Necessary Minimum Background Test [45 pts]

While you are welcome to use online resources, such as Wolfram-Alpha, you should be able to solve these problems by hand.

1 Multivariate Calculus [2 pts]

Consider $y = z\sin(x)e^{-x}$. What is the partial derivative of y with respect to x?

$$y = z\,\mathrm{Sin}(x)e^{-x}$$

$$\boxed{\frac{\partial y}{\partial x} = z\left(\cos(x)e^{-x} - e^{-x}\,\mathrm{Sin}(x)\right)}$$

2 Linear Algebra [8 pts]

Consider the matrix $\boldsymbol{X}$ and the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ below:

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

(a) What is the inner product $\boldsymbol{y}^T\boldsymbol{z}$?

$y^T = [1, 3]$

$[1, 3] \begin{bmatrix} 2 \\ 3 \end{bmatrix} = [2 + 9] = [11]$

(b) What is the product $\boldsymbol{X}\boldsymbol{y}$?

$\begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 + 8 \\ 1 + 4 \end{pmatrix} = \begin{pmatrix} 10 \\ 5 \end{pmatrix}$

(c) Is $\boldsymbol{X}$ invertible? If so, give the inverse; if not, explain why not.

$\begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$

NO, it is not independant so it cannot have an inverse

(d) What is the rank of $\boldsymbol{X}$?

The rank is 1

3 Probability and Statistics [10 pts]

Consider a sample of data S obtained by flipping a coin five times. $X_i, i \in \{1, \dots, 5\}$ is a random variable that takes a value 0 when the outcome of coin flip i turned up heads, and 1 when it turned up tails. Assume that the outcome of each of the flips does not depend on the outcomes of any of the other flips. The sample obtained $S = (X_1, X_2, X_3, X_4, X_5) = (1, 1, 0, 1, 0)$.

(a) What is the sample mean for this data?

$$\bar{X} = \frac{1+1+0+1+0}{5} = \boxed{0.6}$$

(b) What is the unbiased sample variance ?

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{5}\cdot\left((0.4)^2 + (0.4)^2 + (-0.6)^2 + (0.4)^2 + (-0.6)^2\right)$$

$$S^2 = \frac{1.2}{5} = \boxed{0.24}$$

(c) What is the probability of observing this data assuming that a coin with an equal probability of heads and tails was used? (*i.e.*, The probability distribution of X_i is $P(X_i = 1) = 0.5$, $P(X_i = 0) = 0.5$.)

BerONl:

$n = 5$

x = successes or # of 1's

$$P(x) = \tfrac{1}{2}^{x}\left(1 - \tfrac{1}{2}\right)^{n-x}$$

$$= \left(\tfrac{1}{2}\right)^3\left(\tfrac{1}{2}\right)^2 = 0.03125$$

(d) Note the probability of this data sample would be greater if the value of the probability of heads $P(X_i = 1)$ was not 0.5 but some other value. What is the value that maximizes the probability of the sample S? [Optional: Can you prove your answer is correct?]

$P(X_i = 1) = 0.6$

Define

$p(X_i = 1) = X_p$

prove $\frac{\partial}{\partial x}\left(x_p^3 (1 - x_p)^2\right) =$

$$\rightarrow \frac{\partial}{\partial x}\left(x_p^3 \cdot 1 - 2x_p^4 + x_p^5\right) = 5x_p^4 - 8x_p^3 + 3x_p^2)$$

$$= x_p^2(5x_p^2 - 8x_p + 3) \rightarrow x = 0, \quad (5x_p^2 - 8x_p + 3)$$

$$(5x_p - 3)(x_p - 1) = 0$$

$$x_p = 0, 1, \tfrac{3}{5}$$

Since $x_p = 0$ & $x_p = 1$ give a probability density of zero, we get $x_p = \frac{3}{5}$ maximum

(e) Given the following joint distribution between X and Y, what is $P(X = T | Y = b)$?

