MACHINE LEARNING

1. Which of the following are disadvantages of using Hard Margin SVM classifier?
   A) They allow misclassifications, that's why they are not optimal.
   B) They cannot be used when the data is not completely linearly separable while allowing no errors.
   C) They are not optimal to use in case of outliers.
   D) None of the above.

   Ans:- B, C

2. Which of the following statements are true regarding maximal margin classifier?
   A) It is the most optimal classifier in a completely linearly separable data.
   B) It's the classifier for which the margin length or the distance between the closest data-point on eitherside of the classifier and the classifier is maximized.
   C) Any possible classifier which can linearly separate the data of two classes is called maximal marginclassifier.
   D) All of the above.

   Ans:- B

3. Which of the following statements are true regarding soft margin SVM classifier?
   A) They are less sensitive to outliers and can be used even in their presence.
   B) They make sure that there is no data point present in the margin area.
   C) They allow some degree of errors or misclassification.
   D) They can be used in case data is not completely linearly separable.

   ## Ans:- A,C,D

4. Which of the following statements are true regarding SVMs?
   A) They take the data from lower dimensional space to some higher dimensional space in case thedata is not likely to be linearly separable.
   B) They use the kernel tricks to escape the complex computations required to transform the data.
   C) If the data is not linearly separable SVM technique cannot be used.
   D) All of the above.
   Ans:- A, B

5. Which of the following Statements are true regarding the Kernel functions used in SVM?
   A) These functions gives value of the dot product of pairs of data-points in the desired higher. dimensional space without even explicitly converting the whole data in to higher dimensional space.
   B) We have to first convert the whole data in to the higher dimensional space before applying thekernel function.
   C) The data product values given by the kernel functions are used to find the classifier in the higher dimensional space.
   D) None of the above
   Ans:- A,C

6. How can SVM be classified?
   A) It is a model trained using unsupervised learning. It can be used for classification and regression.
   B) It is a model trained using unsupervised learning. It can be used for classification but not for regression
   C) It is a model trained using supervised learning. It can be used for classification and regression.
   D) It is a model trained using supervised learning. It can be used for classification not for regression.
   **Ans:-D**

MACHINE LEARNING

7. The quality of an SVM model depends upon:
   A) Selection of Kernel
   B) Kernel Parameters
   C) Soft Margin Parameter C
   D) All of the above

   Ans:- D

7. The quality of an SVM model depends upon:
   A) Selection of Kernel
   B) Kernel Parameters
   C) Soft Margin Parameter C
   D) All of the above

MACHINE LEARNING

8. The SVM's are less effective when:
   A) The data is linearly separable.
   B) The data is clean and ready to use.
   C) The data is noisy and contains overlapping points.
   D) None of these.
      Ans:- C

9. What would happen when you use very small C (C~0)?
   A) Misclassification would happen.
   B) Data will be correctly classified.
   C) Can't say
   D) None of these.
      Ans:-A

10. What do you mean by generalization error in terms of the SVM?
    A) How far the hyperplane is from the support vectors.
    B) How accurately the SVM can predict outcomes for unseen data.
    C) The threshold amount of error in an SVM.
    D) None of these.
       Ans:- B
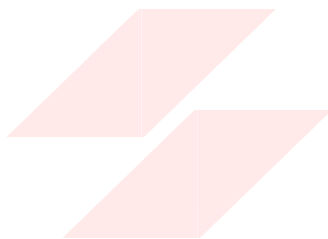
WORKSHEET 2 PYTHON

Q1 to Q7 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is not a core datatype in python?

   A) list   **B) struct**

   C) tuple   C) set

   Ans:- B struct

2. Which of the following is an invalid variable name in python?

   A) _init_   B) no_1

   C) 1_no   D) _1

   Ans:-C 1_no

3. Which one of the following is a keyword in python?

   A) in   B) _init_

   C) on   D) foo

   Ans:- A in

4. **In which of the** following manner are the operators of the same precedence executed in python?

   A) Left to Right   B) BODMAS

   C) Right to Left   D) None of these

   Ans:- A Left to Right

5. Arrange the following in decreasing order of the precedence when they appear in an expression in python?
   i) Multiplication ii) Division iii) Exponential iv) Parentheses

   A) iii – iv – ii – i   B) iii – iv – i – ii

   C) iv – iii – ii – i   D) iii – ii – i – iv

   Ans:- C iv – iii – ii – i

6. (28//6)**3/3%3 = ?

   A) 7.1111…   B) 0

   C) 0.3333…   D) 1

   Ans:- C 0.3333…

7. a = input("Enter an integer"). What will be the data type of a?

   A) int   B) str

   C) float   D) double

   Ans:- B) str

Q8 and Q10 have multiple correct answers. Choose all the correct options to answer your question.

