CSCI 183
Fall 2024
Quiz 1
10/11/2024
Time Limit: 30 Minutes

Name: Arman Mili
SCU ID: ____________

This exam contains 4 pages (including this cover page) and 8 questions.
Total of points is 15.

Grade Table (for teacher use only)

Question	Points	Score
1	1	
2	2	
3	2	
4	2	
5	2	
6	2	
7	2	
8	2	
Total:	15	

1. (1 point) You are conducting a chi-square test to determine if there is an association between a customer's preferred payment method (cash, credit card, or digital wallet) and their age group (under 30, 30-50, over 50) in a retail store. After performing the test, you fail to accept the null hypothesis. What does this result imply?

this means that there was unable to find a relationship between items → High chi = relationship thus the values gotten weren't what was expected

2. (2 points) You have a dataset where all the corr() values of the features and the target are very close to zero. You apply linear regression and your model converges. Is this possible? Explain your answer.

this is possible if the values are very close however if the model converges it's a sign of overfitting to what it trains on. thus "reducing ~~accuracy~~ Precision" of the results, but increasing accuracy. Corr() = 1 is best

3. (2 points) You are using gradient descent for a linear regression model. You initialize the weights to large random values (1000 and 500) and learning rate to 0.001. Will this impact the speed of convergence? Explain your answer.

yes the smaller the learning rate the slower the gradient converges $\theta_j = \theta - \alpha \frac{\partial}{\partial \theta_j}(\theta)$ → the learning rate, will change the weight of the change onto the gradient the smaller the value the slower the change

4. (2 points) Do you think feature scaling is important for gradient descent? If yes, why? If not, why not?

No they would best be fit when trying to find a relationship like in KNN and SVM Algorithms Gradiand decent is an optimization algorithm itself thus this features are not bennefical

5. (2 points) A dataset has the following columns:

   Country (USA, Canada, Mexico)

   Gender (Male, Female)

   Name

   Annual Spending (in USD)

   You want to use **One Hot Encoding:**. How many minimum new columns should be generated? Explain your answer.

4 one gender then 3 (USA, Canada, Mexico) which each have a 1 for true 0 false

G	USA	Canada	Mex
M	1	0	0
F	0	1	0

6. (2 points) **For each of the parts below, which is the best visualization technique to use based on the plots discussed in class? Explain your answer**

   Quant Numerical

   (a) You're a professor teaching Data Science with Python, and you have the grades of the students recorded. You want to visually assess your exam grades {A,B,C,D and F}. Histogram or bar plot to show range, avg, etc. Scatter plot may be good to see changes in relation

   Quant Numerical

   (b) Suppose you own a restaurant and you have recorded the sales (in USD) for each day. You want to visualize if your sales follow a particular distribution.

use a scatter plot or heatmap to plot time, sale, as 2 values to see if there is a distribution.

If just sales and day then ~~Histogram or~~ Bar plot will do

7. (2 points) Regression between foot length (target variable in cm) and height (feature variable in inches) for 33 students resulted in the following regression equation:

   $\hat{y} = 10.9 + 0.23\ x$

   use height

   (a) One student in the sample was 73 inches tall with a foot length of 29cm. What is the predicted foot length for this student? $10.9 + .23 \times 73 = 16.79$ cm

   (b) What would be the error for the example in (a)?

$\frac{1}{m}(\hat{y} - y)^2 = \perp$

$(16.79 - 29) = \frac{-12.21^2}{m} =$

$\frac{1}{m}\sum_{i=1}^{m}(\hat{y}^i - y^i)^2$

5. (2 points) A dataset has the following columns:

   Country (USA, Canada, Mexico)

   Gender (Male, Female)

   Name

   Annual Spending (in USD)

   You want to use **One Hot Encoding:**. How many minimum new columns should be generated? Explain your answer.

4 one gender Then 3 (USA, Canada, Mexico)

which each have a 1 for true 0 false

G	USA	Canada	Mex
M	1	0	0
F	0	1	0

6. (2 points) **For each of the parts below, which is the best visualization technique to use based on the plots discussed in class? Explain your answer**

   Quant Numerical

   (a) You're a professor teaching Data Science with Python, and you have the grades of the students recorded. You want to visually assess your exam grades {A,B,C,D and F}. Histogram or bar plot to show range, avg, etc. Scatter plot may be good to see changes in relation

   Quant Numerical

   (b) Suppose you own a restaurant and you have recorded the sales (in USD) for each day. You want to visualize if your sales follow a particular distribution.

use a scatter plot or heatmap to plot time, sale, as 2 values to see if there is a distribution.

If just sales and day then ~~Histogram or~~ Bar plot will do

7. (2 points) Regression between foot length (target variable in cm) and height (feature variable in inches) for 33 students resulted in the following regression equation:
   $\hat{y} = 10.9 + 0.23\,x$

   use height → (a) One student in the sample was 73 inches tall with a foot length of 29cm. What is the predicted foot length for this student? $10.9 + .23 \times 73 = 16.79\,cm$

   (b) What would be the error for the example in (a)?

$\frac{1}{m}(\hat{y}-y)^2 = \perp$

$(16.79 - 29) = \frac{-12.21^2}{m} =$

$\frac{1}{m}\sum_{i=1}^{m}(\hat{y}^i - y^i)^2$

8. (2 points) **Linear Regression**
Assume that the hypothesis function is $h_\theta(x) = \theta_1 x + \theta_2$. In the standard linear regression case, given an instance $(x^{(i)}, y^{(i)})$ on the training set, the loss function is defined as $L_i(\theta_1, \theta_2) = (h_\theta(x^{(i)}) - y^{(i)})^2$. Imagine that we are still interested in posing the optimization problem (over the dataset) as:

$$min_{(\theta_1,\theta_2)} \sum_{i=1}^{m} L_i(\theta_1, \theta_2)$$

and want to use gradient descent to minimize the function. What if we were to use some slightly different loss functions?

(a) $L_i(\theta_1, \theta_2) = e^{|(h_\theta(x) - y_i)|}$

(b) $L_i(\theta_1, \theta_2) = \lceil (h_\theta(x) - y_i)^2 \rceil$ (ceiling function)

Please answer exactly why these loss functions may or may not be a good choice for regression [0.5 point for each]. Justify your answer [0.5 points for each].

a) will result in a much ~~smaller~~ larger overall loss $e^5 > 5^2$

thus this will result in the gradient being much smaller, and due to the fast nature of c^x will result in a fast ramp to the optimal answer but since fast and large jumps may be inaccurate. ABS value may be negative

b) optimal choice since scales to max and due to not scaling too fast will be able to find the optimal solution. Same as the "normal" loss we have used. Also will take longer than the e^x variant, the non inclusion of the ABS value may be important as well