Assignment no 9

Machine Learning

Q-1:- Which of the following are disadvantages of using Hard Margin SVM classifier?

Ans:- They are not optimal to use in case of outliers.(option-C)

Q-2:- Which of the following statements are true regarding maximal margin classifier?

Ans:- 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 either side of the classifier and the classifier is maximized.

Q-3:- Which of the following statements are true regarding soft margin SVM classifier?

Ans:- A) They are less sensitive to outliers and can be used even in their presence.

C) They allow some degree of errors or misclassification.

D) They can be used in case data is not completely linearly separable.

Q-4:- Which of the following statements are true regarding SVMs?

Ans:- A) They take the data from lower dimensional space to some higher dimensional space in case the data is not likely to be linearly separable.

B) They use the kernel tricks to escape the complex computations required to transform the data.(option A and B are correct)

Q-5:- Which of the following Statements are true regarding the Kernel functions used in SVM?

Ans:- 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.

C) The data product values given by the kernel functions are used to find the classifier in the higher dimensional space.

Q-6:- How can SVM be classified?

Ans:- It is a model trained using supervised learning. It can be used for classification and regression.(option C)

Q-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 (correct one)

Q-8 The SVM's are less effective when,

Ans:- :- The data is noisy and contains overlapping points.(option C)

Q-9:- What would happen when you use very small C (C~0)?

Ans:- Misclassification would happen (option A)

Q-10:- What do you mean by generalization error in terms of the SVM?

Ans:- B) How accurately the SVM can predict outcomes for unseen data.

Python Worksheet 2

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

Ans:- struct (option B)

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

Ans:- 1_no (option c)

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

Ans:- foo(option D)

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

Ans:- left to right(option A)

Q-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

Ans:- iii – ii – i – iv (option-D)

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

Ans:- 0.3333…..(option C)

Q-7. a = input(“Enter an integer”). What will be the data type of a?

Ans:- int(option A)

Q-8. Which of the following statements are correct?

Ans:- A) Division and multiplication have same precedence in python

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

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

Ans:- A) abc = 1,000,000

D) a_b_c = 1,000,000

Q-10. Which of the following is not equal to x16 in python?

Ans:- x^16 (option c)

Q-11. Differentiate between a list, tuple, set and dictionary?

Ans. In Python, a list, tuple, set, and dictionary are all different types of data structures used for storing and organizing data.

Here are the key differences between these four data structures:

1. List: A list is an ordered collection of elements, which can be of different data types. Lists are mutable, which means that you can add, remove, or modify elements in a list. Lists are defined using square brackets [].

   Example: my_list = [1, 2, "three", 4.0]

2. Tuple: A tuple is similar to a list, but it is immutable, which means that you cannot add, remove, or modify elements in a tuple once it is defined. Tuples are defined using parentheses ().

   Example: my_tuple=(1,2,”three”,4.0)

3. Set: A set is an unordered collection of unique elements. Sets are mutable, and you can add or remove elements from a set. Sets are defined using curly braces {} or the set() function.

Example: my_set{1,2,3,5}

4. Dictionary: A dictionary is an unordered collection of key-value pairs, where each key is associated with a value. Dictionaries are mutable, and you can add, remove, or modify key-value pairs in a dictionary. Dictionaries are defined using curly braces {} and colons : to separate the keys and values.

   Examples: my_dict = {"name": "John", "age": 30, "city": "New York"}

Q 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. No, strings are immutable in Python, which means that you cannot change the contents of a string after it is created. However, you can create a new string with the desired modifications.

```
my_string = "I+Love+Python"
new_string = my_string.replace("+", " ")
print(new_string)
```

The **replace()** method is used to replace all occurrences of '+' in the original string with a space. The resulting string is stored in the variable **new_string**. The **print()** function is then used to display the modified string. The output of this code will be:

I Love Python

Q13. 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. In Python, the built-in function **ord()** is used to return the integer value of a given Unicode character. The **ord()** function takes a single argument, which can be a string of length 1 representing a single Unicode character, and returns an integer representing the Unicode code point of that character.