8. Which of the following statements are correct?

   A) Division and multiplication have same precedence in python

   B) Python's operators' precedence is based on PEDMAS

   C) Python's operators' precedence is based on VBODMAS

   D) In case of operators' having the same precedence, the one on the left side is executed

   first.

   Ans:- D

9. Which of the following is(are) valid statement(s) in python?

   A) abc = 1,000,000   B) a b c = 1000 2000 3000

   C) a,b,c = 1000, 2000, 3000   D) a_b_c = 1,000,000

   Ans:-C

FLIP ROBO

10. Which of the following is not equal to x^{16} in python?

A) x**4**4   B) x**16

C) x^16   D) (x**4)**4

Ans:-C x^16

Q11 to Q13 are subjective questions, answer them briefly

11. Differentiate between a list, tuple, set and dictionary.

Ans:- List:

A list is an ordered collection of items, which can be of different data types.

It is mutable, which means that you can change its content after creation.

It is defined by enclosing a comma-separated sequence of items in square brackets.

Example: my_list = [1, 'apple', 3.14, True]

Tuple:

A tuple is similar to a list, but it is immutable, meaning its content cannot be changed after creation.

It is defined by enclosing a comma-separated sequence of items in parentheses.

Example: my_tuple = (1, 'apple', 3.14, True)

Set:

A set is an unordered collection of unique elements.

It is defined by enclosing a comma-separated sequence of items in curly braces or by using the set() function.

Example: my_set = {1, 'apple', 3.14, True}

Dictionary:

A dictionary is an unordered collection of key-value pairs.

It is defined by enclosing a comma-separated sequence of key-value pairs in curly braces or by using the dict() function.

Example: my_dict = {'key1': 'value1', 'key2': 2, 'key3': True}

12. Are strings mutable in python? Suppose you have a string "I+Love+Python", write a small code to replace '+' with space in python.

**Ans:-** string = "I+Love+Python"

new_string = string.replace("+", " ")

print(new_string)

13. What does the function **ord()** do in python? Explain with an example. Also, write down the function for getting the data type of a variable in python.

Ans:- The ord() function in Python returns an integer representing the Unicode character. It takes a single argument, which is a Unicode character, and returns its Unicode code point. For example, the Unicode code point for the character 'a' is 97.

Q14 and Q15 are programming questions. Answer them in Jupyter Notebook.

14. Write a python program to solve a quadratic equation of the form ax^2+bx+c=0. Where a, b and c are to be taken by user input. Handle the erroneous input, such as 'a' should not be equal to 0.

15. Write a python program to find the sum of first 'n' natural numbers without using any loop. Ask users to input the value of 'n'.
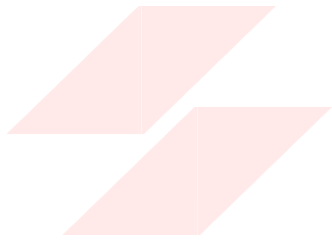
14. Write a python program to solve a quadratic equation of the form ax^2+bx+c=0. Where a, b and c are to be taken by user input. Handle the erroneous input, such as 'a' should not be equal to 0.

```python
import math

# get input values of a, b, and c
a = float(input("Enter the value of a (should not be zero): "))
if a == 0:
    print("a should not be zero")
    exit()
b = float(input("Enter the value of b: "))
c = float(input("Enter the value of c: "))

# calculate discriminant
discriminant = b**2 - 4*a*c

# check if discriminant is positive, zero, or negative
if discriminant > 0:
    # two real and distinct roots
    x1 = (-b + math.sqrt(discriminant)) / (2*a)
    x2 = (-b - math.sqrt(discriminant)) / (2*a)
    print("The roots are", x1, "and", x2)
elif discriminant == 0:
    # one real and repeated root
    x = -b / (2*a)
    print("The root is", x)
else:
    # two complex conjugate roots
    real_part = -b / (2*a)
    imag_part = math.sqrt(-discriminant) / (2*a)
    print("The roots are", real_part, "+", imag_part, "i and", real_part, "-", imag_part
```

```
Enter the value of a (should not be zero): 5
Enter the value of b: 21
Enter the value of c: 5
The roots are -0.2533814687380612 and -3.946618531261939
```

15 Write a python program to find the sum of first 'n' natural numbers without using any loop. Ask users to input the value of 'n'

```python
n = int(input("Enter a positive integer n: "))
if n < 1:
    print("n should be a positive integer")
    exit()

sum = (n * (n + 1)) // 2
print("The sum of the first", n, "natural numbers is", sum)
```

```
Enter a positive integer n: 20
The sum of the first 20 natural numbers is 210
```

STATISTICS WORKSHEET-9

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. The owner of a travel agency would like to determine whether or not the mean age of the agency's customers is over 24. If so, he plans to alter the destination of their special cruises and tours. If he concludes the mean age is over 24 when it is not, he makes a _____ error. If he concludes the mean age is not over 24 when it is, he makes a _____ error.
a. Type II; Type II
b. Type I; Type I
c. Type I; Type II
d. Type II; Type I
   Ans:- C

