p)\, P(x_5 \mid p)$

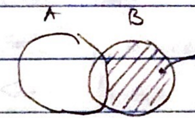
$= p \cdot p \cdot (1-p) \cdot p \cdot (1-p)$

$\hat{p} = \text{argmax } L(p) = \boxed{\frac{3}{5}}$

E) $P(x = T \mid y = b) = \dfrac{P(x = T, y = b)}{P(y = b)} = \dfrac{0.1}{0.1 + 0.15} = \dfrac{0.1}{0.25} = \boxed{0.4}$

④ Probability Axioms

A) $P(A \cup B) = P(A \cap (B \cap A^c))$



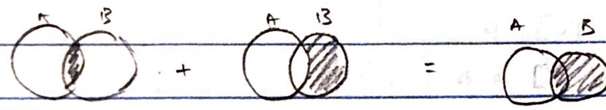
$B \cap A^c$

$\Rightarrow$ $\boxed{\text{FALSE}}$

B) $P(A \cup B) = P(A) + P(B)$

$\Rightarrow$ $\boxed{\text{TRUE}}$ if A and B are disjoint

c) $P(A) = P(A \cap B) + P(A^c \cap B)$



$\Rightarrow$ $\boxed{\text{FALSE}}$

D) $P(A|B) = P(B|A)$

$\Rightarrow$ $\boxed{\text{FALSE}}$

E) $P(A_1 \cap A_2 \cap A_3) = P(A_3 | (A_2 \cap A_1)) P(A_2 | A_1) P(A_1)$

$= P(A_3 | (A_2 \cap A_1)) P(A_2 \cap A_1)$

$P(A \cap B) = P(A|B) P(B)$

$= P(A_3 \cap (A_2 \cap A_1))$

$= P(A_1 \cap A_2 \cap A_3)$ ✓

$\Rightarrow$ $\boxed{\text{TRUE}}$

(5) Discrete and Continuous Distributions

A) Gaussian $\Rightarrow$ (v) $\frac{1}{\sqrt{(2\pi)\sigma^2}} \exp\left(1 - \frac{1}{2\sigma^2}(x-\mu)^2\right)$

B) Exponential $\Rightarrow$ (iv) $\lambda e^{-\lambda x}$ when $x \geq 0$; 0 otherwise

C) Uniform $\Rightarrow$ (ii) $\frac{1}{b-a}$ when $a \leq x \leq b$; 0 otherwise

D) Bernoulli $\rightarrow$ (i) $p^x(1-p)^{1-x}$, when $x \in \{0,1\}$; 0 otherwise

E) Binomial $\Rightarrow$ (iii) $\binom{n}{x} p^x (1-p)^{n-x}$

(6) Mean and Variance

A) mean : $E[x] = p$

variance : $V[x] = p(1-p)$

B) The variance of $2x$ is $4\sigma^2$.

The variance of $X+3$ is σ^2

(7) Algorithms

A) i) $f(n) = \ln(n)$, $g(n) = \lg(n)$

$\Rightarrow$ Both, because $f(n)$ and $g(n)$ are equivalent up to a certain constant.

ii) $f(n) = 3^n$, $g(n) = n^{10}$

$\Rightarrow$ $g(n) = O(f(n))$, because as n gets larger, $f(n)$ becomes significantly larger.

iii) $f(n) = 3^n$, $g(n) = 2^n$

$\Rightarrow$ $g(n) = O(f(n))$, because as n gets larger, $f(n)$ becomes significantly larger

B) You can essentially do a binary search on the array by having 2 indices that check the elements at index i and index $i+1$. You first look at the middle element and if $arr[i] = 0$ and $arr[i+1] = 1$, then return the index i. Otherwise, if $arr[i] = 0$ and $arr[i+1] = 0$, then recursively call the function again but with the new array from the beginning to index i. If $arr[i] = 1$, then recursively call the function from index i to the end. Because of this divide and conquer algorithm, we are essentially splitting the size of the array in half each recursion. So an array of 16 elements will only take at most 4 iterations of the algorithm to find the index and an array of 32 elements will take at most 5 iterations. Thus, the algorithm runs in $O(\log n)$.

