CM146, Winter 2020
Problem Set 0: Math prerequisites
Due Jan 13, 2019

1 Multivariable Calculus

Solution: By using the product rule, we get:

$$\frac{dy}{dx} = zcos(x)e^{-x} - zsin(x)e^{-x}$$

2 Linear Algebra

(a) **Solution:**

$$y^T z = \begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 1 \cdot 2 + 3 \cdot 3 = \boxed{11}$$

(b) **Solution:**

$$Xy = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \cdot 1 + 4 \cdot 3 \\ 1 \cdot 1 + 2 \cdot 3 \end{pmatrix} = \boxed{\begin{pmatrix} 14 \\ 7 \end{pmatrix}}$$

(c) **Solution:** X is not invertible since its rows are linearly dependent (multiply row 2 by 2 and you get row 1).

(d) **Solution:** The rank of X is 1 since the two rows of the matrix are linearly dependent.

3 Probability and Statistics

(a) **Solution:**

$$\overline{X} = \frac{\sum_{i=1}^{5} X_i}{5} = \frac{1+1+0+1+0}{5} = \boxed{\frac{3}{5}}$$

(b) **Solution:**

$$s^2 = \frac{\sum_{i=1}^{5}(X_i - \overline{X})^2}{4} = \frac{(\frac{2}{5})^2 \cdot 3 + (\frac{3}{5})^2 \cdot 2}{4} = \boxed{0.3}$$

(c) **Solution:** This would be the probability that all of these events occur:

$$P(S) = P(X_1 = 1)P(X_2 = 1)P(X_3 = 0)P(X_4 = 1)P(X_5 = 0)$$
$$= 0.5^5 = \boxed{\frac{1}{32}}$$

(d) **Solution:** In general, the probability of the sample S is:

$$P(S) = P(X_1 = 1)P(X_2 = 1)P(X_3 = 0)P(X_4 = 1)P(X_5 = 0) = p^3(1-p)^2$$

where $p = P(X_i = 1)$. Lets find p that maximizes P(S) by using rules from Calculus:

$$\frac{dP(S)}{dp} = 3p^2(1-p)^2 - 2p^3(1-p)$$
$$= p^2(1-p)[3(1-p) - 2p] = p^2(1-p)(3-5p)$$

We can therefore see that $p = \frac{3}{5}, p = 1, p = 0$ are all local minimum/maximum points. $p = 0$ *and* $p = 1$ makes $P(S) = 0$, and we can see by the sign of the derivative around $p = \frac{3}{5}$ that this is indeed the point that maximizes P(S). Therefore, the solution is $\boxed{\frac{3}{5}}$

(e) **Solution:** By using Bayes rule, we get:

$$P(X = T | Y = b) = \frac{P(X = T, Y = b)}{P(Y = b)} = \frac{0.1}{0.1 + 0.15} = \boxed{0.4}$$

4 Probability Axioms

(a) **Solution:** This is false. For example, if A and B are disjoint and are both events with positive probability, this means that $B \subseteq A^c$. Therefore: $B \cap A^c = B$. But now clearly: $P(A \cup B) \neq P(A \cap B) = 0$

(b) **Solution:** This is false. For example, if A and B are not disjoint and are both events with positive probability, this means that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, but $P(A \cap B) \neq 0$.

(c) **Solution:** This is false. For example, if A and B are different events with different probabilites, we get: $P(A \cap B) + P(A^c \cap B) = P(B) \neq P(A)$.

(d) **Solution:** This is false. From Bayes rule, we know that: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. So, for example, if $P(A) = 1$, $P(B) = 0.5$, we get: $P(A|B) = 2P(B|A)$.

(e) **Solution:** This is true. By using the chain rule (applying Bayes rule 2 times), we get: $P(A_1 \cap A_2 \cap A_3) = P(A_3|(A_2 \cap A_1))P(A_2 \cap A_1) = P(A_3|(A_2 \cap A_1))P(A_2|A_1)P(A_1)$

5 Discrete and Continuous Distributions

Solution:
(a) matches (v)
(b) matches (iv)
(c) matches (ii)
(d) matches (i)
(e) matches (iii)

6 Mean and Variance

(a) **Solution:** Lets calculate the mean and variance of a Bernoulli(p) random variable named X:
$E[X] = p \cdot 1 + (1-p) \cdot 0 = \boxed{p}$
$E[X^2] = p \cdot 1 + (1-p) \cdot 0 = p \Rightarrow Var(X) = E[X^2] - (E[X])^2 = p - p^2 = \boxed{p(1-p)}$