$P(X,Y)$		Y: a	b	c
X	T	0.2	0.1	0.2
	F	0.05	0.15	0.3

$P(X = T | Y = b) = 0.1$

4 Probability axioms [5 pts]

Let A and B be two discrete random variables. In general, are the following true or false? (Here A^c denotes complement of the event A.)

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

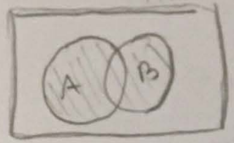
(b) $P(A \cup B) = P(A) + P(B)$ False

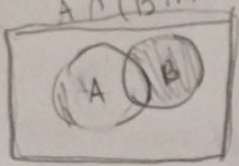
(c) $P(A) = P(A \cap B) + P(A^c \cap B)$ False

(d) $P(A|B) = P(B|A)$ False

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

ⓐ $A \cup B$ $\qquad$ $A \cap (B \cap A^c))$



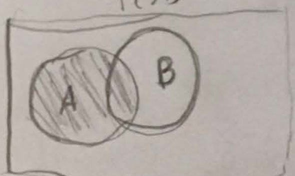


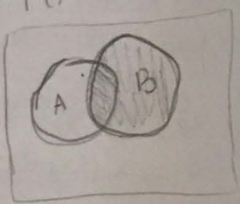
$\boxed{\text{False}}$

ⓑ $\boxed{\text{False}}$

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

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





$\boxed{\text{False}}$

ⓓ $P(A|B) = P(B|A)$ $\boxed{\text{False}}$

$$\frac{P(A \cap B)}{P(B)} \neq \frac{P(B \cap A)}{P(A)}$$

ⓔ $P(A_3|(A_2 \cap A_1)) \cdot P(A_2|A_1)P(A_1)$

$$\frac{P(A_3 \cap A_2 \cap A_1)}{P(A_1 \cap A_2)} \cdot \frac{P(A_2 \cap A_1)}{P(A_1)} \cdot P(A_1) = P(A_3 \cap A_2 \cap A_1)$$

$\boxed{\text{True}}$

5 Discrete and Continuous Distributions[5 pts]

Match the distribution name to its formula.

(a) Gaussian	(i) $p^x(1-p)^{1-x}$, when $x \in \{0,1\}; 0$ otherwi
(b) Exponential	(ii) $\frac{1}{b-a}$ when $a \leq x \leq b; 0$ otherwise
(c) Uniform	(iii) $\binom{n}{x}p^x(1-p)^{n-x}$
(d) Bernoulli	(iv) $\lambda e^{-\lambda x}$ when $x \geq 0; 0$ otherwise
(e) Binomial	(v) $\frac{1}{\sqrt{(2\pi)\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

a → v

b → iv

c → ii

d → i

e → iii

6 Mean and Variance[5 pts]

(a) What is the mean and variance of a *Bernoulli*(p) random variable?

Mean of Bernoulli(p) is p

Variance is p(1 − p)

Expectation: $f_x(x) = p^x (1-p)^{1-x}$ $x \in 1, 0$, $p(1) = p$

$E(f_x(x)) = 1 \cdot p^1 + 0 \cdot (1-p) = \boxed{p}$

Variance: $Var[x] = E(x^2) - (E(x))^2$

$E(x^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$

then

$Var(x) = \boxed{p - p^2}$

(b) If the variance of a zero-mean random variable X is σ^2, what is the variance of $2X$? What about the variance of $X + 3$?

$Var[x] = 6^2$

$Var[2x] = 2^2 \cdot Var[x] = \boxed{4 6^2}$

$Var[x+3] = Var[x] + Var[3] = \boxed{6^2}$

7 Algorithms [10 pts]

(a) **Big-O notation**

For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, or both. Briefly justify your answers.

i. $f(n) = ln(n), g(n) = lg(n)$. Note that ln denotes log to the base e and lg denotes log to the base 2.

Since they are both log functions, they grow in the same manner

So the

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

$O(g(n)) < O(f(n))$, " "

Exponential algorithum runtime grows faster than polynomial. All are false.

