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?

$$\frac{\partial y}{\partial x} = z\cos(x)e^{-x} - z\sin(x)e^{-x}$$

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 $\boldsymbol{y}^T\boldsymbol{z}$?

$$y^T = (1 \quad 3)$$

$$y^T z = (1 \quad 3)\begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$= 1\times 2 + 3\times 3$$

$$= 11$$

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

$$Xy = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 2\times 1 + 4\times 3 \\ 1\times 1 + 2\times 3 \end{pmatrix}$$

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

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

$$\det(X) = 2\times 2 - 1\times 4 = 0$$

∴ Singular, X is not invertible.

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

$$X = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \xrightarrow{①-②\times 2} \begin{pmatrix} 2 & 4 \\ 0 & 0 \end{pmatrix}$$

rank = 1

3 Probability and Statistics [10 pts]

Consider a sample of data S obtained by flipping a coin five times. $X_i, i \in \{1, \ldots, 5\}$ 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?

$\bar{x} = 3/5 = 0.6$

(b) What is the unbiased sample variance ?

$$S^2 = \frac{3\times(1-0.6)^2 + 2\times(0-0.6)^2}{5-1}$$

$= 0.3$

(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 $P(X_i = 1) = 0.5$, $P(X_i = 0) = 0.5$.)

$P(S) = 0.5^5 = 0.03125$

(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?]

Let the value be a

$P(S) = a^3(1-a)^2$

$= a^5 - 2a^4 + a^3$

Take derivative: $5a^4 - 8a^3 + 3a^2 = 0$

$a = \frac{3}{5}$

$P(\frac{3}{5}) = 1080/3125 = 0.3456$

(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

$P(Y=b) = 0.1 + 0.15 = 0.25$

$P(X=T \cap Y=b) = 0.1$

$$P(X=T|Y=b) = \frac{P(X=T \cap Y=b)}{P(Y=b)} = \frac{0.1}{0.25} = 0.4$$

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

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

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

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

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

(a). False

(b). True

(c). True

(d). False

(e). 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 $x \in \{0,1\}; 0$ otherwise
(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?

$\bar{x} = p$

$var = p(1-p)$

(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 $X + 3$?

$2X$: $var = 4\sigma^2$

$X+3$: $var = \sigma^2$

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.

$f(n) = \ln(n) = \frac{\lg(n)}{\lg(e)}$ $\therefore f(n) = O(g(n))$

$g(n) = \lg(n) = \ln(n)/\ln(2)$ $\therefore g(n) = O(f(n))$

Both are true

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

$3^n > n^{10}$

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

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

$3^n > 2^n$

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

(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.

Take the midpoint, if it is 0 and the next one is 1, then it is where transition occurs. If it is 0 and the next one is 0, then it means that transition occurs in the second half. So we can take the midpoint of the second half (divide the second half into two parts). If it is 1, and the previous one is 0, then it is where the transition is. If the previous 1 is 0, then we can divide the first half into two parts and examine the midpoint.

8 Probability and Random Variables [5 pts]

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

$$E[XY] = \int_{X=-\infty}^{X=\infty} \int_{y=-\infty}^{y=\infty} xy f_{x,y}(x,y)\,dxdy$$

$$= \int_{X=-\infty}^{X=\infty} x f_x(x)\,dx \int_{y=-\infty}^{y=\infty} y f_y(y)\,dy = E[X]E[Y]$$

(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.

According to the Large Number Theorem, the number of times 3 shows up is close to the expectation:

$$E(X=3) = 6000 \times \frac{1}{6} = 1000$$

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

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

$\mu = \frac{1}{2}$, $\sigma^2 = \frac{1}{4}$

Central Limit Theorem,

$$\sqrt{n}(\bar{x} - \mu) = N(0, \sigma^2)$$

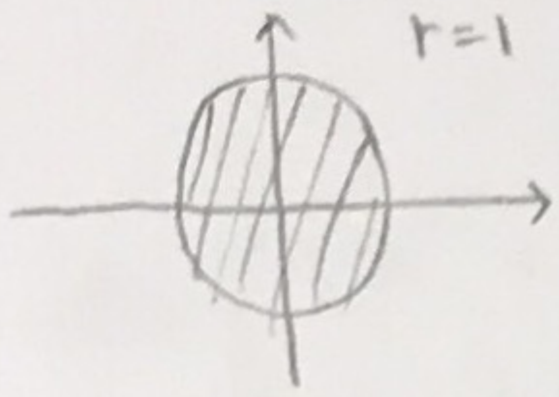
$$\therefore \sqrt{n}(\bar{x} - \tfrac{1}{2}) = N(0, \tfrac{1}{4})$$

