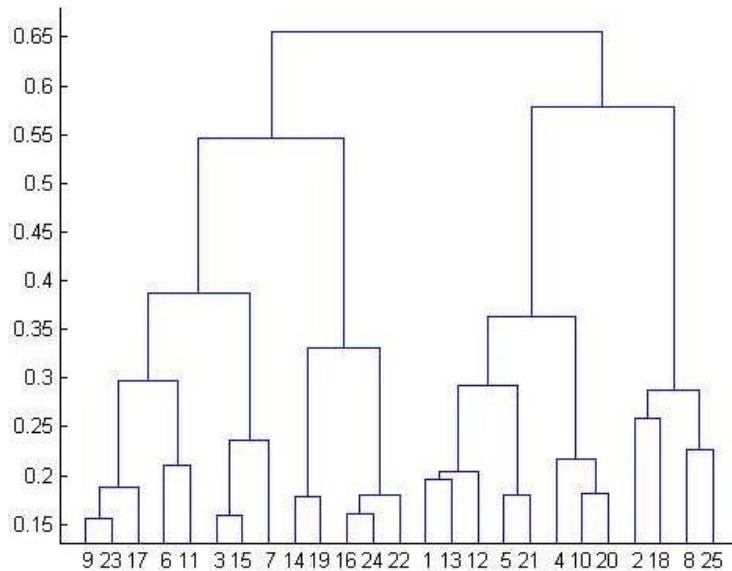


Which analysis can be drawn from the given dendrogram after conducting K-means Clustering? * 1 point



Average-link clustering is used as a proximity function

Number of clusters formed after data points analysis is 4

☐ There are 28 data points in clustering analysis

The above figure interpretation is not true for K-Means clustering

If you want to cluster seven data points into three clusters using K-Means. * 2 points
After 1st iteration clusters, C1, C2, C3 has following observations:

C1: ((2,2), (4,4), (6,6))

C2: ((0,4), (4,0))

C3: ((5,5), (9,9))

Find the cluster centroids for the second iteration?

C1 : (6,6), C2: (4,4), C3: (9,9)

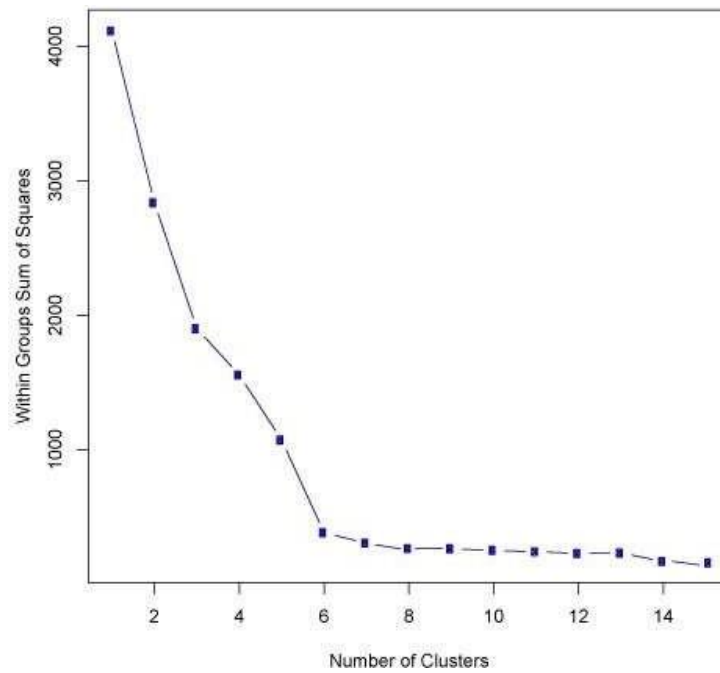
☐ C1: (2,2), C2: (0,0), C3: (5,5)

C1: (4,4), C2: (2,2), C3: (7,7)

