CS M146 Fall 2023 Homework 0 Solutions

UCLA ID: 605900342
Name: Caleb Traxler
Collaborators: N/A

By turning in this assignment, I agree by the UCLA honor code and declare that all of this is my own work.

Problem 1 (Multivariate Calculus)

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

Answer: $y = xsin(z)e^{-x}$

$\frac{\partial y}{\partial x} = \frac{\partial}{\partial x}sin(z)xe^{-x} = sin(z)\frac{\partial}{\partial x}xe^{-x}$. Now allow $u = x, v = e^{-x}$, then I can use the product rule, where $y' = uv' + u'v$. Also note that I can treat $sin(z)$ as a constant.

Upon differentiating u and v, I find that $u' = 1$ and $v' = -e^{-x}$.

Thus, I get $\frac{\partial y}{\partial x} = sin(z)(-xe^{-x} + e^{-x}) \implies \frac{\partial y}{\partial x} = -sin(z)xe^{-x} + sin(z)e^{-x}$

Problem 2 (Linear Algebra)

(a) Consider the matrix X and the vectors y and z below:

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

i. What is the inner product $y^T z$?

$y^T = (1, 3)$, z $= \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, then using the inner product definition, I obtain

$\langle y^T, z \rangle \implies (1)(2) + (3)(3) = 2 + 9 = 11$

Thus, the inner product $\langle y^T, z \rangle = 11$

ii. What is the product Xy?

$X = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix}$, y $= \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, then using matrix multiplication, I find that

$$\text{Xy} = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} (2)(1) + (4)(3) \\ (1)(1) + (3)(3) \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix}$$

Thus, the product $Xy = \begin{pmatrix} 14 \\ 10 \end{pmatrix}$

iii. Is X invertible? If so, give the inverse; if not, explain why not.

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

Recall the definition for the inverse of an invertible 2x2 matrix A, such that for some $a, b, c, d \in R$, I know that generally

$$\text{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \implies A^{-1} = \tfrac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Thus, upon using this definition, I find that

$$X^{-1} = \tfrac{1}{(2)(3)-(4)(1)} \begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{-4}{2} \\ \frac{-1}{2} & \frac{2}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & -2 \\ \frac{-1}{2} & 1 \end{pmatrix}$$

Thus, because the determinant of is nonzero, then the 2x2 matrix is invertible, where

$$X^{-1} = \begin{pmatrix} \frac{3}{2} & -2 \\ \frac{-1}{2} & 1 \end{pmatrix}$$

iv. What is the rank of X?

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

Recall the definition of the rank. I know that the rank of a matrix is the dimension of the vector space generated by its columns. This corresponds to the maximal number of linearly independent columns of A.

$det(X) = (2)(3) - (4)(1) = 2 \neq 0$

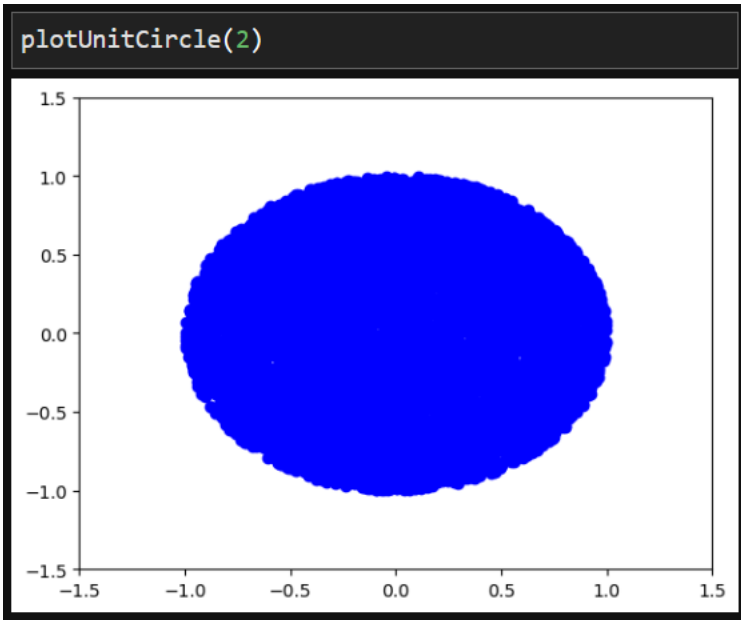
$rank(X) = 2$ because $det(X) \neq 0$ and there exists two columns.

(b) Vector Norms. Draw the regions corresponding to vector $x \in R^2$ with the following norms:

i. $||x||_2 \leq 1$ (Recall $||x||_2 = \sqrt{\sum_i x_i^2}$.)