(b) **Solution:**
$Var(2X) = E[(2X)^2] - (E[2X])^2 = \boxed{4Var(X)}$
Shifting a random variable doesn't change the variance:
$Var(X+3) = \boxed{Var(X)}$

7 Algorithms

(a) Big-O notation

i. **Solution:** $ln(n)$ *and* $lg(n)$ are of the same magnitude in the limit. Therefore: $ln(n) = O(lg(n))$ *and* $lg(n) = O(ln(n))$.

ii. **Solution:** Exponential functions are of bigger magnitude than any polynomial. Therefore: $g(n) = O(f(n))$, but not the other way around.

iii. **Solution:**

$$\lim_{n\to\infty} \frac{g(n)}{f(n)} = \lim_{n\to\infty} \frac{2^n}{3^n} = 0$$

Therefore: $g(n) = O(f(n))$, but not the other way around.

(b) **Solution:** Goal property: the element in the desired index is 0, and the element to its right is 1.
Algorithm: Perform binary search on the array, while looking for an element that holds the goal property. If found, return its index in the array. In any iteration throughout the algorithm, if the element observed doesn't hold the goal property: go right if the element is 0, and go left if the element is 1 (at each iteration divide the array to two parts and continue to either the right or left part).
We are performing binary search, so the algorithm runs in $O(log(n))$. The algorithm is correct because if such element exists, at each iteration we are taking out all elements which could not be the solution, leaving the rest to be considered in following iterations. The algorithm terminates when the required index is found, or when we took out of the array all elements which couldn't be the solution. Therefore, the algorithm is correct.

8 Probability and Random Variables

(a) **Solution:** By definition (and by the property of mutual independence):

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy = \int_{-\infty}^{\infty} f_X(x)dx\int_{-\infty}^{\infty} f_Y(y)dy = E[X]E[Y] \blacksquare$$

(b) i. **Solution:** By the law of large numbers, after 6000 trials, 3 shows up on average about the same as its true expectation as a $Bernoulli(\frac{1}{6})$ random variable. Therefore, the number of times 3 shows up is close to 1000, which means $\frac{1}{6}$ on average.

ii. **Solution:** A fair coin which is tossed n times can be described as a sequence of n i.i.d. $Bernoulli(\frac{1}{2})$ random variables, where 1 means head and 0 means tail ($\sigma^2 = \frac{1}{4}$). Therefore, by the central limit theorem, we get:

$$\sqrt{n}\frac{\overline{X}-\frac{1}{2}}{\sigma} = \sqrt{n}\frac{\overline{X}-\frac{1}{2}}{\frac{1}{2}} \longrightarrow_{n\to\infty} \mathcal{N}(0,1)$$

.

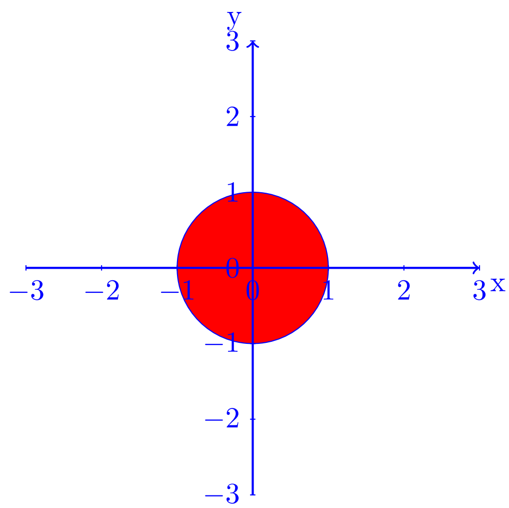
Therefore:

$$\sqrt{n}(\overline{X}-\frac{1}{2}) \longrightarrow_{n\to\infty} \mathcal{N}(0,\frac{1}{4}) \blacksquare$$

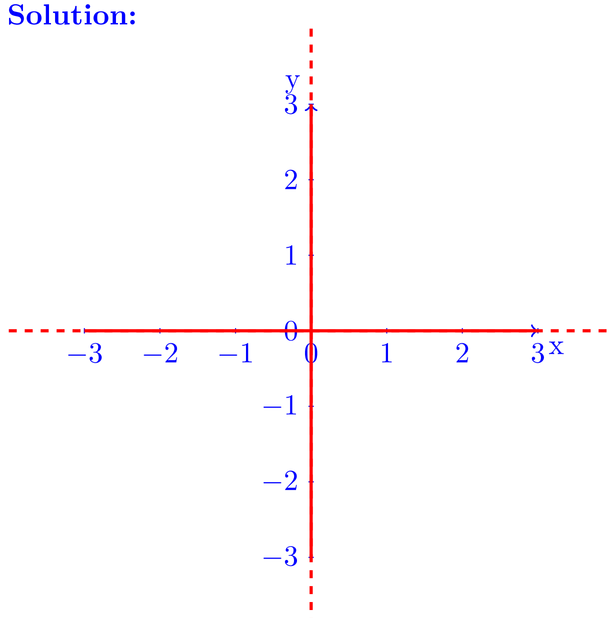
9 Linear Algebra

The regions are marked in red.

