CM146, Fall 2017
Problem Set 0: Math prequisites
Due Oct 09,2017

Author: Shivraj Gill

1 Problem 1

Solution: Solution to problem 1

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

2 Problem 2

(a) Problem 2a

Solution: Solution to problem 2a

$$y^T z = 11$$

(b) Problem 2b **Solution:** Solution to problem 2b

$$Xy = \begin{pmatrix} 14 \\ 10 \end{pmatrix}$$

(c) Problem 2c **Solution:** Solution to problem 2c

X is invertible if and only if we can reduce it to the identity matrix using reduced row echelon form. Steps to reduce X:

$$X = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} => \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} => \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Because the matrix can be reduced to the identity matrix via row operations, we can say the matrix is invertible.

(d) Problem 2d **Solution:** Solution to problem 2d

The rank of X is 2 because it is invertible i.e the matrix is full rank. In other words, the column vectors of the matrix form a basis of $\mathbb{R}^2$

3 Problem 3

(a) Problem 3a

Solution: Solution to problem 3a

The sample mean is $\frac{3}{5} = 0.6$

(b) Problem 3b **Solution:** Solution to problem 3b

$$\sigma^2 = \frac{\sum_{i=1}^{5}(x_i - 0.6)^2}{5} = \frac{(0.4)^2+(0.4)^2+(0.6)^2+(0.4)^2+(0.6)^2}{5} = \frac{1.2}{5} = 0.24$$

(c) Problem 3c **Solution:** Solution to problem 3c

The probability of observing this data is going to be $\frac{1}{2}^5 = \frac{1}{32}$

(d) Problem 3d **Solution:** Solution to problem 3c

We are trying to solve the following optimization problem:

Let $P(X_i = 1) = p$.

$$\begin{aligned} \underset{p}{\text{maximize}} \quad & f(p) = p^3 * (1-p)^2 \\ \text{subject to} \quad & f(p) \geq 0, \end{aligned}$$

Next we take the derivative of f(p) and set it equal to 0 to get $f'(p) = (p-1)p^2(5p-3) = 0$

p can equal 3 different values, plugging them in, we see that the maximum value is obtained when $p = \frac{3}{5}$.

(e) Problem 3e **Solution:** Solution to problem 3e

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

4 Problem 4

(a) Problem 4a **Solution:** Solution to problem 4a True

(b) Problem 4b **Solution:** Solution to problem 4b True

(c) Problem 4c **Solution:** Solution to problem 4c False

(d) Problem 4d **Solution:** Solution to problem 4d False

(e) Problem 4d **Solution:** Solution to problem 4d True

5 Problem 5

(a) Problem 5a **Solution:** Solution to problem 5a (v)

(b) Problem 5b **Solution:** Solution to problem 5b (iv)

(c) Problem 5c **Solution:** Solution to problem 5c (ii)

(d) Problem 5d **Solution:** Solution to problem 5d (i)

(e) Problem 5e **Solution:** Solution to problem 5e (iii)

6 Problem 6

(a) Problem 6a **Solution:** Solution to problem 6a

Let X be a Bernoulli random variable with probability p. $E(X) = p = \mu$ $V(X) = p(1-p)$

(b) Problem 6b **Solution:** Solution to problem 6b

The variance of 2X would be $4\sigma^2$ if Var(X) = σ^2. And Var(X+2) = σ^2 as well because the variance does not change when adding or subtracting by a constant.

7 Problem 7

(a) Problem 7a

i. Problem 7a i. **Solution:** Solution to problem 7a i.

They are the same because converting from one logarithm base to another involves multiplying by a constant factor.

ii. Problem 7a ii. **Solution:** Solution to problem 7a ii.

Only $f(n) = O(g(n))$ since $3^n \geq n^{10}$ for large n.

iii. Problem 7a iii. **Solution:** Solution to problem 7a iii.

Only $f(n) = O(g(n))$ since $3^n \geq 2^n$ for large n.

(b) Problem 7b **Solution:** Solution to problem 7b

1. Select middle element of List L If element is 0, then L = right half of L and go back to step 1

If element is 1, then L = left half of L and go back to step 1

2. If you have reached the last element of the list and it is a 0, then return index of list. Otherwise, return the element to the left of the list.

Explanation: The algorithm is correct because we keep dividing the list in half until we are left with only one element which is either the last 0 or the first 1 in the list.