Moderate Background Test

(8) Probability and Random Variables

A) $E[XY] = E[X]\,E[Y]$

$$E[XY] = \sum_i \sum_j x_i y_j f_{xy}(x_i, y_j)$$

$$= \sum_i x_i f_x(x_i) \cdot \sum_j y_j f_y(y_j)$$

$$\boxed{E[XY] = E[X] \cdot E[Y]}$$ ✓

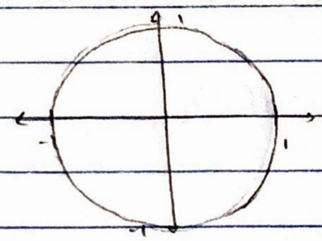
B) i) Because the probability of rolling a fair die and getting a 3 is 1/6, the Law of Large Numbers implies that rolling a fair die 6000 times will thus return a 3 $6000 \cdot 1/6 = 1000$ times

ii) $\sqrt{n}\,(\bar{x} - \frac{1}{2}) \xrightarrow{n \to \infty} N(0, \frac{1}{4})$

The Central Limit Theorem implies that the expression on the left will evaluate to the expression on the right.

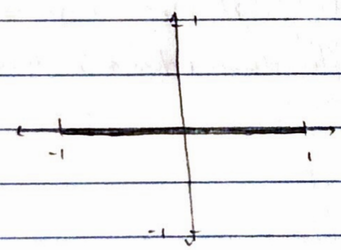
(9) Linear Algebra

A) i) $\|x\|_2 \le 1$ $\qquad \|x\|_2 = \sqrt{\sum_i x_i^2}$

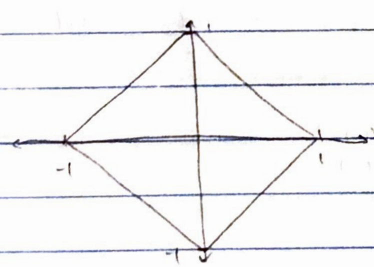


$\sqrt{x_1^2 + x_2^2} = 1$

ii) $\|x\|_0 \le 1$ $\qquad \|x\|_0 = \sum_{i: x_i \neq 0} 1$

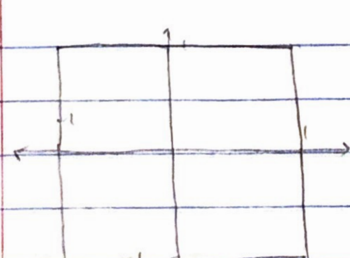


iii) $\|x\|_1 \le 1$ $\qquad \|x\|_1 = \sum_i |x_i|$



$\|x_1, x_2\| = |x_1| + |x_2| = 1$

iv) $\|x\|_\infty \le 1$ $\qquad \|x\|_\infty = \max_i |x_i|$



$\max\{|1|, |x_2|\} = 1 \quad \forall |x_2| \le 1$

$\max\{|-1|, |x_2|\} = 1 \quad \forall |x_2| \le 1$

$\max\{|x_1|, |1|\} = 1 \quad \forall |x_1| \le 1$

$\max\{|x_1|, |-1|\} = 1 \quad \forall |x_1| \le 1$

B) i) In an $n \times n$ square matrix A, an eigenvalue is a scalar λ if there is an eigenvector x such that $Ax = \lambda x$.
An eigenvector is a non-zero vector that changes only by a scalar factor when a linear transformation is applied to it.

ii) $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

$$Ax = \lambda x$$
$$(A - \lambda I_n)x = 0$$
$$= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = 0$$
$$= \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = 0$$