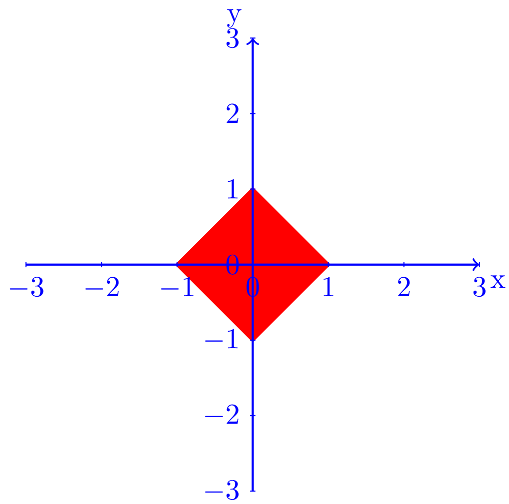
(a) i. **Solution:**



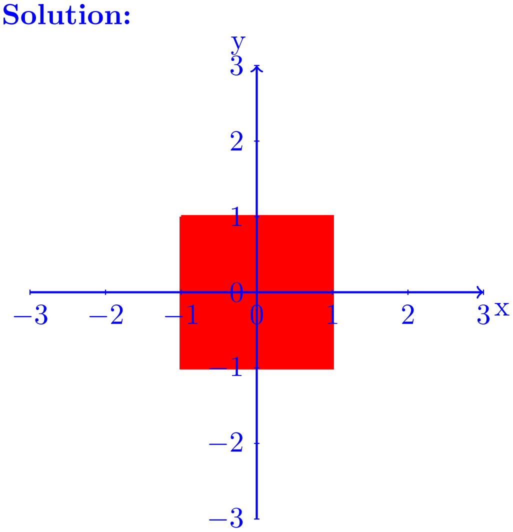
ii. **Solution:**



iii. **Solution:**



iv. **Solution:**



(b) i. **Solution:** Let A be an $n \times n$ square matrix. Let λ be a scalar and v be an $n \times 1$ vector. If $Av = \lambda v$, we say that v is a eigenvector of A, and λ is a eigenvalue of A corresponding to the eigenvector v.

ii. **Solution:** All rows sum to 3, so we conclude that $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector, with $\lambda_1 = 3$ as the corresponding eigenvalue.
In addition, the sum of the diagonal of the matrix equals the

sum of the eigenvalues, so we conclude to have another eigenvalue $\lambda_2 = 1$. From there, it's easy to see that the corresponding eigenvector is $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

In conclusion, the eigenvectors and eigenvalues are:

$$\boxed{\lambda_1 = 3,\ v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

$$\boxed{\lambda_2 = 1,\ v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}}$$

iii. **Solution:** Let v be some eigenvector of A with eigenvalue λ.

$$A^k v = A^{k-1}Av = A^{k-1}\lambda v = \lambda A^{k-2}Av = \lambda A^{k-2}\lambda v = \lambda^2 A^{k-2} v = \cdots = \lambda^k v$$

Therefore, v is an eigenvector of A^k, which corresponds to an eigenvalue λ^k. This is true for every eigenvector of A^k, so the eigenvalues of A^k are $\lambda_1^k, ..., \lambda_n^k$.

(c) i. **Solution:** The first derivative is $\boldsymbol{\alpha}$.

ii. **Solution:** The first derivative is $\boldsymbol{(A + A^T)x}$.
The second derivative is $\boldsymbol{A + A^T}$.

