Raul G. Martinez (PID: A12461871)

DSE 210: Probability and statistics Winter 2020

Worksheet 7 — Linear algebra primer

1. (2 points) Find the unit vector in the same direction as x = (1, 2, 3).

The unit vector has magnitude of one. First find the current magnitude of x:

$$|x| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

Now, let's divide vector x by its current magnitude to find the unit vector **u.**

$$\boldsymbol{u} = \frac{\boldsymbol{x}}{|\boldsymbol{x}|} = \left(\frac{\mathbf{1}}{\sqrt{\mathbf{14}}}, \frac{\mathbf{2}}{\sqrt{\mathbf{14}}}, \frac{\mathbf{3}}{\sqrt{\mathbf{14}}}\right)$$

Finally, verify our unit vector **u** has magnitude of one.

$$|\boldsymbol{u}| = \sqrt{\left(\frac{1}{\sqrt{14}}\right)^2 + \left(\frac{2}{\sqrt{14}}\right)^2 + \left(\frac{3}{\sqrt{14}}\right)^2} = 1$$

2. (2 points) Find all unit vectors in R^2 that are orthogonal to (1, 1).

Define the unit vector of interest as **u** = (1, 1) and all possible orthogonal vectors to **u** as **v** = (v_1, v_2).

All vectors are orthogonal if the dot product is equal to zero < **u**, **v** > = 0, therefore they have to satisfy the following property:

< **u**, **v** > = (1)(v_1) + (1)(v_2) = $v_1 + v_2 = 0$, with solutions v_1 = -1, v_2 = 1 and v_1 = 1, v_2 = -1

Finally, using the solutions v_a = (-1, 1) and v_b = (1, -1) we can find both unit and orthogonal vectors to **u** as:

$$\boldsymbol{v_a} = \left(-\frac{\mathbf{1}}{\sqrt{\mathbf{2}}}, \frac{\mathbf{1}}{\sqrt{\mathbf{2}}}\right) \textit{and } \boldsymbol{v_b} = \left(\frac{\mathbf{1}}{\sqrt{\mathbf{2}}}, -\frac{\mathbf{1}}{\sqrt{\mathbf{2}}}\right)$$

3. (2 points) How would you describe the set of all points $x \in R^d$ with $x \cdot x = 25$?

The dot product of vector x with itself can be described as:

$$< x, x > = x_1 x_1 + x_2 x_2 + x_3 x_3 + \cdots + x_d x_d$$

$$= \boldsymbol{x_1}^{\boldsymbol{2}} + \boldsymbol{x_2}^{\boldsymbol{2}} + \boldsymbol{x_3}^{\boldsymbol{2}} + \cdots + \boldsymbol{x_d}^{\boldsymbol{2}} \boldsymbol{= 25}$$

4. (2 points) The function $f(x) = 2x1 - x2 + 6x3$ can be written as $w \cdot x$ for $x \in R^3$. What is w?

We can re-write f(x) as:

$$f(x) = \boldsymbol{w} \cdot \boldsymbol{x} = \begin{bmatrix} 2 & -1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Therefore,

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{2} & \boldsymbol{-1} & \boldsymbol{6} \end{bmatrix}$$

5. (2 points) For a certain pair of matrices A, B, the product AB has dimension 10×20. If A has 30 columns, what are the dimensions of A and B?

The following property must be satisfied for matrix multiplication:

(m x n) (n x k) equals (m x k), where the order is (row x column)

For AB we have m = 10 and k = 20, and since A has 30 columns we can define n = 30. Therefore the dimensions for A and B are:

A is (10 x 30) and B is (30 x 20)

6. (3 points, 1 each) We have n data points $x^{(1)}, ..., x^{(n)} \in R^d$ and we store them in a matrix X, one point per row.

(a) What is the dimension of X?