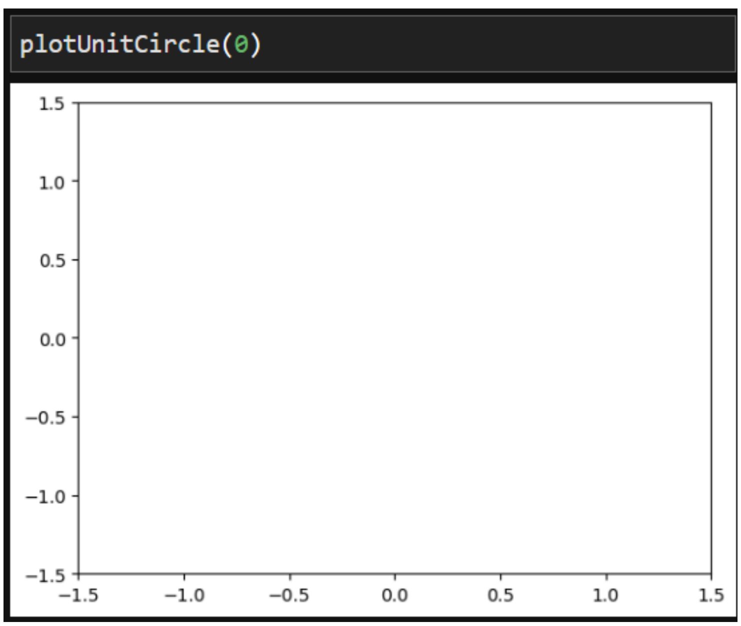
Because $x \in R^2$, then I can assume that there exists two elements in the vector, x_1, $x_2 \leq 1$.

I can further write $(x_1^2 + x_2^2)^{1/2} \leq 1$. Which is the equations of the unit disk bounded by 1.



ii. $||x||_0 \leq 1$ (Recall $||x||_0 = \sum_{i:x\neq 0} 1$.)

$||x||_0 = (\sum_i |x_i|^0)^{\infty=\sum_{i:x\neq 0}} 1 \implies$ the portion of the x-axis between and including $x_1 = -1$ and $x_1 = 1$, with $x_2 = 0$. Also the portion of the y-xis between and including $x_2 = -1$ and $x_2 = 1$ with $x_1 = 0$. Also including the origin.

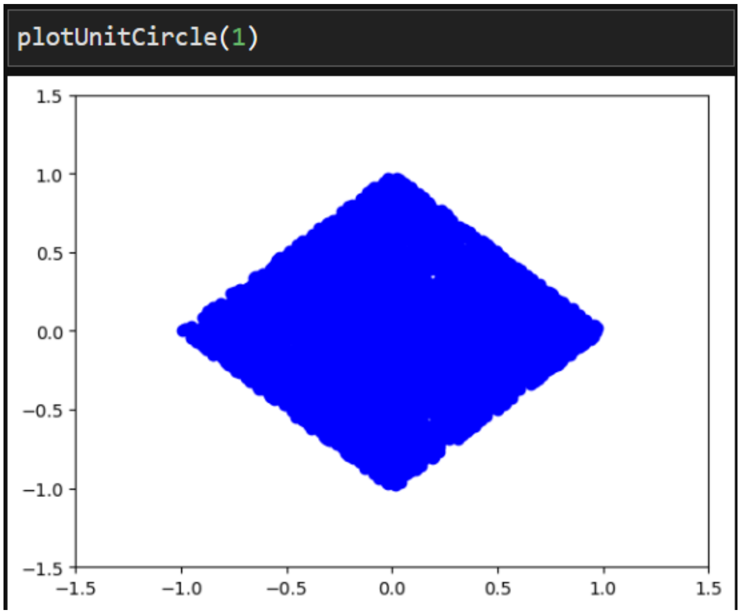


iii. $||x||_1 \leq 1$ (Recall $||x||_1 = \sum_i |x_i|$.)

Similarly to part (i), I know that because the vector $x \in R^2$, then I can represent the vector using two elements, $x_1, x_2 \leq 1$. Then I can further write that $||(x_1, x_2)|| \leq 1 \implies |x_2| + |x_1| \leq 1$. Thus, I obtain the corresponding cases for each quadrant:

Case 1: Quadrant 1 $x_2 \leq 1 - x_1$. Case 2: Quadrant 2 $x_2 \leq 1 + x_1$. Case 3: Quadrant 3 $x_2 \geq -x_1 - 1$.

Case 4: Quadrant 4 $x_2 \geq x_1 - 1$.



iv. $||x||_\infty \leq 1$ (Recall $||x||_\infty = max_i|x_i|$.)

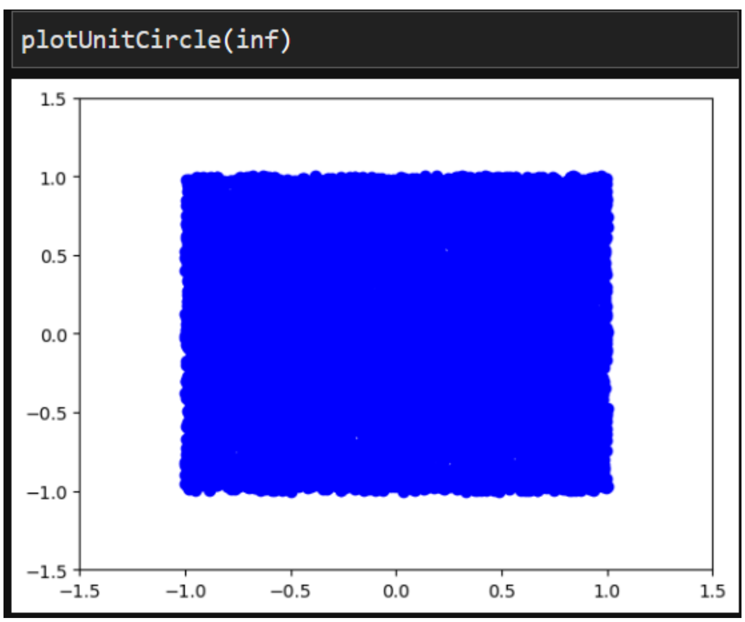
Case 1: Allow $x_1 = 1 \implies max_i\{|1|, |x_2|\} \leq 1$ for all x_2 gives me the vertical line $x_1 = 1$.