Runtime: The runtime of the algorithm is $O(log(n))$ because we keep dividing the list in half, so we only need log(n) steps to find the last 0.

8 Problem 8

(a) Problem 8a **Solution:** Solution to problem 8a

Let's assume X and Y are discrete. (The same logic can be applied if X and Y are continuous)
Then
$E[X] = \sum_j P(X = x_j) * x_j and E[y] = \sum_k P(Y = y_k) * y_k$

Using this we get,
$E[XY] = \sum_{k,j} (P(X = x_j) * x_j * P(Y = y_k) * y_k$
$= \sum_j (P(X = x_j) * x_j * \sum_k (P(Y = y_k) * y_k$
$= E[X] * E[Y]$

(b) Problem 8b

i. Problem 8b i. **Solution:** Solution to problem 8b i.

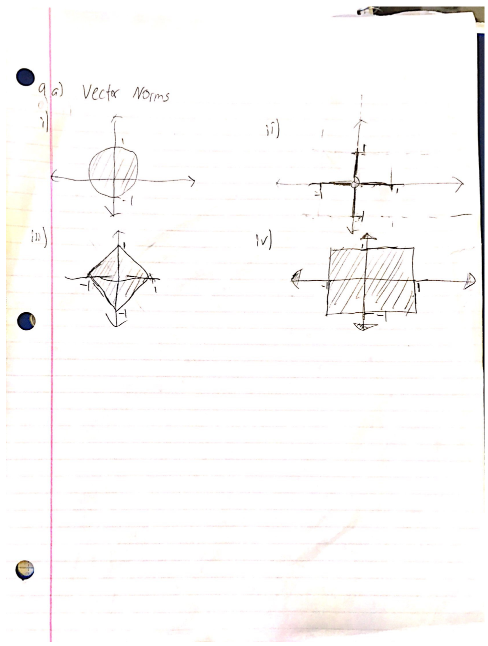
According to the law of large numbers, the average of the results obtained from a large number of trials should be close to the expected value which in our case is $\frac{1}{6} * 6000 = 1000$.

ii. Problem 8b ii. **Solution:** Solution to problem 8b i.

Because each coin toss is an independent Bernoulli Random Variable i.e with a mean and variance, then according to the central limit theorem, $\bar{X}$ follows a normal distribution.

9 Problem 9

(a) Problem 9a



(b) Problem 9b

i. Problem 9b i. **Solution:** Solution to problem 9b i.

   The eigenvalues and eigenvectors of a square matrix A are vectors v and scalars λ such that $Av = \lambda v$

ii. Problem 9b ii. **Solution:** Solution to problem 9b ii.

   $Det(A - \lambda I) = (2 - \lambda)^2 - 1$
   $= \lambda^2 - 4\lambda + 3 = 0$
   $= (\lambda - 3)(\lambda - 1) = 0$

   $= \lambda = 1, 3$

   Now we find the eigenvectors,
   For $\lambda = 1$
   $Av - v = 0$
   1) $v_1 + v_2 = 0$
   2) $v_1 + v_2 = 0$
   so $v_1 = -v_2$
   $v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$
   For $\lambda = 3$
   $Ax - 3x = 0$
   =¿ 1) $-x_1 + x_2 = 0$
   =¿ 2) $x_1 - x_2 = 0$
   =¿ $x_1 = x_2$
   =¿ $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

iii. Problem 9b iii. **Solution:** Solution to problem 9b iii.

   We will prove this by induction. Since this already holds for the base case n = 1, we will assume that this holds for k = n-1. Now let k = n.

$$A^n v = A^{n-1} A v = \lambda A^{n-1} v = \lambda \lambda^{n-1} v = \lambda^n v$$

   Because this holds for k = n, we can say that this holds for all values of n.

(c) Problem 9c

i. Problem 9c i. **Solution:** Solution to problem 9c i.

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

ii. Problem 9c i. **Solution:** Solution to problem 9c i. $\frac{\partial x^T A x}{\partial x} = 2Ax$
$\frac{\partial^2 x^T A x}{\partial x^2} = 2A$

(d) Problem 9d

i. Problem 9d i. **Solution:** Solution to problem 9d i.

