Machine Learning (Assignment # 1)

Murali Yaswanth Motamarri -700741647

Github link: <https://github.com/Yaswanthmm/machine-learning#machine-learning>

Video link: <https://drive.google.com/file/d/1vpjwtKApsNIBKloTCEPlSUdN60yB4b1o/view?usp=share_link>

# Question 1

The following is a list of 10 students ages:

ages = [19, 22, 19, 24, 20, 25, 26, 24, 25, 24]

* Sort the list and find the min and max age
* Add the min age and the max age again to the list
* Find the median age (one middle item or two middle items divided by two)
* Find the average age (sum of all items divided by their number)
* Find the range of the ages (max minus min)

**Solution Screenshot:**

**Text

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

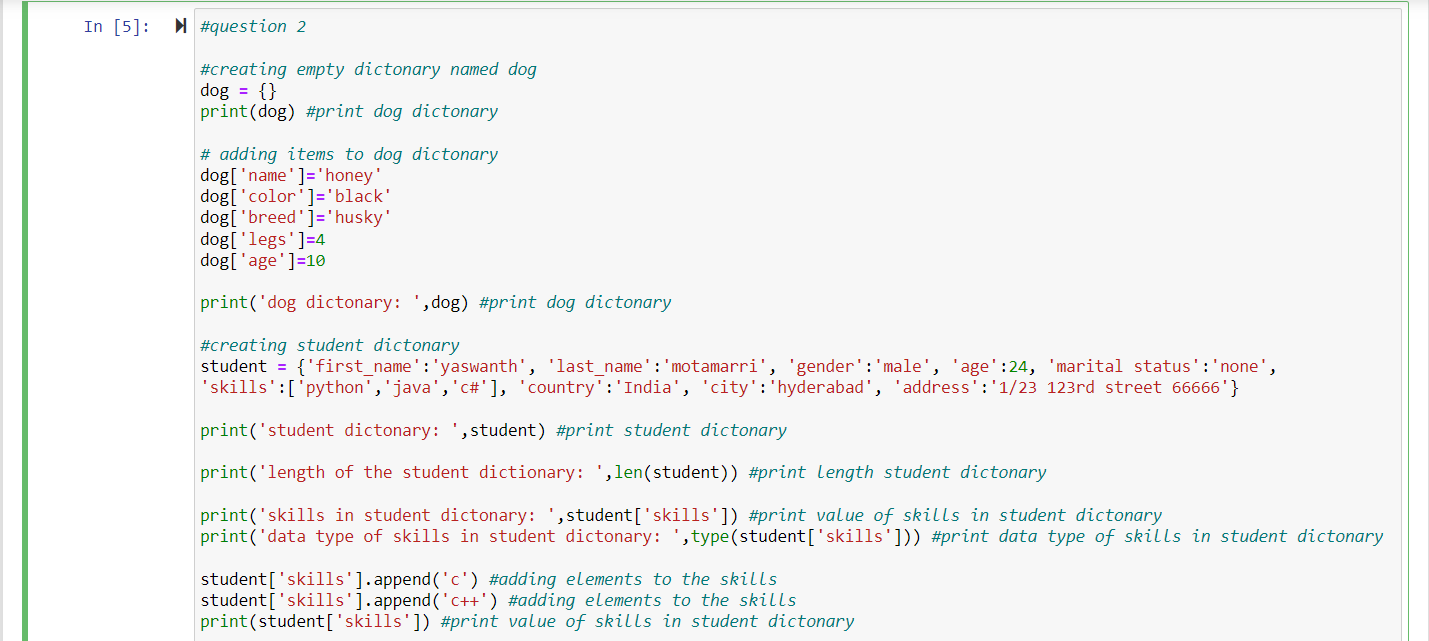
**Description:** In this question I used sort() built-in method to sort the elements, min() method to find the minimum item, max() method to find maximum element, append() method to add elements to the list and len() method to length of the list.