Case 2: Allow $x_2 = 1 \implies max_i\{|x_1|, |1|\} \leq 1$ for all x_1 gives me the horizontal line $x_2 = 1$.

Case 3: Allow $x_1 = -1 \implies max_i\{|-1|, |x_2|\} \leq 1$ for all x_2 gives me the vertical line $x_1 = -1$.

Case 4: Allow $x_2 = -1 \implies max_i\{|x_1|, |-1|\} \leq 1$ for all x_1 gives me the horizontal line $x_2 = -1$.

Also note that $||x||_\infty \leq 1 \implies$ unit square bounded bounded by 1.



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

Definition: Let A be an nxn matrix, where $n \in N$, then an eigenvector of A is a nonzero vector $v \in R^n$ such that $Av = \lambda v$, for some scalar λ.

Definition: Let A be an nxn matrix, where $n \in N$, then an eigenvalue of A is a scalar λ such that the equation $Av = \lambda v$ has a nontrivial solution.

If $Av = \lambda v$ for $v \neq 0$, we say that λ is the eigenvalue for v, and the v is an eigenvector for λ.

ii. Find the eigenvalues and eigenvectors of

A = $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. I can find the eigenvalues, if I first re-write the expression

$Av = \lambda v \implies (\lambda I - A)v = 0$, where I is the identity matrix with n dimensions. Using this form, I can now obtain the characteristic polynomial if I take the determinant of the left side, where

$det(\lambda I - A) = 0$. Thus,

$$det \begin{pmatrix} \lambda - 2 & 1 \\ 1 & \lambda - 2 \end{pmatrix} = 0 \implies (\lambda - 2)(\lambda - 2) - (1)(1) = 0 \implies \lambda^2 - 4\lambda + 3 = 0$$

$$\implies (x-1)(x-3) = 0 \implies x = 1, x = 3$$

Thus, I know that there exists two positive eigenvalues, such that $\lambda = 1$ and $\lambda = 3$ both with a multiplicity of one.

Next, using the eigenvalues, I can further compute the eigenvectors. First, suppose there exists some $x, y \in R$.

For $\lambda = 1$, I obtain that

$$\begin{pmatrix} 1-2 & 1 \\ 1 & 1-2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \implies \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \implies \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \implies x = y = 1$$

Thus, for the eigenvalue, $\lambda = 1$, I have a corresponding eigenvector of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

For the eigenvalue $\lambda = 3$, I obtain that

$$\begin{pmatrix} 3-2 & 1 \\ 1 & 3-2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \implies \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \implies \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \implies x = -1, y = 1$$

Thus, for the eigenvalue, $\lambda = 3$, I have the corresponding eigenvector of $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

iii. Question: For any positive integer k, show that the eigenvalues of A^k are $\lambda_1^k, \lambda_2^k, ..., \lambda_n^k$, the kth powers of the eigenvalues of matrix A, and that each eigenvector for A is still an eigenvector of A^k.

Proof: $\forall k \in N^+$ I want to show that the eigenvalues of A^k are $\lambda_1^k, ..., \lambda_n^k$, such that $n \in N$, the k^{th} powers of the eigenvalues of matrix A.

Recall that I can write, by definition of the characteristic polynomials the following expression $Ax = \lambda x$

I find that using the expression, $Ax = \lambda x$, I can multiply both sides by any A^k for $k \in N^+$.

Base Case:

Suppose $k = 1$, then I obtain

$A^1 Ax = A^1 \lambda x \implies AAx = \lambda Ax$, however because I know that $Ax = \lambda x$ holds, then I can write

$A^2 x = \lambda^2 x$. This holds true for $k = 1$.

Induction step:

Generally, I can do the same thing for the induction step, such that if I assume that k is true then I want to show that $k+1$ holds. Multiply both sides by A^k, then

$A^k Ax = A^k \lambda x \implies A^{k+1} x = \lambda^k (\lambda x) \implies A^{k+1} x = \lambda^{k+1} x$ holds true.

Thus, for an arbitrary number, $n \in N$ amount of eigenvalues, the expression, $A^k x = \lambda^k x$ holds for all n, where eigenvalues of A^k are $\lambda_1^k, \lambda_2^k, ..., \lambda_n^k$ by methods of induction.

