**Text

Description automatically generated**

**WQD7006 MACHINE LEARNING FOR DATA SCIENCE**

**1/2023/2024**

**Final Exam Part 1**

**Name : Soh Pei Lin**

**Matrix Number : S2193368**

**Group : 1**

**Question 1**

**(a)**

The general linear regression model can be written as:

Where

is the constant term or the y-intercept.

is the coefficient for feature *.*

is the coefficient for feature *.*

is the predicted target variable.

**(b)**

For example 1:

For example 2:

For example 3:

**(c)**

For training data (2 example):

For testing data (1 example):

**(d) (i)**

Logistic regression model is better for the condition. The graph shows consistent with the x1 and x2 axis and this model use to predicts the values between 0 and 1. Linear regression is use to predict any values including continuous value.

**(d) (ii)**

Logistic regression classifier used to forms with linear decision surface. According to Figure 1, it is noticeable that the data points is linear separable, hence it could be conclude that they can be perfectly classified.

**(d) (iii)**

If we change the label of point (0,1) from “+” to ”-”, the data points within Figure 1 will be not linearly separable. In such condition, we cannot perfectly classify the examples by using logistic regression classifier.

**Question 2**

**(a)**

Classifier C has the best generalization performance among the four Classifier. This is because Classifier C has the has the lowest absolute error rate on testing data (15%), which is the actual measure of generalization performance.

**(b)**

Classifier A appears to be most underfitting classifier. Classifier A has a relatively high error rate on the training data (25%), which is the highest among the classifiers, indicating that it may not be capturing the training data patterns well.

**(c)**

Classifier B appear as most overfitting classifier. Classifier D has a larger gap difference between training error (5%) and testing error (20%), suggesting it performs well on the training data but poorly on the testing data, which is indicate of overfitting issue.

**(d)**

When using linear regression, if the training set size is increased, which means that more data are input, the model must account with more variation and cannot fit the training data as closely as it could with a smaller dataset. Hence, the mean training error will increase. As for the mean testing error, it will likely decrease because the model is trained with more data, giving it more information to generalize, which usually results in better performance on unseen data.

**Question 3**

**(a)**

Initial Distance Matrix:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| A | 0 | 1 | 4 | 5 |
| B |  | 0 | 2 | 6 |
| C |  |  | 0 | 3 |
| D |  |  |  | 0 |

1. Step A: Find the shortest distance in the matrix.

The shortest distance is 1 (between A and B). Merge clusters A and B into cluster AB.

|  |  |  |  |
| --- | --- | --- | --- |
|  | AB | C | D |
| A | 0 | 3 | 3 |
| B |  | 0 | 2 |
| C |  |  | 0 |

1. Step b:

Now the closest clusters are AB and C with a distance of 2.

Merge AB and C into a new cluster called ABC

|  |  |  |
| --- | --- | --- |
|  | ABC | D |
| ABC | 0 | 3 |
| D | 0 | 0 |

1. Plot graph in google colab - [link](https://colab.research.google.com/drive/1hPp9s7gQCU4SKMdwr6Yz6ln9g65D3XwT?usp=sharing)

A diagram with orange lines

Description automatically generated

**(b)**

Initial Distance Matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| A | 0 | 1 | 4 | 5 |
| B |  | 0 | 2 | 6 |
| C |  |  | 0 | 3 |
| D |  |  |  | 0 |

1. Step a: Combine the closest clusters. From the matrix, it's clear that A and B are the closest with a distance of 1. We combine them into a new cluster AB.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | AB | B | C | D |
| A | 0 | 1 | 4 | 5 |
| B |  | 0 | 2 | 6 |
| C |  |  | 0 | 3 |
| D |  |  |  | 0 |

1. Step b: Update the matrix. The distance from AB to the other clusters is the maximum distance from A or B to the other clusters.
2. Step c: Combine the next closest clusters. C and D are the closest with a distance of 3. We combine them into a new cluster CD.
3. Step d: Update the matrix again. The distance from AB to CD is the maximum distance from AB to C or D, which is 6.

|  |  |  |  |
| --- | --- | --- | --- |
|  | AB | C | D |
| AB | 0 | 3 | 6 |
| C |  | 0 | 4 |
| CD |  |  | 0 |

1. Step e: Combine the last two clusters AB and CD with the distance of 6.

|  |  |  |
| --- | --- | --- |
|  | AB | CD |
| AB | 0 | 6 |
| CD |  | 0 |

1. Plot graph in google colab - [link](https://colab.research.google.com/drive/1hPp9s7gQCU4SKMdwr6Yz6ln9g65D3XwT?usp=sharing)

A diagram of a clustering diagram

Description automatically generated