Name (Last, First):

Vamsi Krishna Poluru

NEU ID:

002924538

Assignment 11

NEU_COE_INFO6105_39506_Spring2024

Please type or handwrite your answers clearly and concisely. Show all work - credit will not be given for numerical solutions that appear without explanation. **[Total 45 = 3 + 42 pts]**

Grading Rubric

Each question is worth 3 points and will be graded as follows:

3 points: Correct answer with work shown

2 points: Incorrect answer but attempt shows some understanding (work shown)

1 point: Incorrect answer but an attempt was made (work shown)

0 points: Left blank or made little to no effort/work not shown

Reflective Journal [3 pts]
(link to your live google doc)

https://docs.google.com/spreadsheets/d/1EVHnjK7z43Ukh_6oYITPlyLyrARNQ5_js8WYGUboyPY/edit#gid=0

I. Multiple Regression (8 x 3 = 24 pts)

Billy's aunt owns a jewellery store, and gives him data on 5000 of the diamonds in her store. For each diamond, we have:

- **carat**: the weight of the diamond, in carats
- **length**: the length of the diamond, in centimeters
- **width**: the width of the diamond, in centimeters
- **price**: the value of the diamond, in dollars

The first 5 rows of the 5000-row dataset are shown below:

carat	length	width	price
0.40	4.81	4.76	1323
1.04	6.58	6.53	5102
0.40	4.74	4.76	696
0.40	4.67	4.65	798
0.50	4.90	4.95	987

Billy has enlisted our help in predicting the price of a diamond given various other features.

1. Suppose we want to fit a linear prediction rule that uses two features, carat and length, to predict price. Specifically, our prediction rule will be of the form

$$\text{predicted price} = w_0 + w_1 \cdot \text{carat} + w_2 \cdot \text{length}$$

We will use least squares to find $\vec{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \\ w_2^* \end{bmatrix}$.

Write out the first 5 rows of the design matrix, X. Your matrix should not have any variables in it.

Answer: (3 pts)

The first 5 rows of the design matrix

1 0.40 4.81
1 1.04 6.58
1 0.40 4.74
1 0.40 4.67
1 0.50 4.90

2. Suppose the optimal parameter vector $\vec{w}^*$ is given by

$$\vec{w}^* = \begin{bmatrix} 2000 \\ 10000 \\ -1000 \end{bmatrix}$$

What is the predicted price of a diamond with 0.65 carats and a length of 4 centimeters? Show your work.

Answer: (3 pts)

predicted price=2000+10000·0.65−1000·4
predicted price=2000+6500−4000
predicted price=4500
Therefore, the predicted price of the diamond is $4500.

3. Suppose $\vec{e} = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$ is the error/residual vector, defined as

$$\vec{e} = \vec{y} - X\vec{w}^*$$

where $\vec{y}$ is the observation vector containing the prices for each diamond.

For each of the following quantities, state whether they are guaranteed to be equal to 0 the scalar, $\vec{0}$ the vector of all 0s, or neither. No justification is necessary.

- $\sum_{i=1}^{n} e_i$
- $||\vec{y} - X\vec{w}^*||^2$
- $X^T X\vec{w}^*$
- $2X^T X\vec{w}^* - 2X^T\vec{y}$

Answer: (12 pts)

1. In ordinary least squares regression, the sum of residuals (errors) is zero because the line of best fit is calculated to minimize these residuals. Hence, this sum is guaranteed to be zero.

2. This expression represents the sum of squared residuals, which is the objective function that ordinary least squares regression seeks to minimize, not necessarily to zero, because there can be an inherent error or noise in the data that cannot be reduced to zero by a linear model.

3. This is part of the normal equation used to solve for the optimal parameters in a linear regression. It is not guaranteed to be zero; instead, it is equal to \(X^T \mathbf{\hat{y}} \) when solved for the optimal weights.

4. This expression can be derived from taking the gradient of the sum of squared residuals with respect to the weights. At the optimal solution (where the gradient is zero), this expression is guaranteed to be zero because it is equal to the gradient of the loss function at the optimum.

1. guaranteed to be zero.
2. is not guaranteed to be zero.
3. is not guaranteed to be zero.
4. is guaranteed to be zero.

4. Suppose we introduce two more features:

- width alone, and
- area, which is defined as length times width

Suppose we also decide to remove the intercept term of our prediction rule. With all of these changes, our prediction rule is now

$$\text{predicted price} = w_1 \cdot \text{carat} + w_2 \cdot \text{length} + w_3 \cdot \text{width} + w_4 \cdot (\text{length} \cdot \text{width})$$