□

Next I want to show that each eigenvector of A is still an eigenvector of A^k

Proof: This proof is very similar to the previous one. Given a matrix A, an eigenvector v corresponding with eigenvalue λ satisfies the equation: $Av = \lambda v$. I want to show that v is also an eigenvector of A^k for all non-negative integer k.

Base case: k=1, then by definition, I know that $Av = \lambda v$ holds true.

Induction step: Assume $k = n$ holds true, such that $A^n v = \lambda v$, then upon multiplying both sides, I find that $A^{n+1}v = A(A^n v) \implies A^{n+1}v = \lambda^{n+1}v$ therefore by induction the statement, "that each eigenvector of A is still an eigenvector of A^k holds true for all $k \geq 1$.

□

(d) Vector and Matrix Calculus. Consider the vectors x and a and the symmetric matrix A.

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

Suppose for some $n \in N$ I can write the vectors with n elements such that,

$x = [x_1, x_2, x_3, ..., x_n]^T$ and $a = [a_1, a_2, a_3, ..., a_n]^T$. The expression, $a^T x$ is simply the following dot product. Note that I will represent this expression using the function f(x).

$f(x) = a_1 x_1 + a_2 x_2 + ... + a_n x_n$

Upon differentiating with respect to x, I find that the following holds true,

$\frac{\partial f(x)}{\partial x_i} = a_i \implies \nabla f(x) = a$

ii. What is the first derivative of $x^T Ax$ with respect to x? What is the second derivative?

I can assume that for some $i, j \in N$ that $x^T Ax = \sum_{i=1}^{n} a_{ij} x_i x_j$. I want to use the fact that A is a symmetric matrix, thus I will re-index this expression to take advantage of the symmetry.

$\implies = \sum_{i=1}^{n} a_{i1} x_i x_1 + \sum_{j=1}^{n} a_{1j} x_1 x_j + \sum_{i=2}^{n} \sum_{j=2}^{n} a_{aij} x_i x_j$

Upon differentiating, I find the resulting expression

$\frac{\partial f}{\partial x_1} = \sum_{i=1}^{n} a_{i1} x_i + \sum_{j=1}^{n} a_{1j} x_j = \sum_{i=1}^{n} a_{1i} x_i + \sum_{i=1}^{n} a_{1i} x_i = 2 \sum_{i=1}^{n} a_{1i} x_i \implies 2Ax.$

Thus, I know that $\frac{\partial}{\partial x} x^T Ax = 2Ax$.

Upon taking a second derivative, I can see trivially that by using the previously calculated that the following calculation holds true

$\frac{\partial^2}{\partial x^2} x^T Ax \implies \frac{\partial}{\partial x} 2Ax = 2A \implies$ the second derivative is simply $2A$.

(e) (Geometry)

i. Show that the vector w is orthogonal to the line $w^T x + b = 0$. (Hint: Consider two points x_1, x_2 that lie on the line. What is the inner product $w^T (x_1 - x_2)$?)

Using the hint, I will consider two points, x_1, x_2 that lie on a line. I can further write two equations in the form $w^T x + b = 0$, such that

$w^T x_1 + b = 0$

$w^T x_2 + b = 0$

Upon subtracting these two equations, I find that I can derive the following expression

$w^T x_1 + b = 0$

$-w^T x_2 - b = 0$

$\implies w^T x_1 - w^T x_2 = 0 \implies w^T (x_1 - x_2) = 0.$

From this, I can conclude that the inner product of w and $x_1 - x_2$ is zero, which means that the vector w is orthogonal to the vector difference $x_1 - x_2$. Furthermore, because $x_1 - x_2$ is a vector lying along the line defined by $w^T x + b = 0$, this implies that w is orthogonal to any vector lying along the line, therefore orthogonal to the line.

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

First, I will multiply both sides of the equation by $\frac{1}{||w||_2}$

$\implies \frac{w^T}{||w||_2}x + \frac{b}{||w||_2}$, such that the expression, $\frac{w^T}{||w||_2}$ is the unit normal vector and $\frac{b}{||w||_2}$ is the distance from the origin to the line $w^T x + b = 0$.

Furthermore, I can subtract the expression to the right hand side and take the absolute value of both side, such that

$|\frac{w^T}{||w||_2}x| = \frac{|b|}{||w||_2}$ holds true, hence concluding the proof.

Problem 3 (Probability and Statistics)

(a) Consider a sample of data S obtained by flipping a coin five times. $X_i, i \in \{1, ..., 5\}$ 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 $S = (X_1, X_2, X_3, X_4, X_5) = (1, 1, 0, 1, 0)$.

i. What is the sample mean for this data?

By definition of the sample mean, I know that generally,

$\bar{X} = \mu = \frac{1}{n}\sum_{i=1}^{n} X_i, \;\; n = 5$ such that n represents the number of elements in the given set of coin flips.

$\implies \bar{X} = \frac{1}{5}(1 + 1 + 0 + 1 + 0) = \frac{3}{5} = 0.6.$

Thus, the sample mean for this data is $\bar{X} = 0.6$.

ii. What is the unbiased sample variance?

By definition of the unbiased sample variance, I know that generally

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, such that n is the number of elements in the given set and $\bar{X}$ is the previously computed mean, where $\bar{X} = 0.6$.