	2	$-\lambda$
2	4	-2λ
$-\lambda$	-2λ	λ^2

$$\det(A - \lambda I_n) = 0$$
$$(2-\lambda)^2 - (1)(1) = 0$$
$$4 - 4\lambda + \lambda^2 - 1 = 0$$
$$\lambda^2 - 4\lambda + 3 = 0$$
$$(\lambda - 3)(\lambda - 1) = 0$$
$$\boxed{\lambda_1 = 3} \quad \boxed{\lambda_2 = 1}$$

3
-3 -1
-4

$A \cdot v_1 = \lambda_1 \cdot v_1$

$(A - \lambda_1)v_1 = 0$

$\begin{pmatrix} 2-3 & 1 \\ 1 & 2-3 \end{pmatrix} v_1 = 0$

$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$-v_{1,1} + v_{1,2} = 0 \qquad v_{1,1} = v_{1,2}$

$v_{1,1} - v_{1,2} = 0$

$\boxed{v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$

$A \cdot v_2 = \lambda_2 \cdot v_2$

$(A - \lambda_2)v_2 = 0$

$\begin{pmatrix} 2-1 & 1 \\ 1 & 2-1 \end{pmatrix} v_2 = 0$

$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$v_{2,1} + v_{2,2} = 0$

$v_{2,1} = -v_{2,2}$

$\boxed{v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}}$

iii) If we start with our original formula to find eigenvalues, we have:

$$Ax = \lambda x$$

where λ is our eigenvalue and x is our eigenvector.

If we multiply each side by A, we get:

$$A \cdot A x = \lambda x \cdot A$$
$$A^2 x = \lambda (Ax) \Longleftarrow Ax = \lambda x$$
$$A^2 x = \lambda (\lambda x)$$
$$A^2 x = \lambda^2 x$$

We can then generalize this by multiplying each side by A^{k-1}:

$$(A^{k-1}) \cdot A x = \lambda x \cdot (A^{k-1})$$
$$A^k x = \lambda (A^{k-1} x)$$
$$A^k x = \lambda (\lambda^{k-1} x)$$
$$A^k x = \lambda^k x$$

Thus, the eigenvalues of A^k are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$ and the eigenvector x remains the same, proving that each eigenvector of A is still an eigenvector of A^k.

(c) i) $u(x) = a^T x$ $\quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ $\quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$a^T = (a_1 \dots a_n)$

$a^T x = (a_1 \dots a_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\quad \frac{\partial}{\partial x} a^T x = \left(\frac{\partial}{\partial x_1} a^T x_1 \quad \dots \quad \frac{\partial}{\partial x_n} a^T x_n \right)$

$= (a_1 \quad \dots \quad a_n)$

$$\boxed{\frac{\partial a^T x}{\partial x} = a^T}$$

ii) $u(x) = x^T A x$

$u(x+h) = (x+h)^T A (x+h)$

$= x^T A x + h^T A x + x^T A h + h^T A h$

$r_x(h) = h^T A h$ $\quad = u(x) + x^T (A + A^T) h + r_x(h)$

$r_x(h) = o(\|h\|)$, $h \to 0$

$\nabla u(x) = x^T (A + A^T)$

$\nabla u(x)(h) = z^T h$

$z = (A + A^T) x$

$A = A^T$ $\quad \boxed{\frac{\partial}{\partial x} x^T A x = 2Ax}$

$\frac{\partial^2 x}{\partial x^2} x^T A x = \frac{\partial}{\partial x} 2Ax$

$$\boxed{\frac{\partial^2 x}{\partial x^2} = 2A}$$

(D) i) $w^T x + b = 0$

$(x_1 - x_2)$ is a vector parallel to the line

$w^T x_1 + b = 0$

$w^T x_2 + b = 0$

$w^T x_1 + \cancel{b} = w^T x_2 + \cancel{b}$

$w^T x_1 = w^T x_2$

$w^T x_1 - w^T x_2 = 0$

$w^T (x_1 - x_2) = 0$

$\Rightarrow$ $\boxed{\text{w is orthogonal to the line } w^T x + b = 0}$

ii) $w^T x + b = 0$ Let p = point on $w^T x + b = 0$ line that's closest to origin