$f(n)$ 3^n $g(n)$ 10^n

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

$O(3^n) = O(2^n)$ as both are exponential

so $f(n) = O(g(n))$ Both

$g(n) = O(f(n))$

(b) **Divide and Conquer**

Assume that you are given an array with n elements all entries equal either to 0 or +1 such that all 0 entries appear before +1 entries. You need to find the index where the transition happens, *i.e.*, you need to report the index with the last occurrence of 0. Give an algorithm that runs in time $O(\log n)$. Explain your algorithm in words, describe why the algorithm is correct, and justify its running time.

I use a binary Search algorithom which will divide the entries by 2 each time and Search the entry pile with the transition. Ex if I land on a 1, I split off all values after that one and Search in all values lower. My algorithm is recursive and the amount in the array is divided by 2 each time giving a complexity of $O(\log n)$

8 Probability and Random Variables [5 pts]

(a) **Mutual and Conditional Independence** If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]E[Y]$.

$$E[XY] = \int_{-\infty}^{\infty} f_{xy}(x,y) \cdot xy \, dx dy = \int_{-\infty}^{\infty} f_x(x) f_y(y) \, xy \, dx dy$$

$\rightarrow$ Split integrals $= \int_{-\infty}^{\infty} f_x(x) x \, dx \cdot \int_{-\infty}^{\infty} f_Y(y) y \cdot dy = E[x] \cdot E[y]$

(b) **Law of Large Numbers and Central Limit Theorem**
Provide one line justifications.

i. If a fair die is rolled 6000 times, the number of times 3 shows up is close to 1000.

$P(\text{any value of the Die}) = \frac{1}{6}$ then $E[3] = np$ (binomial)
$= \frac{1}{6} \cdot 6000 = 1000.$

ii. If a fair coin is tossed n times and $\bar{X}$ denotes the average number of heads, then the distribution of $\bar{X}$ satisfies

$$\sqrt{n}(\bar{X} - \tfrac{1}{2}) \xrightarrow{n \to \infty} \mathcal{N}(0, \tfrac{1}{4})$$

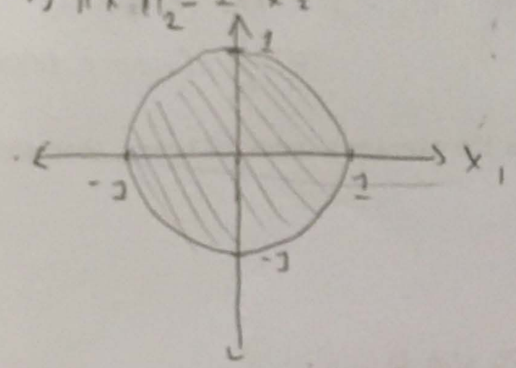
A binomial distribution - like the one here - can be expressed as a normal distribution as sample size approaches ∞. with $Var[X] = \frac{1}{2}^2 = \frac{1}{4}$

9 Linear Algebra [20 pts]

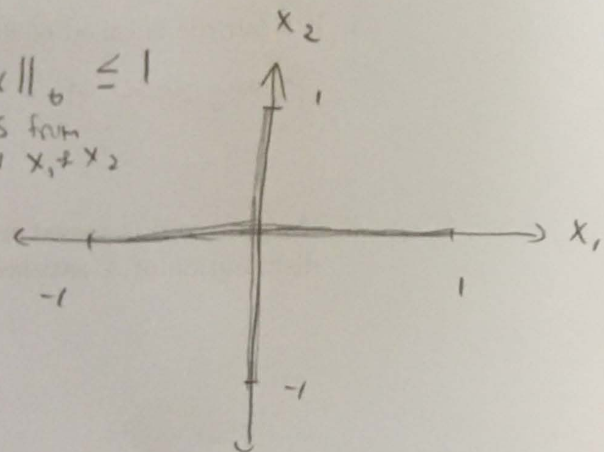
(a) **Vector Norms** [4 pts]
Draw the regions corresponding to vectors $\boldsymbol{x} \in \mathbb{R}^2$ with following norms (you can hand draw or use software for this question):