$\implies s^2 = \frac{1}{5-1}[(1-0.6)^2 + (1-0.6)^2 + (0-0.6)^2 + (1-0.6)^2 + (0-0.6)^2] = 0.3$

Thus, the unbiased sample variance is $s^2 = 0.3$.

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

To compute the probability of obtaining the sample given, $S = (X_1, X_2, X_3, X_4, X_5) = (1, 1, 0, 1, 0)$. I will simply assume that each X_i, for $i \in \{1, ..., 5\}$ occurs sequentially, thus I get the following probability computation

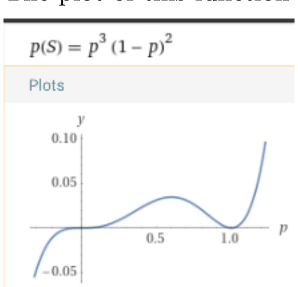
$p(S) = 0.5 * 0.5 * 0.5 * 0.5 * 0.5 = \frac{1}{32} = 0.03125 = 3.125\%$

iv. 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?]

First, I want to write an expression to represent the general probability for the given five element sample

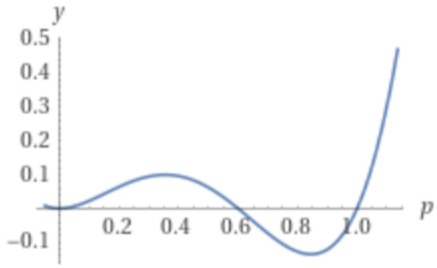
$p(S) = p * p * (1-p) * p * (1-p) = p^3(1-p)^2$

The plot of this function looks like the following:



Now, I want to compute the max of the function, $p(S) = p^3(1-p)^2$ to determine the value that maximizes the probability of the sample S.

$\frac{dp(S)}{dp} = \frac{d}{dp}(p^3(1-p)^2) = 0 \implies 0 = p^2(5p^2 - 8p + 3)$



From this graph, I can see that the plot crosses the x-axis at 0.6 $\implies$ when $p = 0.6$, this means that the likelihood of obtaining $S = (1,1,0,1,0)$ is maximized when $p = 0.6$.

v. 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

First, recall that I can re-write the given joint probability as the following expression

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

I see that $P(X = T \cap Y = b) = 0.1$ and $P(Y = b) = 0.1 + 0.15 = 0.25$

$\implies \frac{0.1}{0.25} = 0.4$ Thus, $P(X = T|Y = b) = 0.4$.

(b) Match the distribution name to its formula.

(a) Gaussian: (v) $\frac{1}{\sqrt{(2\pi)\sigma^2}} exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$

(b) Exponential: (iv) $\lambda e^{\lambda x}$ when $x \geq 0; 0$ otherwise.

(c) Uniform: (ii) $\frac{1}{b-a}$ when $a \leq x \leq b; 0$ otherwise.

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

(e) Binomial: (iii) $\binom{n}{x}p^x(1-p)^{n-x}$

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

Suppose that I flip a coin, such that the probability of head is p. Suppose that there exists a random variable, $Z \sim Bernoulli(p)$, where $Z \in \{0,1\}$, $P(Z = 1) = p$ and $P(Z = 0) = 1 - p$.

Thus, the expectation (mean) is $E(Z) = 0 * (1-p) + 1 * p = p$.

Furthermore, I can compute the variance, such that

$Var(Z) = (0-p)^2 * (1-p) + (1-p)^2 * p$

$= p(1-p)[p + (1-p)] = p(1-p)$.

$E(Z^2) = p$

$\implies Var(Z) = E(Z^2) - E(Z)^2 = p - p^2 = p(1-p)$

Thus, the mean and variance of a $Bernoulli(p)$ random variable are $E(Z) = p$ and $Var(Z) = p(1-p)$.

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

Generally, I know that for some $a, b \in R$ the following linear transformation for the variance hold true.

$Var(aX + b) = E[((aX + b) - E(aX + b))^2] = E[(aX + b - (aE(X) + b))^2] = E[(a(X - E(X)))^2] = a^2 E[(X - E(X))^2] = a^2 Var(X)$

Thus, upon using the general case above, I can compute

$Var(2X) = 2^2 Var(X) = 4Var(X) = 4\sigma^2$.

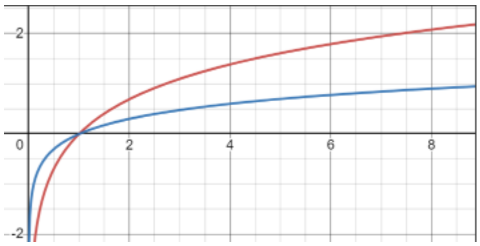
$Var(X + 2) = 1^2 Var(x) = Var(X) = \sigma^2$.

Problem 4 (Algorithms)

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.

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

Note that $f(n) = ln(n)$ is the red curve and $g(n) = log(n)$ is the blue curve.



I can see from the graph that the functions, $f(n) = ln(n)$ and $g(n) = log(n)$ have similar growth rate. I also know that generally by the definition of logarithms that $log_b(n) = \frac{ln(n)}{ln(b)}$ such that $f(n)$ and $g(n)$ have very similar growth rates related which are only different by a constant factor, thus both $f(n) = O(g(n))$, $g(n) = O(f(n))$ hold true.