None of the above

What is the best choice for no. of clusters based on the following figure? *

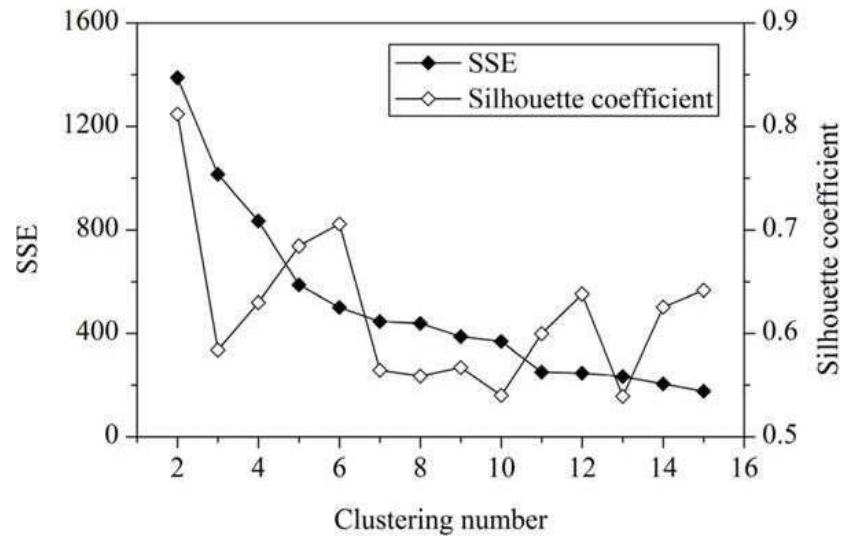
1 point



- ☒ 6
- ☐ 5
- ☐ 14
- ☐ 15



What is the best choice for no. of clusters based on the following figure? * 1 point
where SSE is Sum of Squared Error metric.



- ☐ 4
- ☐ 2
- ☐ 5
- ☒ 6
- ☐ 8
- ☐ 13

Find the entropy for the eight actual values of target variable in the train file. * 2 points

[0,0,0,1,1,1,1,1]

- ☐ $5/8 \log(5/8) + 3/8 \log(3/8)$
- ☐ $5/8 \log(3/8) - 3/8 \log(5/8)$
- ☒ $-(5/8 \log(5/8) + 3/8 \log(3/8))$
- ☐ $3/8 \log(5/8) + 5/8 \log(3/8)$



A classification model predicted true for a class which actual value was false. Then this is:

* 1 point

- ☐ False negative
- ☒ False positive
- ☐ True positive
- ☐ True negative

The true positive value is 20 and the false negative value is 5, find the value of recall:

2 points

- ☐ 0.6
- ☐ 0.7
- ☐ 0.5
- ☒ 0.8

Calculate the precision when the true positive value is 10 and the false positive value is 15.

2 points

- ☐ 0.5
- ☒ 0.4
- ☐ 0.6
- ☐ 0.8

The learning requires self assessment to identify patterns within data is: *

1 point

- ☐ supervised learning
- ☐ semi supervised learning
- ☒ unsupervised learning
- ☐ None



Major issue of Leave-One-Out-Cross-Validation (LOOCV) is? *

1 point

- ☐ faster runtime compared to k-fold cross validation
- ☐ slower runtime compared to normal validation
- ☒ low variance
- ☐ high variance

In Python, a syntax error is detected by the _____ at _____ *

2 points

- ☐ compiler/at compile time
- ☒ interpreter/at runtime
- ☐ compiler/at runtime
- ☐ interpreter/at compile time

What will be the output of the following piece code? *

1 point

```
str1 = 'hello'
str2 = ', '
str3 = 'world'

str1[-1:]
```

- ☐ olleh
- ☐ hello
- ☐ h
- ☒ o



What will be the output of the following piece code? *

1 point

```
1 veggies = ['carrot', 'broccoli', 'potato', 'asparagus']
2 veggies.insert(veggies.index('broccoli'), 'celery')
3 print(veggies)
```

- ☐ ['carrot', 'celery', 'broccoli', 'potato', 'asparagus']
- ☒ ['carrot', 'celery', 'potato', 'asparagus']
- ☐ ['carrot', 'broccoli', 'celery', 'potato', 'asparagus']
- ☐ ['celery', 'carrot', 'broccoli', 'potato', 'asparagus']

What will be the output of the following piece code? *

1 point

```
1 t = (1, 2, 4, 3, 8, 9)
2 print([t[i] for i in range(0, len(t), 2)])
```

- ☐ [2, 3, 9]
- ☐ [1, 2, 4, 3, 8, 9]
- ☐ [1, 4, 8]
- ☐ (1, 4, 8)

What will be the output of the following piece code? *

1 point

```
1 a = {1: "A", 2: "B", 3: "C"}
2 for i, j in a.items():
3     print(i, j, end=" ")
```

- ☐ 1 A 2 B 3 C
- ☐ 1 2 3
- ☐ ABC
- ☐ 1."A" 2."B" 3."C"