i. $\|\boldsymbol{x}\|_2 \leq 1$ (Recall $\|\boldsymbol{x}\|_2 = \sqrt{\sum_i x_i^2}$.)
ii. $\|\boldsymbol{x}\|_0 \leq 1$ (Recall $\|\boldsymbol{x}\|_0 = \sum_{i: x_i \neq 0} 1$.)
iii. $\|\boldsymbol{x}\|_1 \leq 1$ (Recall $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$.)
iv. $\|\boldsymbol{x}\|_\infty \leq 1$ (Recall $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$.)

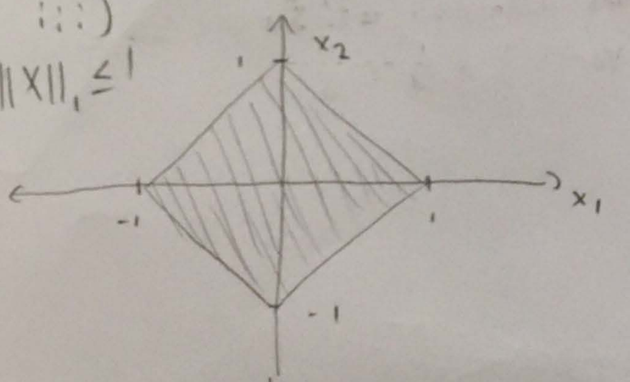
i) $\|x\|_2 \leq 1$



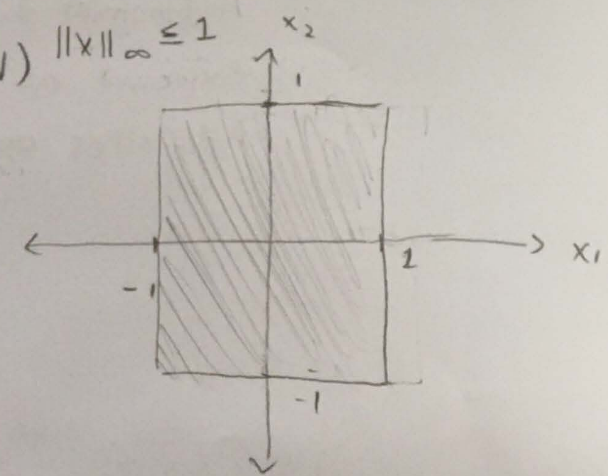
ii) $\|x\|_0 \leq 1$
-axis from
1 → -1 x_1 & x_2



iii)
$\|x\|_1 \leq 1$



iv) $\|x\|_\infty \leq 1$



(b) **Matrix Decompositions [6 pts]**

i. Give the definition of the eigenvalues and the eigenvectors of a square matrix.

An Eigenvector $\vec{V}$ of Matrix A is a non-zero vector where A applies to $\vec{V}$ equals $\vec{V}$ multiplied by a Scaler λ in its Same Space. This Scaler (λ) is an eigenvalue of A. So $A\vec{v} = \lambda\vec{V}$ therefore the linear transformation A on $\vec{V}$ can be defined by λ.

ii. Find the eigenvalues and eigenvectors of

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

$\text{Det}(A - \lambda I) = 0$

$\text{Det}\begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = 0$

$(2-\lambda)^2 - 1 = 0$

$\lambda^2 - 4\lambda + 4 - 1 = 0$

$\lambda^2 - 4\lambda + 3 = 0$

$(\lambda - 3)(\lambda - 1) = 0$

$\boxed{\lambda_1 = 3 \quad \lambda_2 = 1}$

$\vec{V}_2$

$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 1 \cdot \begin{pmatrix} a \\ b \end{pmatrix}$