We can store n data points (i.e. x1, x2, …, xn) in a matrix, each with d elements (i.e. x1_1, x1_2, …, x1_d), as follows:

$$X = \begin{bmatrix} x1 \\ x2 \\ \vdots \\ xn \end{bmatrix} = \begin{bmatrix} x1\_1 & x1\_2 & \cdots & x1\_d \\ x2\_1 & x2\_2 & \cdots & x2\_d \\ \vdots & \vdots & \ddots & \vdots \\ xn\_1 & xn\_2 & \cdots & xn\_d \end{bmatrix}$$

Therefore, the dimension for X equals to (n x d).

(b) What is the dimension of XX^T ?

Considering the matrix X from (a), the dimension for XX^T is equal to (n x n).

(c) What is the (i, j) entry of XX^T , simply?

Let's first find a few terms for XX^T with matrix multiplication:

$$XX^T = \begin{bmatrix} x1\_1 * x1\_1 + x1\_2 * x1\_2 + \cdots & x1\_1 * x2\_1 + x1\_2 * x2\_2 + \cdots & \cdots \\ x1\_1 * x2\_1 + x1\_2 * x2\_2 + \cdots & x2\_1 * x2\_1 + x2\_2 * x2\_2 + \cdots & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

From the simplification above we can say that the (i, j) entry of XX^T is equal to the inner product for xi and xj datapoints. In other words:

$$\mathbf{XX^T ij = < xi, xj >}$$

7. (2 points) Vector x has length 10. What is $x^T xx^T xx^T x$?

For this special case, use the property:

$$x^T x = \|x\|^2 = 10$$

Then, using the associate properties we can group the expression from above as:

$$\boldsymbol{\|x\|^2 \|x\|^2 \|x\|^2 = 10^2\, 10^2\, 10^2 = 10^6}$$

8. (2points)For x=(1, 3, 5)compute $x^T x$ and xx^T.

Define both x and its transpose as:

$$x = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \quad , \quad x^T = \begin{bmatrix} 1 & 3 & 5 \end{bmatrix}$$

Then,

$$x^T x = \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = (1)(1) + (3)(3) + (5)(5) = \mathbf{35}$$

$$x\,x^T = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} (1)(1) & (1)(3) & (1)(5) \\ (3)(1) & (3)(3) & (3)(5) \\ (5)(1) & (5)(3) & (5)(5) \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{3} & \mathbf{5} \\ \mathbf{3} & \mathbf{9} & \mathbf{15} \\ \mathbf{5} & \mathbf{15} & \mathbf{25} \end{bmatrix}$$

9. (2 points) Vectors $x, y \in R^d$ both have length 2. If $x^T y = 2$, what is the angle between x and y?

We can define the dot product of x^T and y as:

$$x^T \cdot y = \|x^T\| \, \|y\| \, cos\theta$$

Therefore the angle can be calculated as:

$$\theta = cos^{-1}\left(\frac{x^T \cdot y}{\|x^T\| \, \|y\|}\right) = cos^{-1}\left(\frac{2}{(2)(2)}\right) = cos^{-1}\left(\frac{1}{2}\right) = \boldsymbol{60\ degrees}$$

10. (2 points) The quadratic function $f : R^3 \to R$ given by

$$f(x) = 3x^2_1 + 2x1x2 - 4x1x3 + 6x^2_3$$

can be written in the form $x^T M x$ for some **symmetric** matrix M . What is M ?

Let's define x and its transpose as:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad , \quad x^T = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$$

Then, we can re-write the quadratic function as follows:

$$[x_1 \quad x_2 \quad x_3]\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 3{x_1}^2 + 2x_1x_2 - 4x_1x_3 + 6{x_3}^2$$

First matrix multiplication gives:

$$[x_1 \quad x_2 \quad x_3]\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = [ax_1 + dx_2 + gx_3 \quad bx_1 + ex_2 + hx_3 \quad cx_1 + fx_2 + ix_3]$$

Second matrix multiplication gives:

$$[ax_1 + dx_2 + gx_3 \quad bx_1 + ex_2 + hx_3 \quad cx_1 + fx_2 + ix_3]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$
$$= a{x_1}^2 + dx_1x_2 + gx_1x_3 + bx_1x_2 + e{x_2}^2 + hx_2x_3 + cx_1x_3 + fx_2x_3 + i{x_3}^2$$