2. Suppose we wish to test H_0: $\mu = 53$ vs H_1: $\mu > 53$. What will result if we conclude that the mean is greater than 53 when its true value is really 55?
a. We have made a Type I error
b. We have made a correct decision
c. We have made a Type II error
d. None of the above are correct
   Ans:- b

3. The value that separates a rejection region from an acceptance region is called a _____.
a. parameter
b. critical value
c. confidence coefficient
d. significance level
   Ans:- b

4. A hypothesis test is used to prevent a machine from under filling or overfilling quart bottles of beer. On the basis of sample, the machine is shut down for inspection. A thorough examination reveals there is nothing wrong with the filling machine. From a statistical point of view:
a. Both Type I and Type II errors were made.
b. A Type I error was made.
c. A Type II error was made.
d. A correct decision was made.
   Ans:- d

5. Suppose we wish to test H_0 : $\mu = 21$ vs H_1 : $\mu > 21$. Which of the following possible sample results gives the most evidence to support H_1 (i.e., reject H_0)? Hint: Compute Z-score.
a. x = 23 s , = 3
b. x = 19 s , = 4
c. x = 17 s , = 7
d. x = 18 s , = 6
 Ans:- d

6. Given H_0: $\mu = 25$, H_1: $\mu \neq 25$, and P-value = 0.041. Do you reject or fail to reject H0 at the 0.01 level of significance?
a. fail to reject H0

b. not sufficient information to decide

c. reject H_0

Ans:- d

7. A bottling company needs to produce bottles that will hold 12 ounces of liquid. Periodically, the company gets complaints that their bottles are not holding enough liquid. To test this claim, the bottling company randomly samples 36 bottles. Suppose the p-value of this test turned out to be 0.0455. State the proper conclusion.
a. At $\alpha = 0.085$, fail to reject the null hypothesis.
b. At $\alpha = 0.035$, accept the null hypothesis.
c. At $\alpha = 0.05$, reject the null hypothesis.
d. At $\alpha = 0.025$, reject the null hypothesis.

   Ans:- c

8. If a hypothesis test were conducted using $\alpha = 0.05$, for which of the following p-values would the null hypothesis be rejected?
a. 0.100
b. 0.041
c. 0.055
d. 0.060

 Ans:- b

9 . For $H_1: \mu > \mu_0$ p-value is 0.042. What will be the p-value for Ha: $\mu < \mu o$?
a. 0.084
b. 0.021
c. 0.958
d. 0.042

 Ans:- d

10. The test statistic is $t = 2.63$ and the p-value is 0.9849. What type of test is this?
a. Right tail
b. Two tail
c. Left tail
d. Can't tell

   Ans:- c

11. The test statistic is $z = 2.75$, the critical value is $z = 2.326$. The *p*- value is …
a. Less than the significance level
b. Equal to the significance level
c. Large than the significance level

   Ans:- a

12. The area to the left of the test statistic is 0.375. What is the probability value if this is a left tail test?
a. 0.750
b. 0.375
c. 0.1885
d. 0.625

 Ans:-b

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly.

13. What is T distribution and Z distribution?

Ans:- The t-distribution and z-distribution are both probability distributions that are commonly used in statistics for hypothesis testing and confidence interval estimation.

The z-distribution is a normal distribution with a mean of 0 and a standard deviation of 1. It is used when the population standard deviation is known and the sample size is large (typically n > 30). The z-distribution is often used in hypothesis testing when the population variance is known.

The t-distribution, also known as Student's t-distribution, is a family of probability distributions that arise from the estimation of the mean of a normally distributed population in situations where the sample size is small (typically n < 30) or the population standard deviation is unknown. The t-distribution is similar in shape to the normal distribution but has heavier tails, which reflects the increased uncertainty due to the smaller sample size. The t-distribution is often used in hypothesis testing when the population variance is unknown.

14.Is the T distribution normal?

Ans:-The T distribution is similar to the normal distribution in shape, but it has fatter tails, which means it has more variability. As the sample size increases, the T distribution approaches the normal distribution. Therefore, the T distribution is not exactly normal, but it becomes more similar to the normal distribution as the sample size increases.

15.What does the T distribution tell us?
Ans:- The t-distribution is a probability distribution used in hypothesis testing, which provides a way to estimate population parameters when the sample size is small or the population variance is unknown. The t-distribution tells us how the sample mean deviates from the population mean when the population standard deviation is unknown and is estimated from the sample. It is similar to the

standard normal distribution, but it has heavier tails, which means that it provides wider confidence intervals and larger p-values for the same level of significance compared to the normal distribution. The shape of the t-distribution depends on the degrees of freedom, which is determined by the sample size. As the sample size increases, the t-distribution approaches the normal distribution.