$2a + b = a$

$a + 2b = b$

$a = -b$

I freely choose $a = 1$ then $b = -1$

therefore $\boxed{\vec{V}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}}$ for $\lambda_2 = 1$

$\vec{V}_1$

$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 3\begin{pmatrix} a \\ b \end{pmatrix}$

$2a + b = 3a$

$b = a$

Since I choose $a = 1, b = 1$

$\boxed{\vec{V}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$ for $\lambda_1 = 3$

iii. For any positive integer k, show that the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$, the k^{th} powers of the eigenvalues of matrix $\boldsymbol{A}$, and that each eigenvector of $\boldsymbol{A}$ is still an eigenvector of $\boldsymbol{A}^k$.

Eigenvector $\vec{V}$ & eigenvalue λ are defines by

if $A\vec{V}_1 = \lambda_1 \vec{V}_1$

then $k = 2$

$$A^2 \vec{V} = A(A\vec{V}_1)$$
$$= A \cdot \lambda_1 \vec{V}_1$$
$$= \lambda_1 A \vec{V}$$
$$= \lambda_1 \cdot \lambda_1 A$$
$$= \lambda_1^2 A$$

In all cases of k

$$A^{k-1} \cdot A\vec{V}_1 = A^{k-1} \lambda_1 \vec{V}_1 \Leftrightarrow A^k \vec{V}_1 = \lambda_1 (\lambda_1^{k-1} \vec{V}_1)$$

$$A^k \vec{V}_1 = \lambda_1^k \vec{V}_1$$ holds true for all k positive integers

(c) **Vector and Matrix Calculus [5 pts]**
Consider the vectors $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{a}}$ and the symmetric matrix $\boldsymbol{A}$.

i. What is the first derivative of $\boldsymbol{a}^T\boldsymbol{x}$ with respect to $\boldsymbol{x}$?

$$\vec{a}^T\vec{x} = [a_1 \cdots a_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + \dots + a_nx_n$$

$$\frac{\partial \vec{a}^T\vec{x}}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}(a_1x_1 + \dots + a_nx_n) \\ \vdots \\ \frac{\partial}{\partial x_n}(a_1x_1 + \dots + a_nx_n) \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \boxed{\vec{a}}$$

ii. What is the first derivative of $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ with respect to $\boldsymbol{x}$? What is the second derivative?

A is Symmetric

chain rule

$$\vec{\nabla}(x^TAx) = \frac{\partial(\vec{x}^TA\vec{x})}{\partial\vec{x}^T} = \vec{x}^T \cdot A^T + \vec{x}^TA = \vec{x}^T(A^T+A)$$

w.r.t. x

- Since A is Symmetric, $A^T + A = 2A$ so $= \boxed{2A\vec{x}}$

$$\vec{\nabla}''(\vec{x}^TA\vec{x}) = \vec{\nabla}\, 2A\vec{x} = 2 \cdot \begin{bmatrix} \frac{\partial}{\partial x_1} \cdot \vec{x}^TA \\ \vdots \\ \frac{\partial}{\partial x_n}\vec{x}^TA \end{bmatrix} = \boxed{2 \cdot A}$$

$2 \cdot \vec{x}^TA$

(d) **Geometry [5 pts]**

i. Show that the vector $\boldsymbol{w}$ is orthogonal to the line $\boldsymbol{w}^T\boldsymbol{x} + b = 0$. (Hint: Consider two points $\boldsymbol{x}_1, \boldsymbol{x}_2$ that lie on the line. What is the inner product $\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$?)

$\vec{X}_1 = \begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix}$ $\vec{X}_2 = \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix}$ $\vec{W}^T = (W_1 \ W_2)$

$\vec{X}_1$ & $\vec{X}_2$ lie on the line

$A = W^T \vec{X}_1 + b = 0$ then $A - B = W^T(X_1 - X_2) = 0$

$B = W^T \vec{X}_2 + b = 0$

Since the dot product of $W^T(X_1 - X_2) = 0$ W^T must be orthogonal to $X_1 - X_2$ and the line they represent.

ii. Argue that the distance from the origin to the line $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ is $\frac{b}{\|\boldsymbol{w}\|_2}$.