9 Linear Algebra [20 pts]

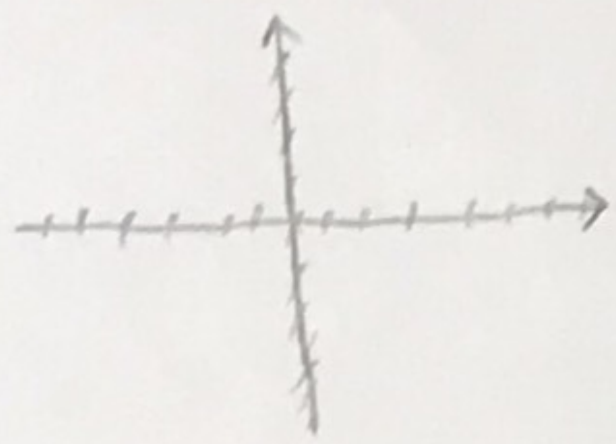
(a) **Vector Norms** [4 pts]
Draw the regions corresponding to vectors $\boldsymbol{x} \in \mathbb{R}^2$ with following norms (you can hand draw or use software for this question):

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 $||\boldsymbol{x}||_\infty = \max_i |x_i|$.)

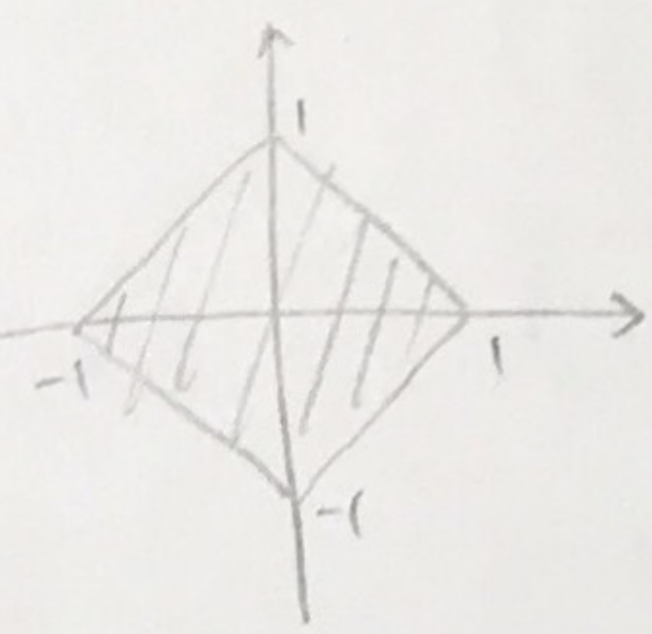
i.



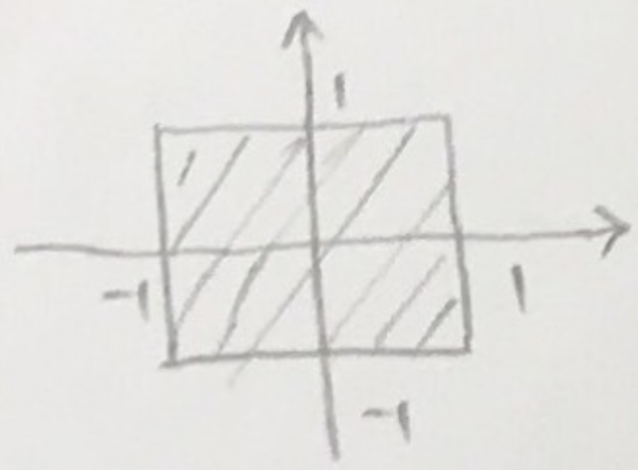
ii.



iii.



iv.