$O(g(n)) = O(f(n)) = O(log(n)) \implies f(n) = O(g(n))$ and $g(n) = O(f(n))$.

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

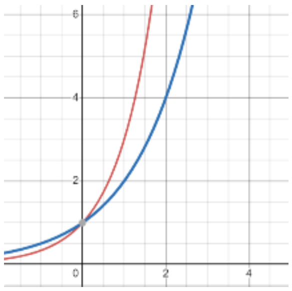
Note that $f(n) = 3^n$ is the red curve and $g(n) = n^{10}$ is the blue curve.

I can see that $f(n) = 3^n$ is an exponential function and $g(n) = n^{10}$ is a polynomial function. I know that by definition, exponential functions grow faster than any polynomial functions. Thus, for large n, i.e. as $n \to \infty \implies 3^n > n^{10}$.

From this, I can conclude that $f(n) = 3^n$ is not $O(n^{10})$ because the exponential function $f(n)$ grows faster than the polynomial $g(n)$. However, $g(n) = n^{10}$ is $O(3^n)$ since a polynomial function like $g(n)$ grows slower than an exponential function like $f(n)$ for large n.

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

Note that $f(n) = 3^n$ is the red curve and $g(n) = 2^n$ is the blue curve.



I can see that both $f(n)$ and $g(n)$ are exponential functions, but they have different bases. I see that $3 > 2 \implies f(n)$ grows faster than $g(n)$ for large n.

Thus, I know that $f(n) = 3^n$ is not in $O(2^n)$. This is because f(n) will grow faster than any constant multiple of $g(n)$. I also see that $g(n) = 2^n$ is in $O(3^n)$. This is because as $n \to \infty$ g(n) will always be bounded above by a constant multiple of $f(n)$.

Problem 5 (Programming Skills)

(a) Sampling from multivariate probability distributions

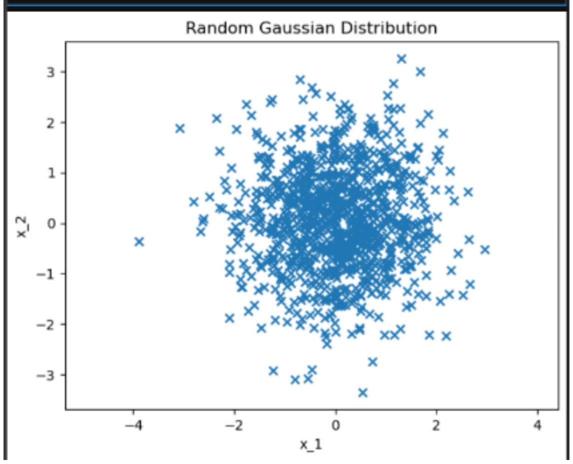
i. Draw 1000 samples $x = (x_1, x_2)^T$ from a 2-dimensional Gaussian distribution with mean $(0, 0)^T$ and identity covariance matrix, i.e. $p(x) = \frac{1}{(2\pi)^2} exp(-\frac{||x||^2}{2})$, and make a scatter plot ($x_1 vs. x_2$).

```
import numpy as np
import matplotlib.pyplot as plt

mean = np.array([0, 0])
cov = np.identity(2)

samples = np.random.multivariate_normal(mean, cov, 1000)

plt.scatter(samples[: , 0], samples[: , 1], marker='x')
plt.axis('equal')
plt.xlabel('x_1')
plt.ylabel('x_2')
plt.title('Random Gaussian Distribution')
plt.show()
```



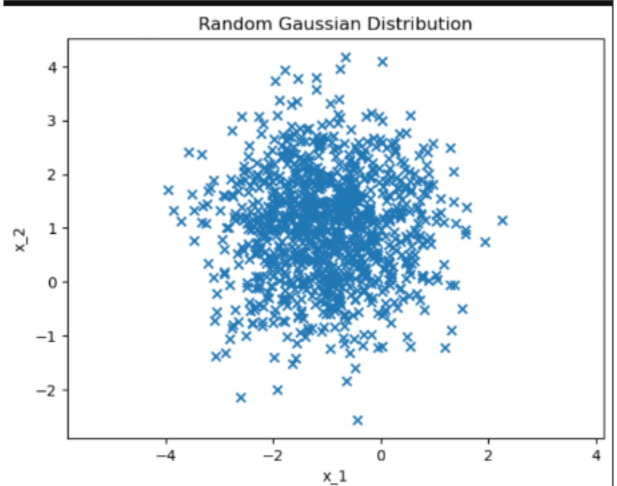
ii. How does the scatter plot change if the mean is $(-1, 1)^T$?