The distance from a point to a hyperplane ($W^T\vec{X} + b = 0$ is a hyperplane) is defined as $aX + b = d$ closest distance $= \frac{|d - b|}{\|a\|_2}$

I argue that $W^TX + b = 0$ has a distance from the origin of $\frac{|b|}{\|W\|_2}$ following the above formula.

as well,

$W^T\vec{X} + b = 0 \Rightarrow$ $\vec{X}$ can be represented as αW for some Scaler α because W is orthogonal to the line.

Now, $\alpha W^T W + b = 0 \Rightarrow \alpha = \frac{-b}{\|W\|_2^2}$

therefore $\|\vec{X}\| = \|\alpha W\| = \alpha \|W\| = \frac{-b\|W\|_2}{\|W\|_2^2} = \frac{-b}{\|W\|_2}$ where $-b$ is assumed to be negative giving $\frac{b}{\|W\|_2}$

11 Eigendecomposition [2.5 pts]

Write a python program to compute the eigenvector corresponding to the largest eigenvalue of the following matrix and submit the computed eigenvector.

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix}$$

```
A = np.array([[1,0],[1,3]])
e_val, e_vect = np.linalg.eig(A)
evalmax_Vect = e_vect[e_val.tolist().index(max(e_val))]
evalmax_vect
```

```
>>> array([0. , 0.89442719])
```

12 Data [5 pts]

There are now lots of really interesting data sets publicly available to play with. They range in size, quality and the type of features and have resulted in many new machine learning techniques being developed.

Find a public, free, supervised (i.e. it must have features *and* labels), machine learning dataset. You may NOT list a data set from 1) The UCI Machine Learning Repository or 2) from Kaggle.com. Once you have found the data set, provide the following information:

(a) The name of the data set.

(b) Where the data can be obtained.

(c) A brief (i.e. 1-2 sentences) description of the data set including what the features are and what is being predicted.

(d) The number of examples in the data set.

(e) The number of features for each example. If this is not concrete (i.e. it is text), then a short description of the features.

For this question, do not just copy and paste the description from the website or the paper; reference it, but use your own words. Your goal here is to convince the staff that you have taken the time to understand the data set, where it came from, and potential issues involved.

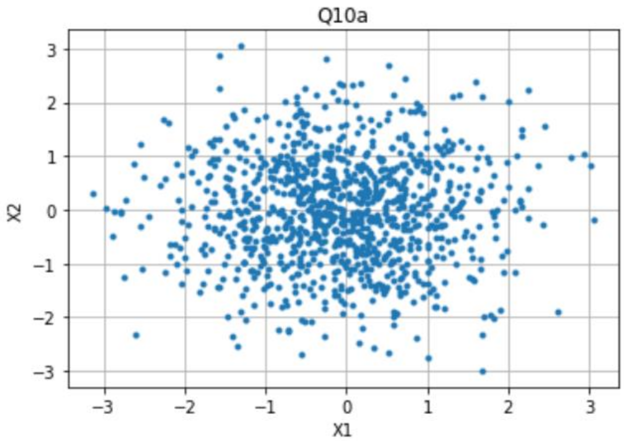
a) Laboratory for Intelligent & Safe Automobiles: Vehicle Detection Data Set

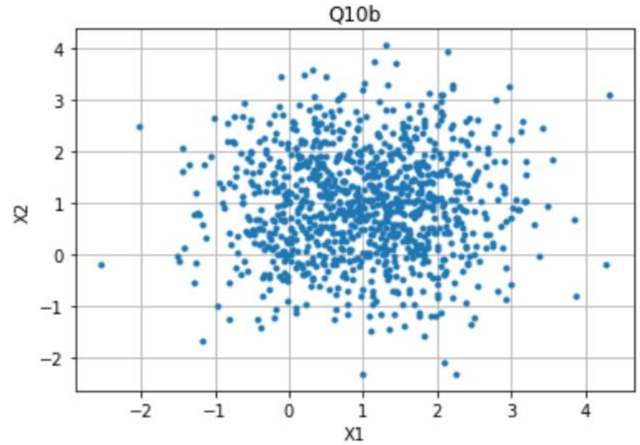
b) cvrr.ucsd.edu/LISA/vehicledetection.html