$p = \min x^T x$ $2p = \lambda w$

$p = \frac{1}{2}\lambda w$

$w^T x + b = 0$

$w^T \left(\frac{1}{2}\lambda w\right) + b = 0$

$\frac{1}{2}\lambda w^T w + b = 0$

$\lambda = \frac{-2b}{w^T w}$

$p = \frac{1}{\cancel{2}}\left(\frac{-\cancel{2}b}{w^T w}\right) w$

$= \frac{-b}{w^T w} w$

$D = \sqrt{p^T p} = \sqrt{\left(\frac{-b}{w^T w}\right)^2 w^T w}$

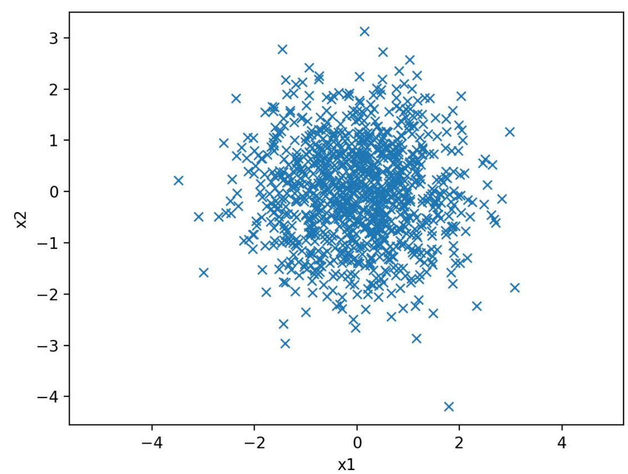
$= \frac{b}{w^T w}\sqrt{w^T w}$

$= \frac{b}{\sqrt{w^T w}}$

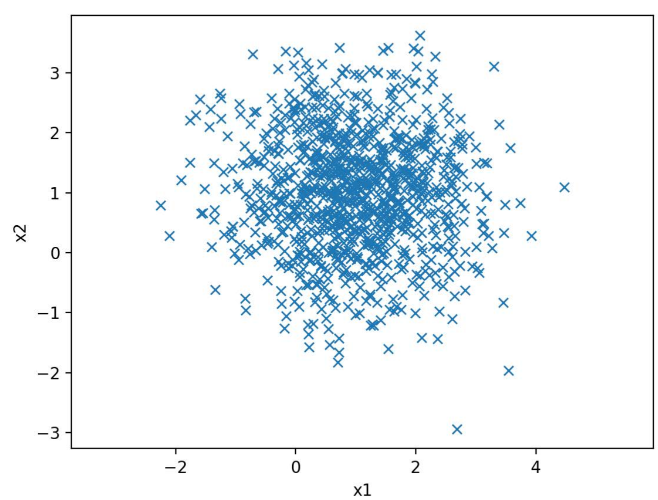
$$\boxed{D = \frac{b}{\|w\|}}$$

10 Sampling from a Distribution

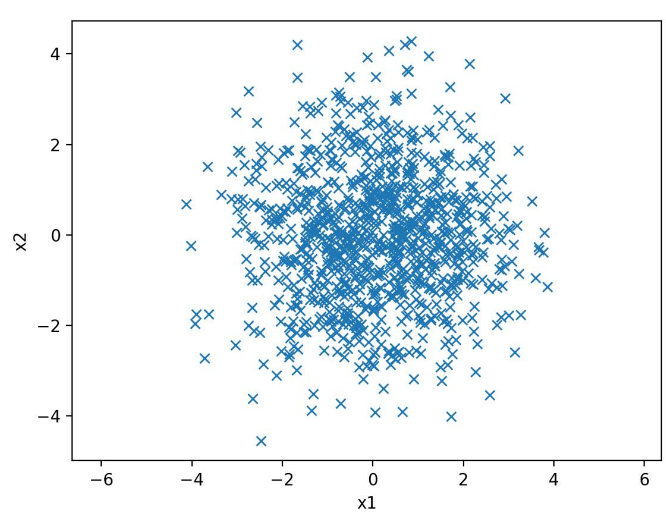
(a)