Suppose we have two points $x_1 and x_2$ that satisfy the equation of the line. Then,

$$w^T x_1 + b = 0$$

$$w^T x_2 + b = 0$$

By subtracting the two equations, we get

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

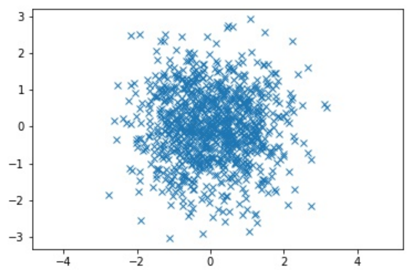
which means w is orthogonal to every point on the line.

ii. Problem 9d ii. **Solution:** Solution to problem 9d ii.

The distance from the origin to the line must be $\frac{b}{|w|}$ because w is normal to the line and hence gives us the distance to the origin.

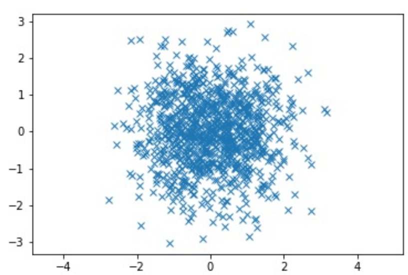
10 Problem 10

(a) Problem 10 a. **Solution:** Problem 10 a.



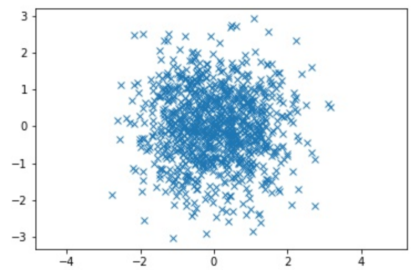
(b) Problem 10 b. **Solution:** Problem 10 b.

Figure 1: The plot is now centered at the point (1,1), but the shape stays roughly around the same.



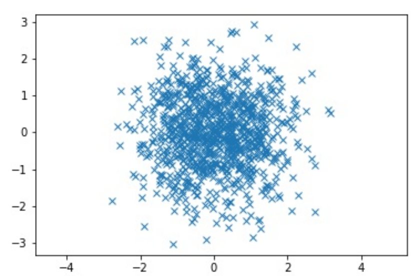
(c) Problem 10 c. **Solution:** Problem 10 c.

Figure 2: The points become more spread out, and hence the shape of the curve covers more area.



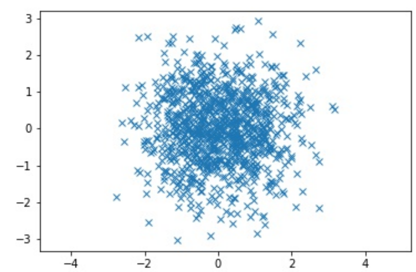
(d) Problem 10 d. **Solution:** Problem 10 d.

Figure 3: The scatter plot represents a line with slope $\frac{1}{2}$. Note that there are more points on the line that are close to the origin because the mean is 0.



(e) Problem 10 e. **Solution:** Problem 10 e.

Figure 4: The scatter plot represents a line with slope -$\frac{1}{2}$. Note that there are more points on the line that are close to the origin because the mean is 0.



11 Problem 11

The solution from running the command np.linalg.eig is v = (0, 1), where v is the eigenvector associated with the largest eigenvalue, $\lambda = 3$.

12 Problem 12

(a) 12 a. **Solution:** 12 a.

The data set is called Titanic dataset.

(b) 12 b. **Solution:** 12 b.

The data can be obtained at https://www.kaggle.com/c/titanic

(c) 12 c. **Solution:** 12 c.

The dataset has the following features: gender, pclass, sex, age, number of siblings, number of parents, ticket number, fare, cabin, and port of embarkation for each passenger. We are predicting whether each passenger survived or not.

(d) 12 d. **Solution:** 12 d.

The training set has 891 examples and the testing set has 418 examples.

(e) 12 e. **Solution:** 12 e.

Gender: M/F (Discrete)
$pclass : Upper/Middle/Lower(Discrete)$
age: 1 - max(age) (Discrete)
$sibsp : 0 - max(numberofsiblings)(Discrete)$
parch: 0 - max(number of parents) (Discrete)
$ticket - 6digitnumber(Discrete)$
fare - Price of Fair (Continuous)
$cabin - String(CategorizedasDiscreteNumber)$
embarked - 1 Letter code (Categorized as Discrete Number)