Using the given quadratic function and the equation from the second matrix multiplication above we can solve for the variables of M:

a = 3, b = 1, c = -2, d = 1, e = 0, f = 0, g = -2, h = 0, i = 6

Therefore, the symmetric matrix for M (where M = M^T) equals to:

$$\mathbf{M} = \begin{bmatrix} \mathbf{3} & \mathbf{1} & \mathbf{-2} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{-2} & \mathbf{0} & \mathbf{6} \end{bmatrix}$$

11. (4 points, 1 each) Which of the following matrices is necessarily symmetric?

(a) AA^T for arbitrary matrix A. **Symmetric.**

(b) $A^T A$ for arbitrary matrix A. **Not Symmetric, when A is a column vector.**

(c) $A + A^T$ for arbitrary square matrix A. **Symmetric.**

(d) $A - A^T$ for arbitrary square matrix A. **Not Symmetric, elements become negative.**

Below is some code in Python I used to verify the statements above in question 11.

```
import numpy as np

x = np.matrix([[1],[-10],[5],[6]])
x_T = np.transpose(x)

a = np.matmul(x,x_T)
a_T = np.transpose(a)

b = np.matmul(x_T,x)
b_T = np.transpose(b)

c = x + x_T
c_T = np.transpose(c)

d = x - x_T
d_T = np.transpose(d)
```

12. (4 points, 2 each) Let A = diag(1, 2, 3, 4, 5, 6, 7, 8).

(a) What is |A|?

The determinant of a diagonal matrix is just the product along the diagonal, therefore:

|A| = (1)(2)(3)(4)(5)(6)(7)(8) = **40,320**

(b) What is A^{-1}?

The inverse of the diagonal matrix A is another diagonal matrix calculated as:

$$A^{-1} = \begin{bmatrix} \frac{1}{1} & & & & & & & \\ & \frac{1}{2} & & & & & & \\ & & \frac{1}{3} & & & & & \\ & & & \frac{1}{4} & & & & \\ & & & & \frac{1}{5} & & & \\ & & & & & \frac{1}{6} & & \\ & & & & & & \frac{1}{7} & \\ & & & & & & & \frac{1}{8} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{1} & & & & & & & \\ & \mathbf{0.5} & & & & & & \\ & & \mathbf{0.33} & & & & & \\ & & & \mathbf{0.25} & & & & \\ & & & & \mathbf{0.2} & & & \\ & & & & & \mathbf{0.17} & & \\ & & & & & & \mathbf{0.14} & \\ & & & & & & & \mathbf{0.13} \end{bmatrix}$$

13. (4 points, 2 each) Vectors $u_1, \ldots, u_d \in R^d$ all have unit length and are orthogonal to each other. Let U be the d × d matrix whose rows are the u_i.

(a) What is UU^T?

Since U is a square and orthogonal matrix with orthogonal unit vectors, we can say that:

$$\mathbf{UU^T = U^TU = I}\text{ , where I is the identity matrix}$$

(b) What is U^{-1}?

Considering the above, and that a matrix is orthogonal if its transpose is equal to its inverse, therefore:

$$\mathbf{U^{-1} = U^T}$$

14. (2 points) Matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & z \end{pmatrix}$ is singular. What is z?

For a singular matrix, which is not invertible, the determinant is zero. Therefore we can calculate z as:

|A| = (1) (z) – (2)(3) = 0 -> z – 6 = 0

$$\mathbf{z = 6}$$

15. **Programming Question**: (10 points) Code following matrix operations in Python:

(a) Find the vector matrix product of A and M (where A is a vector and M is a matrix). where,

A = [1, -1, 0] ; M = [[1, 2, 3], [4, 5, 6], [7, 8, 9]]

```
# Find the vector matrix product of A and M (where A is a vector and M is a matrix)

A = np.matrix([1,-1,0])
M = np.matrix([[1, 2, 3], [4, 5, 6], [7, 8, 9]])

AM = np.dot(A,M)

print("A shape :\t",A.shape)
print("M shape :\t",M.shape)
print("AM shape :\t",AM.shape)
print("AM :\t",AM)
```

```
A shape :        (1, 3)
M shape :        (3, 3)
AM shape :       (1, 3)
AM :     [[-3 -3 -3]]
```