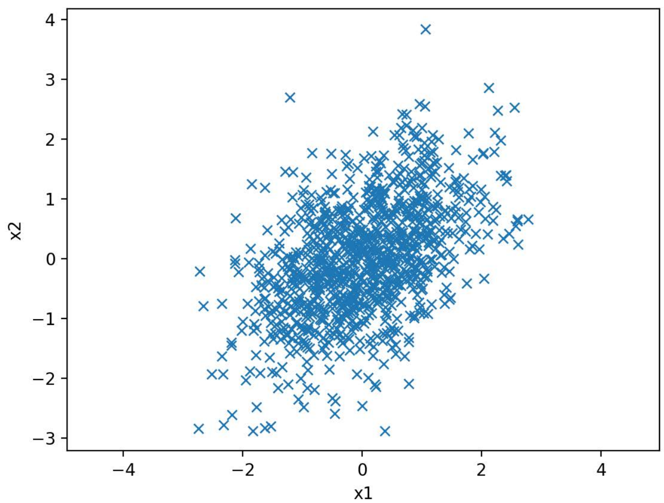
(b)



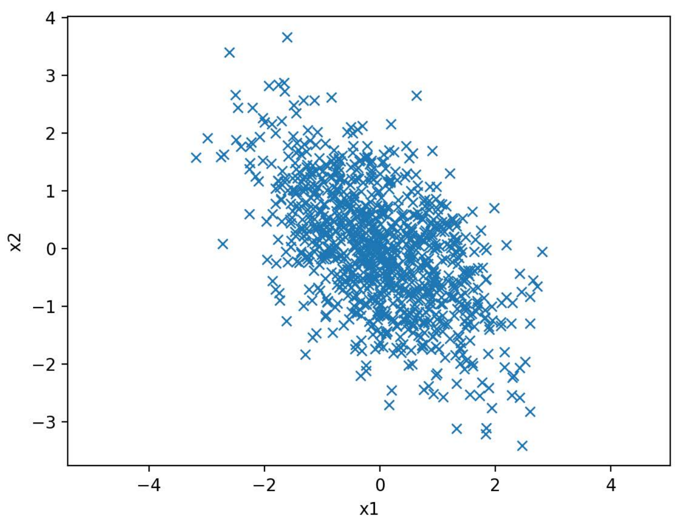
(c)



(d)



(e)



11 Eigendecomposition

```python
import numpy as np
from numpy import linalg as LA

def eigendecomp():
    index = 0
    large_index = 0
    large_eig = 0

    w, v = LA.eig(np.array([[1,0],[1,3]]))
    for i in w:
        if i > large_eig:
            large_eig = i
            large_index = index
        index += 1

    print(v[:,large_index])

eigendecomp()
```

(12) Data

A) Stanford Dogs Dataset

B) vision.stanford.edu/aditya86/ImageNetDogs/

C) This dataset has images of 120 breeds of dogs that was built using thousands of images and annotations from ImageNet in order to categorize the breed of a dog based on an image of one.

D) 120 categories, 20,580 images

E) To construct this database, the SIFT (Scale-invariant Feature Transform) algorithm was used on the training images in order to detect and describe certain features in images. SIFT works by constructing a scale space, using the Laplacian of Gaussian to find interesting key points in an image, and then using these key points to generate SIFT features for each image. There can be thousands of SIFT features per image (such as the shape of a body part) and these features can ultimately be used to detect and identify objects in an image. In this case, these features are used to identify the breed of a dog based on an image of it.