What will be the output of the following piece code? *

5 points

```
1 class A:
2     def __init__(self):
3         self._x = 5
4
5 class B(A):
6     def display(self):
7         print(self._x)
8
9 def main():
10     obj = B()
11     obj.display()
12
13 main()
```

Your answer

The effectiveness of an SVM depends upon:

1 point

- ☐ Selection of Kernel
- ☐ Kernel Parameters
- ☐ Soft Margin Parameter C
- ☐ All of the above

Which one is correct? *

1 point

- ☐ The most positively correlated features are good features
- ☐ The most negatively correlated features are good features
- ☐ The features with correlation zero are good features
- ☐ A, B

Which of these about a set is not true? *

1 point

- ☐ Mutable datatype
- ☐ Allows duplicate values
- ☐ Data type with unordered values
- ☐ mmutable data type



What is the output of the following code? *

2 points

```
class test:

    def finite(self,a):

        self.a=a

    def display(self):

        print(self.a)

obj=test()

obj.display()
```

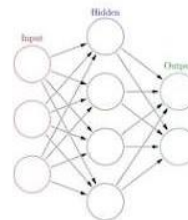
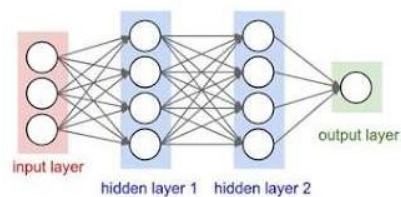
- ☐ Runs normally, doesn't display anything
- ☐ Displays 0, which is the automatic default value
- ☐ Error as one argument is required while creating the object
- ☐ Error as display function requires additional argument

1. For a binary classification problem, which of the following architecture would you choose?

* 2 points

Left figure with 2 hidden layers is (a)

Right figure with 1 hidden layer is (b)



- ☐ a
- ☐ b
- ☐ Any of these
- ☐ None



Increasing the dimensionality of our data always decreases our misclassification rate

* 1 point

- ☐ TRUE
- ☐ FALSE

Importance of using PCA before the clustering? Choose the most complete answer. * 1 point

- ☐ Find the explained variance
- ☐ Avoid bad features
- ☐ Find good features to improve your clustering score
- ☐ Find which dimension of data maximize the feature variance

Following the steps to run a PCA's algorithm, why is so important standardize your data?

* 1 point

- ☐ Standardize data allows other people understand better your work
- ☐ Use the best practices of data wrangling
- ☐ Find the features which can best predicts Y
- ☐ Make the training time more fast

It is not necessary to have a target variable for applying dimensionality reduction algorithms.

* 1 point

- ☐ True
- ☐ False



Which of the following techniques would perform better for reducing dimensions of a data set?

* 1 point

- ☐ Removing columns which have too many missing values
- ☐ Removing columns which have high variance in data
- ☐ Removing columns with dissimilar data trends
- ☐ None of these

If you have J input features and J target feature in a machine learning problem. You have to select 100 most features based on the relationship between input features and the target features.

Do you think, this is an example of dimensionality reduction?

- ☐ Yes
- ☐ No

Identify the type of a decision tree_____*

1 point

- ☐ Categorical
- ☐ Continuous
- ☐ Both a and b
- ☐ None

_____ is performed for trimming the tree in Decision trees.*

1 point

- ☐ Stemming
- ☐ Pruning
- ☐ Defining threshold

The measure of the randomness in the information being processed in the Decision Tree by *

1 point

- ☐ Entropy
- ☐ Information gain



Find the correct statement from below: *

1 point

- ☐ A decision tree is a graphical representation of all the possible solutions to a decision based on certain conditions.
- ☐ Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- ☐ A decision tree model consists of a set of rules for dividing a large heterogeneous population into smaller, more homogenous (mutually exclusive) classes.
- ☐ All of the above.
- ☐ None of the above

Select the correct sequence of typical decision tree structure down below: * 2 points

- (I) Take the entire data set as input
- (II) Divide the input data into two part
- (III) Reapply the split to every part recursively
- (IV) Stop when meeting desired criteria
- (V) Cut the tree when we went too far while doing splits

(I, II, V), (V), G5

- ☐ (V), (I), (III), (II), (IV).
- ☐ (I), (II), (IV), (V)
- ☐ (I), (II), (V), (IV)