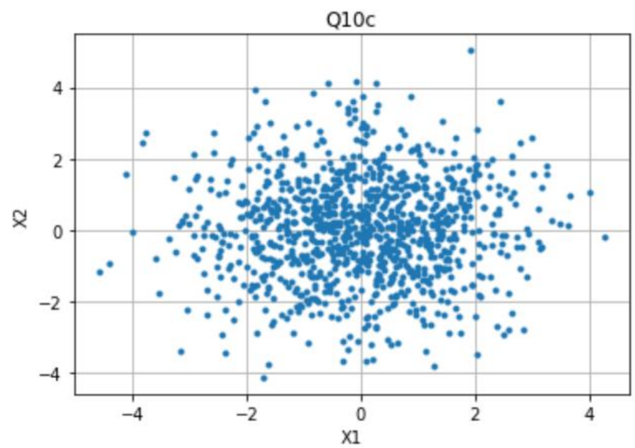
c) The data set consists of a three on-road videos taken from a driven car. The features are obtain through video/image processing and include: time of day, traffic level, Color of cars, driving environment
It is predicting kinds of cars at a specific time and environment (urban vs rural)

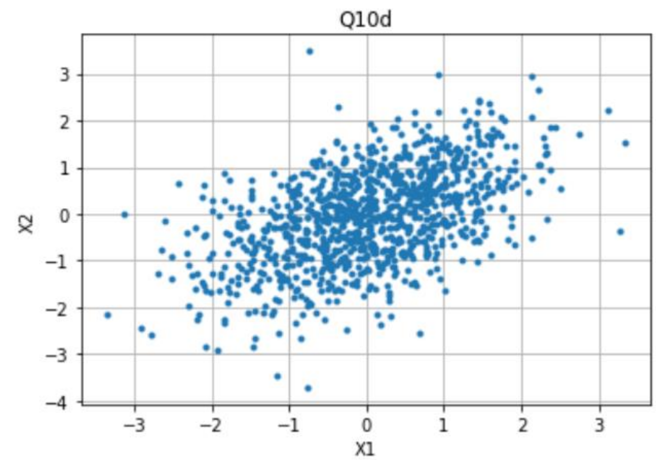
d) 1600 examples are in the set.

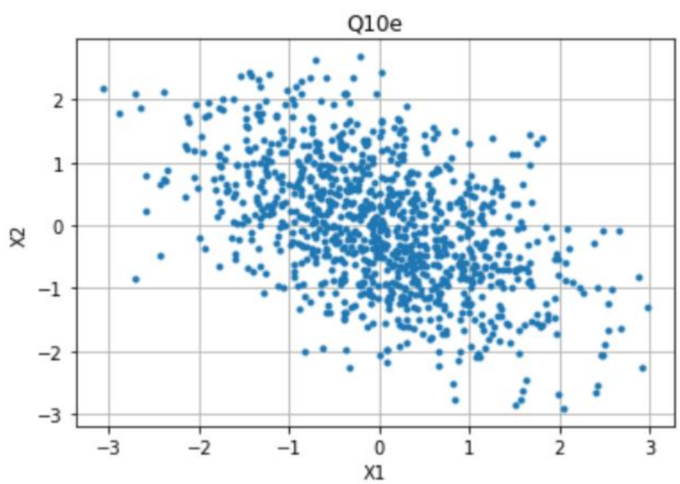
e) There are 10 features for each example


Q10a
X1
X2


Q10b
X1
X2


Q10c
X1
X2


Q10d
X1
X2


Q10e
X1
X2