To find median, I used the if-else statement with length of list in the condition and median formula. And for finding range subtracted max(ages) - min(ages)

# Question 2

* Create an empty dictionary called dog
* Add name, color, breed, legs, age to the dog dictionary
* Create a student dictionary and add first\_name, last\_name, gender, age, marital status, skills, country, city and address as keys for the dictionary
* Get the length of the student dictionary
* Get the value of skills and check the data type, it should be a list
* Modify the skills values by adding one or two skills
* Get the dictionary keys as a list
* Get the dictionary values as a list

**Solution Screenshot:**



Scatter chart

Description automatically generated

**Description:**

In this question first created empty dog dictionary and added elements to the dictionary. Created a student dictionary and found length of student dict. Printed skills and the datatype of student dictionary. Added skills to the dictionary and printed the keys and values using keys(), values() command.

# Question 3

* Create a tuple containing names of your sisters and your brothers (imaginary siblings are fine)
* Join brothers and sisters tuples and assign it to siblings
* How many siblings do you have?
* Modify the siblings tuple and add the name of your father and mother and assign it to family\_members

**Solution Screenshot:**

**Text

Description automatically generated**

**Description:**

In this question first created two tuples sisters, brothers. Joined sisters and brothers and assigned it to siblings. Found the length of siblings tuple using len() menthod. Converted the tuple to list and added father and mother name using append() method and converted it back to tuple and assigned it to family\_members.

# Question 4

it\_companies = {'Facebook', 'Google', 'Microsoft', 'Apple', 'IBM', 'Oracle', 'Amazon'} A = {19, 22, 24, 20, 25, 26}

B = {19, 22, 20, 25, 26, 24, 28, 27}

age = [22, 19, 24, 25, 26, 24, 25, 24]

* Find the length of the set it\_companies
* Add 'Twitter' to it\_companies
* Insert multiple IT companies at once to the set it\_companies
* Remove one of the companies from the set it\_companies
* What is the difference between remove and discard
* Join A and B
* Find A intersection B
* Is A subset of B
* Are A and B disjoint sets
* Join A with B and B with A
* What is the symmetric difference between A and B
* Delete the sets completely
* Convert the ages to a set and compare the length of the list and the set.

**Solution Screenshot:**

**Text

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Description:**

In this question I found the length of set it\_companies using len() method. Added companies ‘Twitter’ using add() method and 'cognizant','infosys','wipro' using update() method. Removed company using remove() method. Joined sets A and B using union() method. Printed intersection of sets A and B using intersection() method. Checked if A and B are subsets using issubset() method. Checked if A and B are disjoint sets using isdisjoint() method. Found symmetric difference using symmetric\_difference() method. Converted the list age to set and compared the lengths of set and list using len() method. Deleted the sets using del command.

What is the difference between remove and discard?

-->discard() is a built in method which is used to remove a specific element from a set, if the element is not present in the set, it prints the original set instead of raising an error or exception.'''

-->remove() is also a built in method which is used to remove a specific element from a set, but it will throw an error or exception if the element is not present in the set'''

# Question 5

The radius of a circle is 30 meters.

* Calculate the area of a circle and assign the value to a variable name of \_area\_of\_circle\_
* Calculate the circumference of a circle and assign the value to a variable name of

\_circum\_of\_circle\_

* Take radius as user input and calculate the area.

**Solution Screenshot:**

**Text

Description automatically generated**

**Description:**

Imported math class and found area and circumference. Took input from user using input command.

# Question 6

“I am a teacher and I love to inspire and teach people”

* How many unique words have been used in the sentence? Use the split methods and set to get the unique words.

**Solution Screenshot:**

**Graphical user interface, application

Description automatically generated**

**Description:**

Using split() method converted the string to list and then converted the list to set to find the number of unique words. Used the len() method to find the length of set.

# Question 7

Use a tab escape sequence to get the following lines.

**Name Age Country City**

**Asabeneh 250 Finland Helsinki**

**Solution Screenshot:**

**Graphical user interface, scatter chart

Description automatically generated**

**Description:**

Printed the lines given using tab escape sequence(\t).