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

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

If there is a vector $x \in R^n \neq 0$ such that $Ax = \lambda x$ for some scalar value λ, then λ is the eigenvalue of A, and x is the eigenvector.

ii. Find the eigenvalues and eigenvectors of

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

$$\det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3$$

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

$$\lambda_1 = 1, \quad \lambda_2 = 3$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$\therefore$ 1, 3

$\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

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 $\boldsymbol{A}^k$.

$$Ax = \lambda x$$

$$S = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix} \leftarrow \text{eigenvectors}$$

$$B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \leftarrow \text{eigenvalues}$$

$$A = SBS^{-1}$$

$$A^k = (SBS^{-1})(SBS^{-1}) \cdots (SBS^{-1})$$

$$= SB(S^{-1}S)B(S^{-1}S) \cdots (BS^{-1})$$

$$= SB^kS^{-1}$$

$$B^k = \begin{pmatrix} \lambda_1^k & \cdots & 0 \\ \vdots & \lambda_2^k & \vdots \\ 0 & \cdots & \lambda_n^k \end{pmatrix}$$

$\therefore$ Eigenvectors are unchanged, eigenvalues of A^k are $\lambda_1^k, \lambda_2^k, \cdots \lambda_n^k$,

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

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

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?

$2Ax$

$2A^T$

(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 $\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$?)

$$W^T X + b = 0$$

$$\therefore W^T X_1 + b = 0 \;, \quad W^T X_2 + b = 0$$

$$W^T (X_1 - X_2) = 0$$

$$W^T (X_1 - X_2) = W \cdot (X_1 - X_2)$$

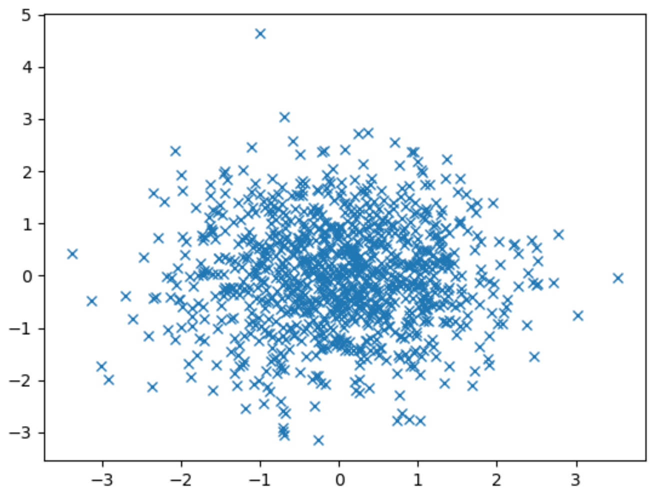
$$\therefore W \cdot (X_1 - X_2) = 0$$

W orthogonal to the line

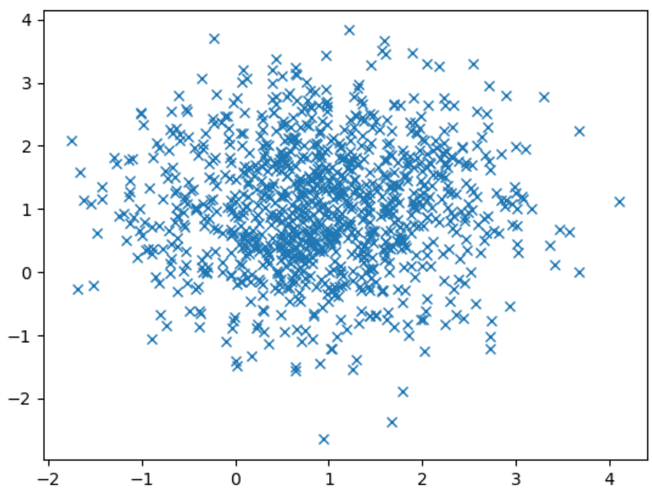
ii. Argue that the distance from the origin to the line $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ is $\frac{b}{||\boldsymbol{w}||_2}$.