(b) Find the matrix product of 3 matrices A, B and C (i.e. A dot B dot C). where,

A = [[1, 2]] ; B = [[2, 3, 4], [5, 6, 7]] ; C = [[-1, 1, -1, 1], [0, 0, 0, 0], [1, 1, 1, 1]]

```
# Find the matrix product of 3 matrices A, B and C (i.e. A_dot_B_dot_C)

A = np.matrix([1, 2])
B = np.matrix([[2,3,4],[5,6,7]])
C = np.matrix([[-1, 1, -1, 1], [0, 0, 0, 0], [1, 1, 1, 1]])

A_dot_B = np.dot(A,B)
A_dot_B_dot_C = np.dot(A_dot_B,C)

print("A shape :\t",A.shape)
print("B shape :\t",B.shape)
print("C shape :\t",C.shape)
print("A_dot_B shape :\t",A_dot_B.shape)
print("A_dot_B_dot_C shape :\t",A_dot_B_dot_C.shape)
print("A_dot_B :\t",A_dot_B)
print("A_dot_B_dot_C :\t",A_dot_B_dot_C)
```

```
A shape :        (1, 2)
B shape :        (2, 3)
C shape :        (3, 4)
A_dot_B shape :  (1, 3)
A_dot_B_dot_C shape :    (1, 4)
A_dot_B :        [[12 15 18]]
A_dot_B_dot_C :  [[ 6 30  6 30]]
```

16. (5 points) A man has two possible moods: happy and sad. The prior probabilities of these are:

$$\pi(happy) = \frac{3}{4}\,, \quad \pi(sad) = \frac{1}{4}$$

His wife can usually judge his mood by how talkative he is. After much observation, she has noticed that:

- When he is happy,

$$Pr(talks\ a\ lot) = \frac{2}{3}\,, \quad Pr(talks\ a\ little) = \frac{1}{6}\,, \quad Pr(completely\ silent) = \frac{1}{6}$$

- When he is sad,

$$Pr(talks\ a\ lot) = \frac{1}{6}\,, \quad Pr(talks\ a\ little) = \frac{1}{6}\,, \quad Pr(completely\ silent) = \frac{2}{3}$$

(a) Tonight, the man is just talking a little. What is his most likely mood?

We can define the features and labels as:

labels = [happy, sad]; features = [talks a lot, talks a little, completely silent]

Let's first start by finding the Pr(talks a little):

$$\begin{aligned}\Pr(talks\ a\ little)& \\ &= \Pr(talks\ a\ little|\ happy)\Pr(happy) \\ &+ \Pr(talks\ a\ little|\ sad)\Pr(sad)\end{aligned}$$

$$\Pr(talks\ a\ little) = \left(\frac{1}{6}\right)\left(\frac{3}{4}\right)+\left(\frac{1}{6}\right)\left(\frac{1}{4}\right)=\frac{3}{24}+\frac{1}{24}=\frac{1}{6}$$

Now, using Naive Bayes Theorem we can find the probability for each label given the feature:

$$\Pr(label \mid feature) = \frac{\Pr(label)\Pr(feature \mid label)}{\Pr(feature)}$$

$$\Pr(happy|\ talks\ a\ little) = \frac{\Pr(happy)\Pr(talks\ a\ little \mid happy)}{\Pr(talks\ a\ little)} = \frac{\frac{3}{4} * \frac{1}{6}}{\frac{1}{6}} = \mathbf{\frac{3}{4}}$$

$$\Pr(sad|\ talks\ a\ little) = \frac{\Pr(sad)\Pr(talks\ a\ little \mid sad)}{\Pr(talks\ a\ little)} = \frac{\frac{1}{4} * \frac{1}{6}}{\frac{1}{6}} = \mathbf{\frac{1}{4}}$$