If the mean is $(-1, 1)^T$, then the center of the distribution of points in the scatter plot will shift to the coordinate $(-1, 1)^T$. The overall shape, spread and orientation of the points remains the same.

```python
mean = np.array([-1, 1])
cov = np.identity(2)

samples = np.random.multivariate_normal(mean, cov, 1000)

plt.scatter(samples[:, 0], samples[:, 1], marker='x')
plt.axis('equal')
plt.xlabel('x_1')
plt.ylabel('x_2')
plt.title('Random Gaussian Distribution')
plt.show()
```



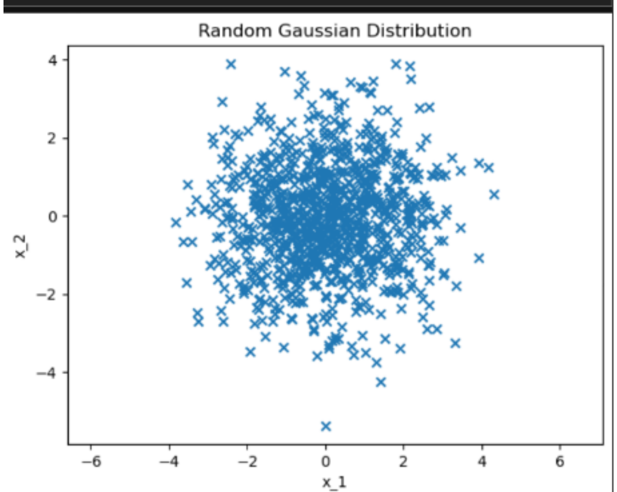
iii. How does the (original) scatter plot change if you double the variance of each component?

When I double the covariance matrix, I find that the distribution of the random points become more spread apart, encompassing more area on the graph.

```python
mean = np.array([0, 0])
cov = [[2, 0] ,[ 0, 2]]

samples = np.random.multivariate_normal(mean, cov, 1000)

plt.scatter(samples[:, 0], samples[:, 1], marker='x')
plt.axis('equal')
plt.xlabel('x_1')
plt.ylabel('x_2')
plt.title('Random Gaussian Distribution')
plt.show()
```



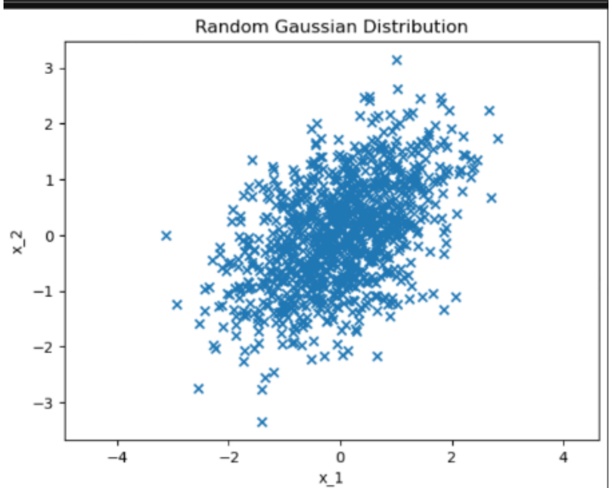
iv. How does the (original) scatter plot change if the covariance matrix is change to $(1, 0.5; 0.5, 1)$?

I notice that when I change the covariance to $(1, 0.5; 0.5, 1)$, the distribution becomes slightly increasing as x_1 and x_2 get large. This results in a slanted or tilted upward position, such that it almost doesn't look completely random, as it takes on more of a positive elliptical shape.

```
mean = np.array([0, 0])
cov = [[1, 0.5] ,[ 0.5, 1]]

samples = np.random.multivariate_normal(mean, cov, 1000)

plt.scatter(samples[:, 0], samples[:, 1], marker='x')
plt.axis('equal')
plt.xlabel('x_1')
plt.ylabel('x_2')
plt.title('Random Gaussian Distribution')
plt.show()
```



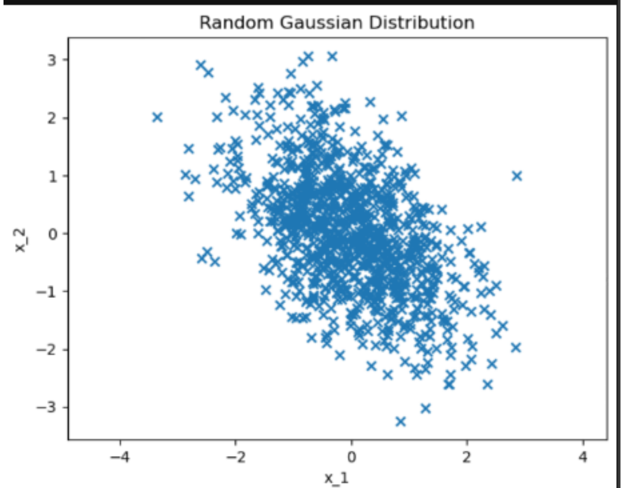
v. How does the (original) scatter plot change if the covariance matrix is changed to $(1, -0.5; -0.5, 1)$

I notice that when I change the covariance to $(1, -0.5; -0.5, 1)$, the distribution becomes slightly decreasing as x_1 and x_2 get large. This results in a slanted or tilted downward position, such that it almost doesn't look completely random, as it takes on more of a negative elliptical shape.

```
mean = np.array([0, 0])
cov = [[1, -0.5] ,[ -0.5, 1]]

samples = np.random.multivariate_normal(mean, cov, 1000)

plt.scatter(samples[:, 0], samples[:, 1], marker='x')
plt.axis('equal')
plt.xlabel('x_1')
plt.ylabel('x_2')
plt.title('Random Gaussian Distribution')
plt.show()
```