# Question 8

Use the string formatting method to display the following:

radius = 10

area = 3.14 \* radius \*\* 2

“The area of a circle with radius 10 is 314 meters square.”

**Solution Screenshot:**

Graphical user interface, application, chat or text message

Description automatically generated

**Description:** Calculated the area of circle and printed the area of circle using string formatting method.

# Question 9

Write a program, which reads weights (lbs.) of N students into a list and convert these weights to kilograms in a separate list using Loop. N: No of students (Read input from user)

Ex: L1: [150, 155, 145, 148]

Output: [68.03, 70.3, 65.77, 67.13]

**Solution Screenshot:**

Graphical user interface, text, application, chat or text message, email

Description automatically generated

**Description:**

Using loop and append() method got the weights of students in lb and converted the weights to kg.

# Question 10

The diagram below shows a dataset with 2 classes and 8 data points, each with only one feature value, labeled f. Note that there are two data points with the same feature value of 6. These are shown as two x’s one above the other. Provide stepwise mathematical solution, do not write code for it.

A picture containing clock

Description automatically generated

1. Divide this data equally into two parts. Use first part as training and second part as testing. Using KNN classifier, for K=3, what would be the predicted outputs for the test samples? Show how you arrived at your answer.
2. Compute the confusion matrix for this and calculate accuracy, sensitivity and specificity values.
3. Tabulating data points from the above diagram

|  |  |
| --- | --- |
| **Data point** | **Class** |
| [1,0] |  |
| [2,0] |  |
| [3,0] | **X** |
| [6,0] | **X** |
| [6,0] | **X** |
| [7,0] |  |
| [10,0] |  |
| [11,0] |  |

Dividing datapoints into training and test datasets equally

|  |  |
| --- | --- |
| **Training Dataset** | |
| **Data point** | **Class** |
| [1,0] |  |
| [3,0] | **X** |
| [6,0] | **X** |
| [7,0] |  |

|  |  |
| --- | --- |
| **Test Dataset** | |
| **Data point** | **Class** |
| [2,0] |  |
| [6,0] | **X** |
| [10,0] |  |
| [11,0] |  |

Using the KNN classifier, for K=3. We can get the Euclidian distance between the points (x1, y1) and (x2,y2) using the formula ((x2-x1)^2 – (y2-y1)^2)^1/2. We can get the class of the test data point

For [2,0] :- The nearest three test points are

|  |  |
| --- | --- |
| **Data point** | **Class** |
| [1,0] |  |
| [3,0] | **X** |
| [6,0] | **X** |

Here the predicted class for the test point is **X**

For [6,0] :- The nearest three test points are

|  |  |
| --- | --- |
| **Data point** | **Class** |
| [3,0] | **X** |
| [6,0] | **X** |
| [7,0] |  |

Here the predicted class for the test point is **X**

For [10,0] :- The nearest three test points are

|  |  |
| --- | --- |
| **Data point** | **Class** |
| [3,0] | **X** |
| [6,0] | **X** |
| [7,0] |  |

Here the predicted class for the test point is **X**

For [11,0] :- The nearest three test points are

|  |  |
| --- | --- |
| **Data point** | **Class** |
| [3,0] | **X** |
| [6,0] | **X** |
| [7,0] |  |

Here the predicted class for the test point is **X**

The actual and predicted classes for the test data sets are

|  |  |  |
| --- | --- | --- |
| **Data point** | **Actual Class** | **Predicted Class** |
| [2,0] |  |  |
| [6,0] | **X** |  |
| [10,0] |  |  |
| [11,0] |  |  |

1. The confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Actual values** | | |
| **Predicted values** |  | **X** |  |
| **X** | **TP=0** | **FP=1** |
|  | **FN=0** | **TN=3** |

Accuracy =(TP+TN)/(TP+FP+FN+TN)

=3/4=0.75

Sensitivity=TP/(TP+FN)

=0/0=0

Specificity=TN/(TN+FP)

=3/4=0.75