According to the calculations above, **his most likely mood is happy with probability of 3/4.**

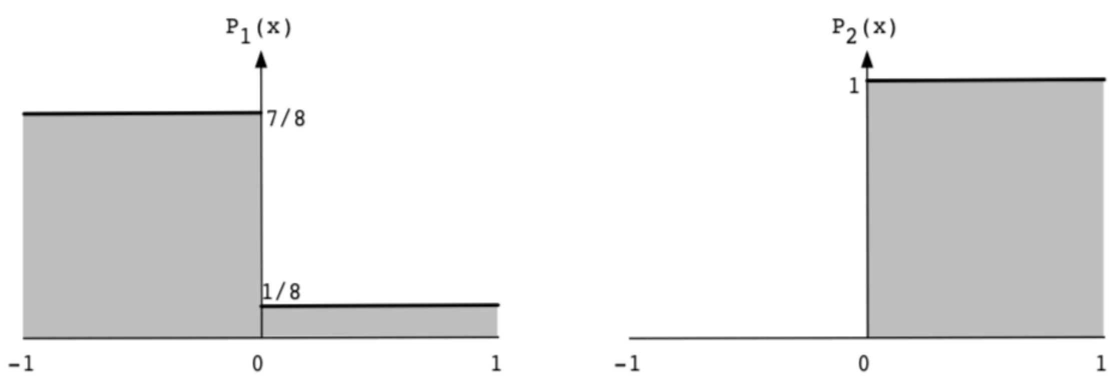
(b) What is the probability of the prediction in part (a) being incorrect?

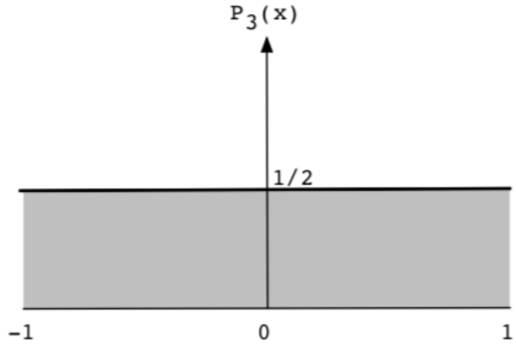
The probability of part (a) being incorrect is the complement for the probability of being happy given that he talks a little, in other words, it should be equal to the probability of being sad which is **1/4.**

17. (5 points) Suppose X = [−1, 1] and Y = {1, 2, 3}, and that the individual classes have weights

$$\pi_1 = \frac{1}{3}, \quad \pi_2 = \frac{1}{6}, \quad \pi_3 = \frac{1}{2}$$

and densities P_1 , P_2 , P_3 as shown below.





What is the optimal classifier h^*? Specify it exactly, as a function from X to Y.

Find the probabilities for each classifier by splitting x from -1 to 0 and 0 to 1:

When Y = 1,

$$\pi_1 P_1(-1 \leq X < 0) = \frac{1}{3} * \frac{7}{8} = \mathbf{\frac{7}{24}}$$

$$\pi_1 P_1(0 < X \leq 1) = \frac{1}{3} * \frac{1}{8} = \frac{1}{24}$$

When Y = 2,

$$\pi_2 P_2(0 < X \leq 1) = \frac{1}{6} * 1 = \frac{1}{6}$$

When Y = 3,

$$\pi_3 P_3(-1 \leq X < 0) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$\pi_3 P_3(0 < X \leq 1) = \frac{1}{2} * \frac{1}{2} = \mathbf{\frac{1}{4}}$$

Based on the Bayes-optimal prediction $h^*(X) = \arg max_j\, \pi_j P_j(X)$ and the probabilities from above, the optimal classifier is:

Class Y = 1 for the range x = [-1, 0) with probability: $\boldsymbol{\pi_1 P_1(-1 \leq X < 0) = \frac{7}{24}}$

Class Y = 3 for the range x = (0, 1] with probability: $\boldsymbol{\pi_3 P_3(0 < X \leq 1) = \frac{1}{4}}$