(d) i. **Solution:** Let $\boldsymbol{x_1, x_2}$ be to points on the line $\boldsymbol{w^T x} + b = 0$. Therefore:
$\boldsymbol{w^T(x_1 - x_2) = w^T x_1 - w^T x_2 = w^T x_1 - w^T x_2} + b - b = 0$.
The final equality holds since $\boldsymbol{x_1, x_2}$ are on the above line. The inner product is 0, and so $\boldsymbol{w}$ is orthogonal to the given line.

ii. **Solution:** Let x_0 be a point on the line such that the vector x_0 is orthogonal to the line, and thus parallel to w. By properties from geometry, we know that $||x_0||_2$ is the distance from the origin to the line. Since w and x_0 are parallel, we know that if w and x_0 are pointing to the opposite direction:

$$w^T x_0 = -||w||_2||x_0||_2 = -b$$

. in which case $b \geq 0$. And if they are pointing to the same direction:

$$w^T x_0 = ||w||_2||x_0||_2 = -b$$

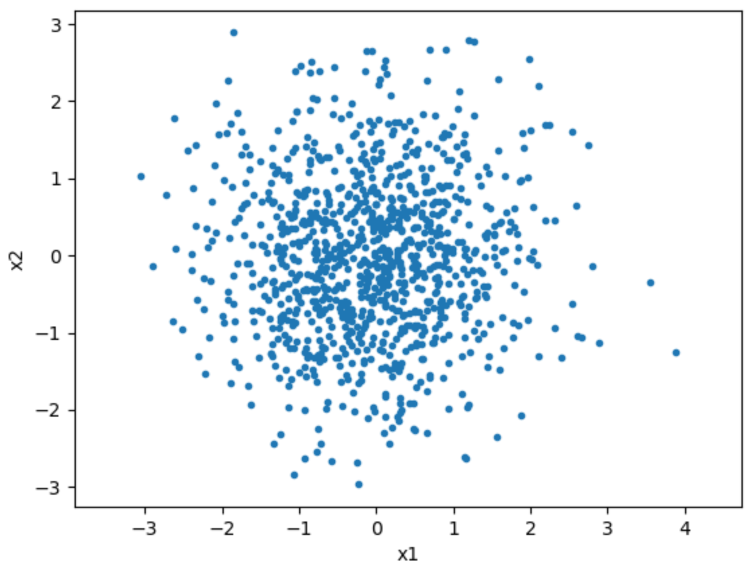
. in which case $b \leq 0$. Therefore, we get:

$$||x_0||_2 = ||w||_2 \frac{||x_0||_2}{||w||_2} = \frac{|b|}{||w||_2}$$

which is also the distance from the origin to the line.