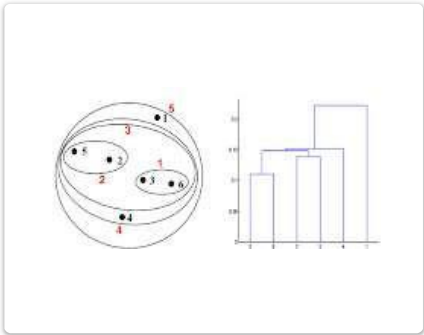
Given, six points below select the cluster representations and dendrogram * 5 points using complete link proximity function.

point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points



☐ Option 1



☐ Option 2



☐ Option 3



☐ Option 4



High entropy means that the partitions in classification are? *

1 point

- ☐ pure
- ☐ not pure
- ☐ useful
- ☐ useless

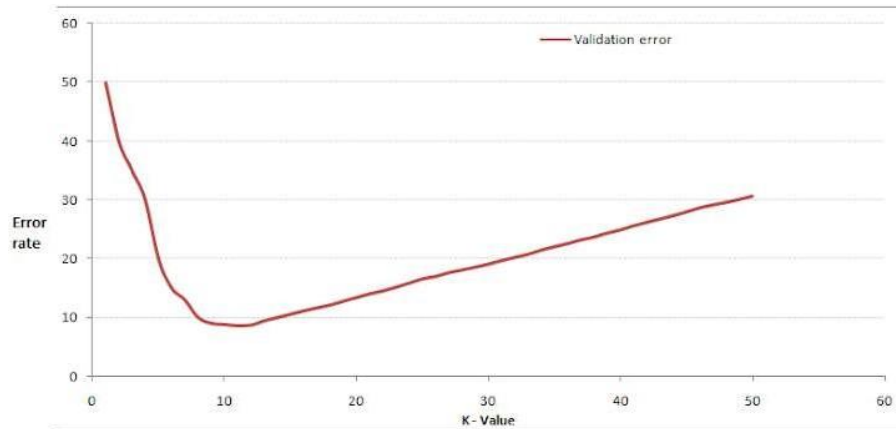
While performing regression/classification, which of the following is the correct sequence to process the data?

* 1 point

- ☐ Normalize the data — PCA — training
- ☐ PCA —+ normalize PCA output — training
- ☐ Normalize the data —+ PCA —+ normalize PCA output —+ training
- ☐ None of the above

Given the image below, which would be the best value for K assuming that the model you are using is KNN. *

1 point



- ☐ 3
- ☐ 20
- ☐ 5
- ☐ 10
- ☐ 7

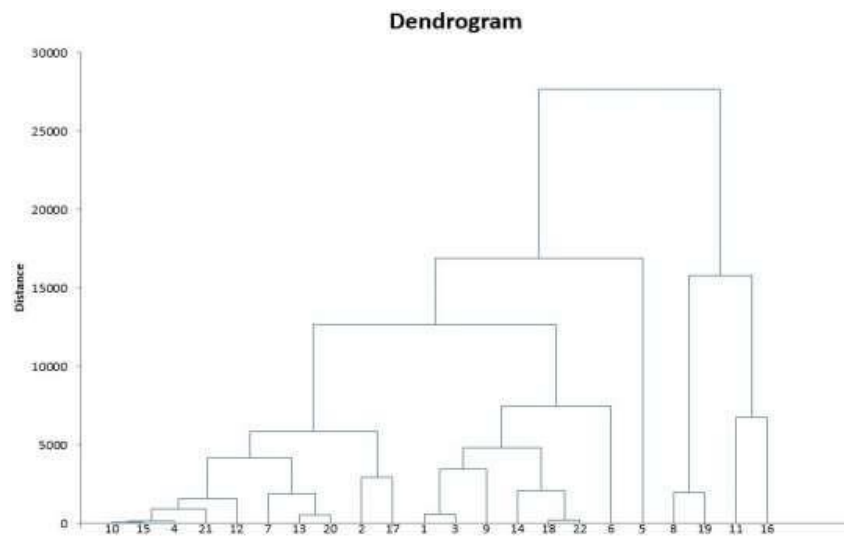


Suppose, you have given the following data where x and y are the 2 input variables and Class is the dependent variable. Suppose, you want to predict the class of new data point $x=1$ and $y=1$ using Euclidean distance in 7NN. In which class this data point belong to? * 5 points

x	y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+

- ☐ + Class
- ☐ - Class
- ☐ Can't say
- ☐ None of these

Consider the dendrogram below and find the number of clusters if threshold is defined on 10000 * 1 point



- ☐ 7
- ☐ 3
- ☐ 4
- ☐ 5

Given are five data points having two attributes x and y.