Here's an example of how to use the **ord()** function:

```
my_char = 'A'
my_char_code = ord(my_char)
print(my_char_code)
```

To get the data type of a variable in Python, you can use the built-in function **type()**. The **type()** function takes a single argument, which can be any Python object, and returns the type of that object.

Here's an example of how to use the **type()** function:

```
my_var = 42
my_var_type = type(my_var)
print(my_var_type)
```

STATISTICS WORKSHEET 9

Q-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.

Ans: Type I; Type II (option C)

Q-2. Suppose we wish to test H0: μ =53 vs H1: μ > 53. What will result if we conclude that the mean is greater than 53 when its true value is really 55?

Ans:- We have made a Type ll error (option C)

Q-3. The value that separates a rejection region from an acceptance region is called a _____.

Ans. Critical Value(option B)

Q-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:

Ans:- A correct decision was made.(option d)

Q-5. Suppose we wish to test H0 : μ =21 vs H1 : μ > 21. Which of the following possible sample results gives the most evidence to support H1 (i.e., reject H0)? Hint: Compute Z-score.

Ans:- x = 23 s , = 3 (option A)

Q-6. Given H0: μ = 25, H1: μ ≠ 25, and P-value = 0.041. Do you reject or fail to reject H0 at the 0.01 level of significance?

Ans:- fail to reject H0(option A)

Q-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.

Ans:- At α = 0.085, fail to reject the null hypothesis (option A)

Q-8. If a hypothesis test were conducted using $\alpha = 0.05$, for which of the following p-values would the null hypothesis be rejected?

Ans. 0.041 (option-B)

Q-9. For H1: $\mu > \mu 0$ p-value is 0.042. What will be the p-value for Ha: $\mu < \mu o$?

Ans:- 0.021 (option b)

Q-10. The test statistic is t = 2.63 and the p-value is 0.9849. What type of test is this?

Ans:- two tail test(option B)

Q-11. The test statistic is z =2.75, the critical value is z = 2.326. The p- value is …

Ans:- Less than the significance level(option-A)

Q-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?

Ans:- 0.375 (option B)

Q-13. What is T distribution and Z distribution?

Ans:- T distribution and Z distribution are both probability distributions that are commonly used in statistical inference. However, they have some important differences.

Z distribution, also known as the standard normal distribution, is a probability distribution with a mean of 0 and a standard deviation of 1. It is a continuous distribution that is symmetrical and bell-shaped. Z distribution is used in hypothesis testing and confidence interval estimation when the population standard deviation is known.

T distribution, also known as the Student's t-distribution, is a probability distribution with a mean of 0 and a standard deviation greater than 1. It is a continuous distribution that is also symmetrical and bell-shaped, but has more variability than the Z distribution. T distribution is used in hypothesis testing and confidence interval estimation when the population standard deviation is unknown and must be estimated from the sample data.

The main difference between the two distributions is that the t-distribution has fatter tails than the normal distribution, meaning it has more probability in the tails. This is because the t-distribution is based on smaller sample sizes than the normal distribution. As sample size increases, the t-distribution becomes closer to the normal distribution.

Q-14. Is the T distribution normal?

Ans:- The t-distribution is similar to the normal distribution in that it is a symmetrical, bell-shaped distribution. However, it is not exactly the same as the normal distribution.

The t-distribution has more variability than the normal distribution, which means that it has fatter tails. This is because the t-distribution is based on smaller sample sizes than the normal distribution. As the sample size increases, the t-distribution becomes closer to the normal distribution.

Therefore, while the t-distribution is not exactly the same as the normal distribution, it is similar and can be used as an approximation of the normal distribution under certain conditions, such as when the sample size is large or the sample standard deviation is close to the population standard deviation.

Q-15. What the T distribution tell us?

Ans:- The t-distribution is a probability distribution that is used in statistical inference to estimate population parameters based on sample statistics. In particular, the t-distribution is used to:

1. Construct confidence intervals for population means when the population standard deviation is unknown and must be estimated from the sample data.
2. Test hypotheses about population means when the population standard deviation is unknown and must be estimated from the sample data.

The t-distribution is also used to account for the increased uncertainty that arises from estimating the population standard deviation from the sample data. As the sample size increases, the t-distribution becomes closer to the normal distribution.