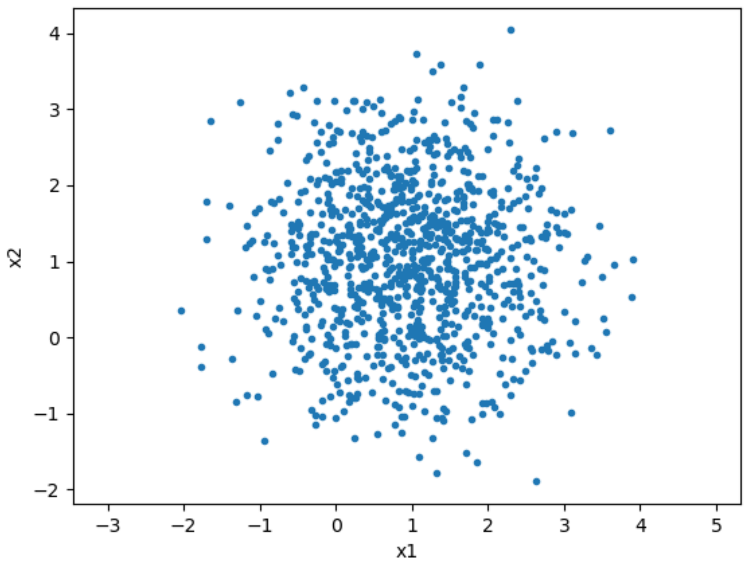
10 Sampling from a Distribution

(a) **Solution:**

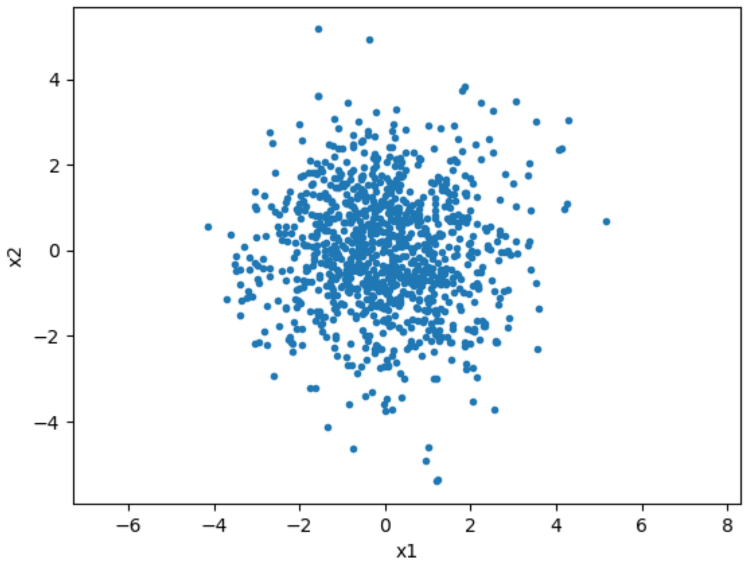


(b) **Solution:**

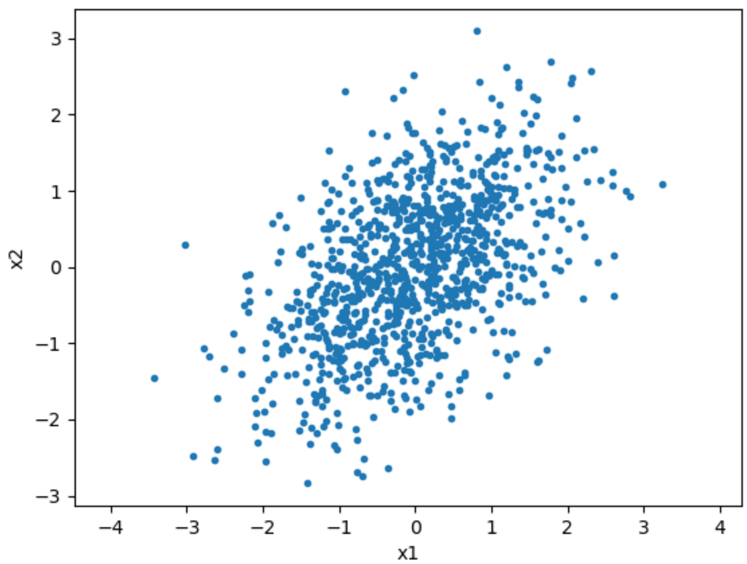
In this case, the dots are centered around $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.



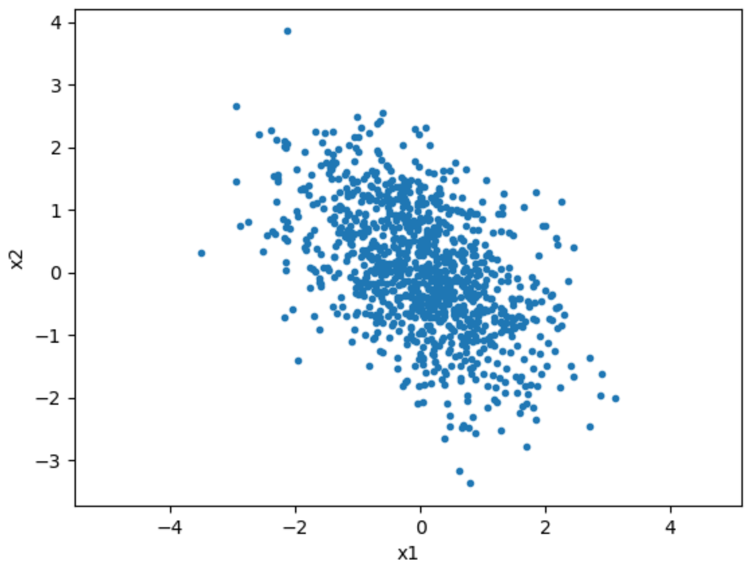
(c) **Solution:**
In this case, the dots are more spreaded.



(d) **Solution:**
In this case, the dots will be twisted.



(e) **Solution:**
In this case, the dots will be twisted, but in the other direction.



11 Eigendecomposition

Solution: The computed eigenvector is $\begin{pmatrix} 0 \\ 0.89442719 \end{pmatrix}$, and its corresponding eigenvalue is 3.

12 Data

Solution:

(a) The name of the data set is: The Olivetti faces dataset.

(b) The data can be obtained from the scikit-learn python library, and is downloaded from AT&T. To fetch the dataset, you can use the function sklearn.datasets.fetch_olivetti_face.

(c) The dataset contains a set of face images, each of which contains 4096 features (each picture is 64x64 pixels). The identity of a the person in the picture is being predicted, out of 40 people in total (40 classes).

(d) The number of examples in the dataset is 400 (a low amount of examples).

(e) The number of features for each example is 4096, which is the dimensionallity of each image.