(b) 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. For example, you can use one of the open datasets released by the US government here: https://catalog.data.gov/dataset Once you have found the data set, provide the following information:

i. The name of the data set and its link

I used professors recommendation such that through the US Government dataset website, https://catalog.data.gov/datas
I found a very interesting and detailed dataset that talks about the fruit and vegetable prices generally in the US. Here is the specific link to the dataset from the government website: https://catalog.data.gov/dataset/fruit-and-vegetable-prices

Here is the website for the actual government website which contains all of the information about the data set. This website is owned by the Economic Research Service, US Department of Agriculture: https://www.ers.usda.gov/data-products/fruit-and-vegetable-prices.aspx

This website contains a csv which can be downloaded through the blue link titled: ALL FRUITS - Average prices (CSV format).

Lastly, here is the head for the data set.

Fruit	Form	RetailPrice	RetailPrice	Yield	CupEquiva	CupEquiva	CupEquival
Apples	Fresh	1.5193	per pounc	0.9	0.2425	pounds	0.4094
Apples, ap	Canned	1.066	per pounc	1	0.5401	pounds	0.5758
Apples, re	Juice	0.7804	per pint	1	8	fluid ounc	0.3902
Apples, fr	Juice	0.5853	per pint	1	8	fluid ounc	0.2926
Apricots	Fresh	2.9665	per pounc	0.93	0.3638	pounds	1.1603
Apricots, p	Canned	1.6905	per pounc	1	0.5401	pounds	0.9131
Apricots, p	Canned	2.06	per pounc	0.65	0.4409	pounds	1.3974
Apricots	Dried	6.6188	per pounc	1	0.1433	pounds	0.9485
Bananas	Fresh	0.5249	per pounc	0.64	0.3307	pounds	0.2712
Berries, m	Frozen	3.5585	per pounc	1	0.3307	pounds	1.1768
Blackberri	Fresh	6.0172	per pounc	0.96	0.3197	pounds	2.0037
Blackberri	Frozen	3.6362	per pounc	1	0.3307	pounds	1.2025
Blueberrie	Fresh	4.1739	per pounc	0.95	0.3197	pounds	1.4045
Blueberrie	Frozen	3.3898	per pounc	1	0.3307	pounds	1.121
Cantaloup	Fresh	0.5767	per pounc	0.51	0.3748	pounds	0.4238
Cherries	Fresh	3.4269	per pounc	0.92	0.3417	pounds	1.2729
Cherries, p	Canned	4.5257	per pounc	0.65	0.4409	pounds	3.07
Clementin	Fresh	1.3847	per pounc	0.77	0.463	pounds	0.8326
Cranberrie	Dried	4.6513	per pounc	1	0.1232	pounds	0.5729
Dates	Dried	5.5713	per pounc	1	0.1653	pounds	0.9212
Figs	Dried	6.8371	per pounc	0.96	0.1653	pounds	1.1776
Fruit cockt	Canned	1.7198	per pounc	1	0.5401	pounds	0.9289
Fruit cockt	Canned	1.5932	per pounc	0.65	0.4409	pounds	1.0808
Grapefruit	Fresh	1.1695	per pounc	0.49	0.463	pounds	1.105
Grapefruit	Juice	1.0415	per pint	1	8	fluid ounc	0.5208

ii. A brief (2-3 sentence) description of the data set including what the features are and what is being predicted.

This dataset contains the estimated average prices for more than 150 of the most commonly consumed fruits and vegetables. Various grocery stores, supermarkets, super-centers, convenience stores, drug stores and liquor stores across the US provided weekly retail sales data (revenue and quantity). The data set's columns are labeled as the fruit name, the form of the fruit, i.e. fresh, canned, juice, etc. the retail price, the yield, and the cup equivalence. The goal with this data set is to examine the quantity and variety of fruits and vegetables that a household can afford with a limited budget.

iii. The number of examples in the data set.

There are approximately 29 number of fruits, which are; apples, apricots, bananas, berries blackberries, blueberries, cantaloupe, cherries, clementine, cranberries, dates, figs, fruit cocktail, grapefruit, grapes, honeydew, kiwi, mangoes, nectarines, oranges, papaya, peaches, pears, pineapple, plums, pomegranates, raspberries, strawberries and watermelon. Most of these fruits have different variations which are also counted as examples. In total, there are 62 different types of fruits in the data set.

iv. The number of features for each example.

Each of the fruits can have various kinds of features, which include the following; Form: fresh, canned, dried, frozen and juice. Retail Price: numbers within the range $0 \leq retailprice \leq 10.5527$. Yeild: a number within the range $0 \leq yeild \leq 1$. Cup Equivalence Size: a number within the range $CES < 8$. Cup Equivalence Unit: pounds and fluid ounces. Cup Equivalent Price: a number within the range $0 \leq CEP \leq 3.07$.