$$\frac{W^T 0 + b}{||W||_2} = \frac{b}{||W||_2}$$

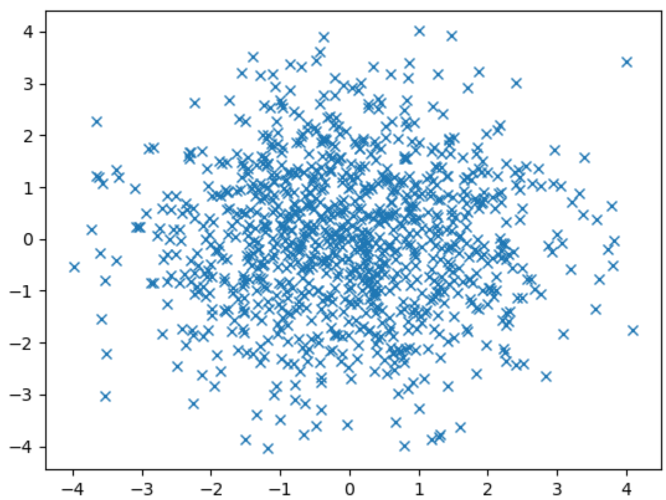
10. (a)



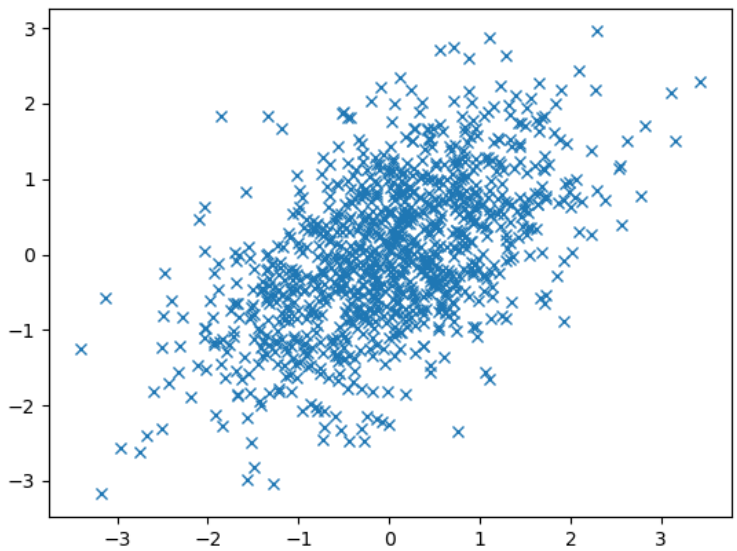
(b)



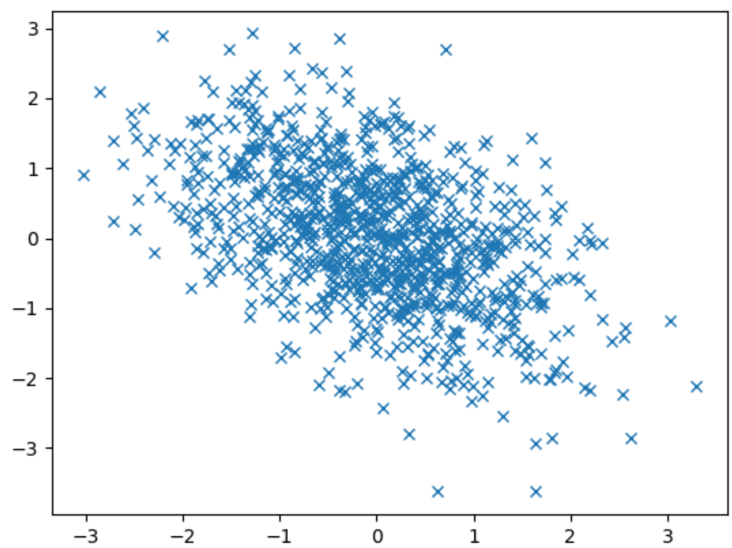
(c)



(d)



(e)



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.

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

.

```
import numpy as np
from numpy import linalg as LA
A = np.array([[1,0],[1,3]])
W.V = LA.eig(A)
result = v[np.argmax(w)]
print(result)
```

- Printed Result:

[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) Milk RadNet Laboratory Analysis

(b) It can be obtained from U.S. Environmental Protection Agency.

(c) This data set records the amount of different radioactive chemicals that are present in milk throughout cities in the U.S. and the milk radiation net is being predicted.

(d). There are 65 examples in the data set.

(e). For each example, there are date posted, date collected, unit, location, states, and 13 radioactive chemical contents in each milk example.