**Logo

Description automatically generated**

**San Francisco Bay University**

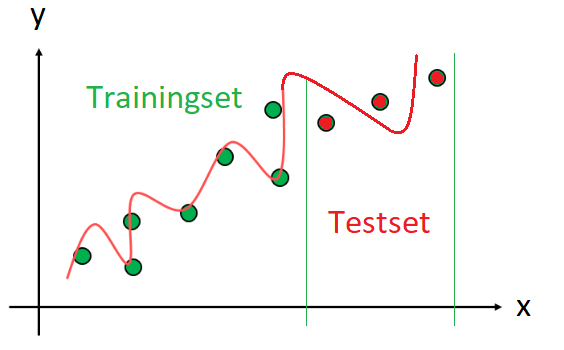
**CS483 - Fundamentals of Artificial Intelligence**

**2022 Summer Midterm Exam**

**Student Name: Student ID:**

**Instruction**

1. **Put your answer right after each question in the answer sheet**
2. **Make copy & paste for your program in text mode, NOT image onto the answer sheet.**
3. **Excel is preferred for hand calculation**
4. If we want to fit **trainingset** (X, y) with **single** feature in a squiggle curve as follows, please create **hypothesis/loss/cost** functions and **partial derivative** expression *vs* each coefficient for future gradient descent algorithm implementation first. But if you find that this hypothesis function will generate quite large errors for the testset as follows after modeling, how to tweak it?



***\*Hint:*** *high order polynomial hypothesis function is one of options. You should make decision about what the highest order is based on your fitting observation.*

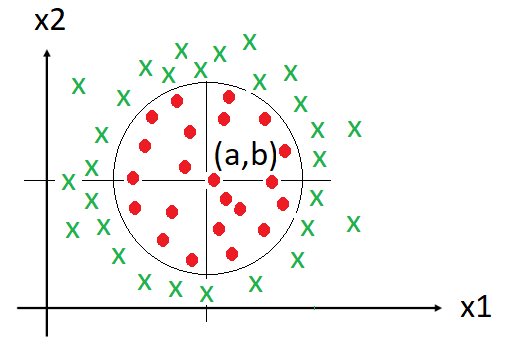
1. Write python program to make feature scaling for "***Alcohol***" and "***Malic acid***" in the following dataset

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Alcohol | Malic acid | Class label |
| 0 | 14.23 | 1.71 | 0 |
| 1 | 13.2 | 1.78 | 1 |
| 2 | 13.16 | 2.36 | 1 |
| 3 | 14.37 | 1.95 | 0 |
| 4 | 13.24 | 2.59 | 0 |

1. Given the dataset as below, write python program or hand calculation to find in linear regression hypothesis function by **normal** equation method

|  |  |
| --- | --- |
| X | y |
| 1 | 7 |
| 2 | 9 |
| 3 | 12 |
| 4 | 15 |
| 5 | 16 |

1. In the process of applying gradient descent algorithm to find max value for each coefficient in hypothesis function, appropriate learning rate α is very important. Please discuss about how it makes the impacts on your training result
2. A trainingset with *2* features (, ) and binary class labeled in "o" and "x" as *0* and *1* respectively is visualized as follows. In terms of our observation, circle decision boundary in **logistic** regression is a good option for binary classification. Please buildup hypothesis/cost function and partial derivative expression vs each coefficient for gradient descent algorithm application



***\*Hint:*** *circle equation is*

1. KNN algorithm is one of supervised learning algorithms. If *K=1*, what is error rate in given "**Testset**" after training based on "**Trainingset**" from "iris" dataset as follows?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trainingset** |  |  |  |  |
| **sepal length** | **sepal  width** | **petal length** | **petal width** | **Species (label)** |
| 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 4.9 | 3 | 1.4 | 0.2 | setosa |
| 6.6 | 2.9 | 4.6 | 1.3 | versicolor |
| 5.2 | 2.7 | 3.9 | 1.4 | versicolor |
| 6.9 | 3.1 | 5.1 | 2.3 | virginica |
| 5.8 | 2.7 | 5.1 | 1.9 | virginica |
|  |  |  |  |  |
| **Testset** |  |  |  |  |
| **sepal length** | **sepal  width** | **petal length** | **petal width** | **Species**  **(label)** |
| 7 | 3.2 | 4.7 | 1.4 | versicolor |
| 5 | 3.3 | 1.4 | 0.2 | setosa |
| 5.9 | 3 | 5.1 | 1.8 | virginica |

1. If ***K = 2***, please classify the follows points (, ) in K-Means algorithm to different classes and calculate **total** "within cluster sum of square" (WCSS) value in either python program or hand calculation

|  |  |  |
| --- | --- | --- |
| ID | X1 | X2 |
| A1 | 2 | 10 |
| A2 | 8 | 4 |
| A3 | 5 | 8 |
| A4 | 6 | 4 |
| A5 | 1 | 2 |