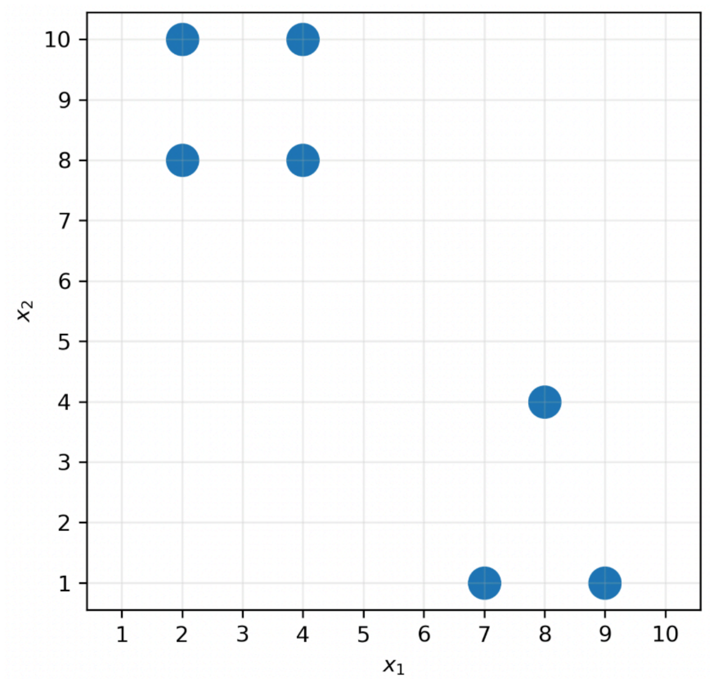
- Write out just the first 2 rows of the design matrix X for this new prediction rule. You do **not** need to simplify the numbers in your matrix, it is fine if they involve the multiplication symbol.
- Is the optimal coefficient for carat, w_1^*, for this new prediction rule guaranteed to be equal to 10000, the optimal coefficient for carat in our original prediction rule? No justification is necessary.

Answer: (6 pts)

First diamond: carat = 0.40, length = 4.81, width = 4.76
Second diamond: carat = 1.04, length = 6.58, width = 6.53
The first 2 rows of the design matrix X are:=
[0.40 4.81 4.76 (4.81×4.76)
1.04 6.58 6.53 (6.58×6.53)]
2nd answer , in this new prediction rule is not guaranteed to be equal to 10000, the optimal coefficient for carat in the original prediction rule. Changing the model by adding new features (like width and area) and removing the intercept term can significantly alter the relationship between the features and the target variable, thereby affecting the optimal coefficients.

II. Clustering (2 x 3 = 6 pts)

Consider the following dataset of 7 points.



Suppose we want to identify $k = 2$ clusters in this dataset using k-Means Clustering.

1.

Determine the centroids $\vec{\mu}_1$ and $\vec{\mu}_2$ that minimize inertia. (Let $\vec{\mu}_1$ be the centroid with a smaller x_1 coordinate.) Justify your answers.

Note: You don't have to run the k-Means Clustering algorithm to answer this question.

Answer: (3 pts)

Cluster 1 (lower cluster): (7,1), (8,4), (9,1)
Cluster 2 (upper cluster): (2,8), (2,10), (4,8), (4,10)
μ 1=((7+8+9)/3,(1+4+1)/3)=(8,2)
μ 2=((2+2+4+4)/4,(8+10+8+10)/4)=(3,9)

2.

What is the total inertia for the centroids you chose in part (a)? Show your work.

Answer: (3 pts)

Inertia =∑(distance from each point to µ)
For Cluster 1 (8,2)
Inertia Cluster 1=(7–8)^ 2+(1–2) ^2 +(8–8)^2 +(4–2)^2 +(9–8)^2 +(1–2)^2=8
Inertia Cluster 2 =(2–3)^2 +(8–9)^2 +(2–3)^2 +(10–9)^2 +(4–3)^2 +(8–9)^2 +(4–3)^2 +(10–9)^2=8

Total Inertia =8+8=16

III. Classification (4 x 3 = 12 pts)

Below, we have a (real) sample of 12 passengers on the RMS Titanic, a British cruise ship that tragically sank in 1912. For each passenger in our sample, we have the following information:

- **survived**: 1 if they survived the sinking of the Titanic, 0 if they did not
- **sex**: either Male or Female
- **class**: the cabin class they were in, either First, Second, or Third (with First being the most luxurious)
- **age**: either child (for those under 18) or adult (for those 18 and older)

survived	sex	class	age
0	female	Third	adult
0	male	First	adult
0	male	Third	adult
1	female	First	adult
1	female	First	adult
1	female	First	child
1	female	Second	adult
1	female	Second	adult
1	female	Third	adult
1	female	Third	child
1	male	First	child
1	male	Second	adult

1. Selma Augusta Emilia was a 38 year old female in Third class. Using the Naive Bayes classification algorithm with no smoothing, predict whether Selma Augusta Emilia survived the sinking of the Titanic. You must show all intermediate steps, including the probabilities you used to make your classification, and you must clearly state what your prediction is.

Answer: (9 pts)

P(Survived)= 9/12
P(Not Survived)= 3/12
P(Female|Survived)= 7/9
P(Third Class|Survived)= 2/9
P(Adult|Survived)=6/9
P(Female|Not Survived)= 1/3
P(Third Class|Not Survived)=2/3
P(Adult|Not Survived)=1

P(Survived|Female, Third Class, Adult) ∝ 7/9* 2/9* 6/9*9/12=7/81
P(Not Survived|Female, Third Class, Adult) 1/3*2/3*1*1/4=1/18
14/81> 1/18

Naive Bayes classifier, we predict that Selma Augusta Emilia would have survived.

2. If you try to use Naive Bayes to classify whether Selma Augusta Emilia's 8 year old daughter (who also sat with her in Third class) survived the Titanic, one of the two probabilities that you'd compare would turn out to be 0.

Using smoothing, compute $P(\text{child} \mid \text{survived})$ and $P(\text{child} \mid \text{didn't survive})$. Show your work. Note that all you need to do is find these two probabilities; you don't need to make any predictions.

Answer: (3 pts)

P(child|survived)= Count of survived children+1/
Total count of survivors+Number of unique classes in age

P(child|didn't survive)= Count of non-survived children+1/
Total count of non-survivors+Number of unique classes in age

P(child|survived)=3+1/9+1=2/5
P(child|not survived)= 0+1/3+1 = 1/4