* 2 points

The distance matrix of the points, indicating the Euclidean distance between points. Use hierarchical clustering with minimum distance function to answer the question.

a) How many clusters are there initially (before any fusions have happened)?

Observation	x	y
1	3	2
2	3	5
3	5	3
4	6	4
5	6	7

Label	1	2	3	4	5
1	0.00	3.00	2.24	3.61	5.83
2	3.00	0.00	2.83	3.16	3.61
3	2.24	2.83	0.00	1.41	4.12
4	3.61	3.16	1.41	0.00	3.00
5	5.83	3.61	4.12	3.00	0.00

☐ 4

☐ 3

☐ 7

☒ 5

☐ 1

b) Which two clusters would be fused first? (refer to above figure) *

2 points

☐ 1 and 2

☐ 3 and 4

☐ 1 and 5

☐ 3 and 5



Which clusters will be fused in last iteration? (refer to above figure) *

2 points

- ☐ 1 fuses with (2, 3, 4, 5)
- ☐ 2 fuses with (1, 3, 4, 5)
- ☐ 5 fuses with (1, 2, 3, 4)
- ☐ 4 fuses with (1, 2, 3, 5)

Select the correct statement(s) from the following *

1 point

- ☐ The clusters formed by k-means algorithm do not depend on the initial selection of cluster centers.
- ☐ The results of k-means algorithm get impacted by outliers and range of the attributes.

K-means clustering automatically selects the most optimum value of k

k-means algorithm can be applied to both categorical and numerical variables.

Select the centroid of the following 5 data points *

2 points

X	Y	Z
12	23	45
31	31	31
17	15	25
19	27	45
13	11	27

- ☐ 18.4, 21.4, 32.1
- ☐ 18.4, 21.4, 34.6
- ☐ 34.6, 21.4, 18.4
- ☐ 21.4, 32.1, 18.4



Select the clustering task *

1 point

- ☐ A dataset containing various features of ACs - length, width, height, split/window is given. A new design of AC is introduced in the market. The algorithm needs to predict whether the new AC is a split/window AC given its dimensions.
- ☒ A baby is given some toys to play. These toys consist of various animals, vehicles and houses, but the baby is unaware of these categories. The baby chooses different toys and starts making different groups with the toys based on what he feels are similar toys.

The number of input dimensions are equal to principal components in PCA *

1 point

- ☒ True
- ☐ False
- ☐ Not relevant

In a classification system, a patient with the disease is *
classified as not having the disease.

1 point

This is an example of a:

- ☐ True positive
- ☐ True negative
- ☒ False positive
- ☐ False negative

Disadvantage(s) of Naive Bayes Classifier *

1 point

- ☒ Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.
- ☐ It performs well in Multi-class predictions as compared to the other Algorithms.
- ☐ Naive Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- ☐ It is the most popular choice for text classification problems.



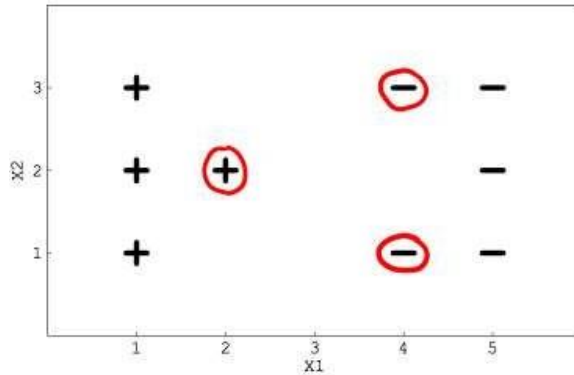
The benefit of Naive Bayes includes: *

1 point

- ☐ NaTve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- ☐ It is the most popular choice for text classification problems.
- ☐ It can be used for Binary as well as Multi-class Classifications.
- ☐ Alloftheabove

Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors. * 2 points

a) If you remove the following any one red points from the data. Does the decision boundary will change?



- ☐ Yes
- ☐ No
- ☐ N/A

b) If you remove the non-red circled points from the data, the decision boundary will change? (refer above figure) * 2 points

- ☐ Yes
- ☐ No



You trained a binary classified model which gives very high accuracy on the training data, but much lower accuracy on validation data. The following may be true: * 2 points

This is an instance of overfitting

This is an instance of underfitting

The training was not well regularized

The training and testing examples are sampled from different distributions

Using the kernel trick, one can get non-linear decision boundaries using algorithms designed originally for linear models. * 1 point

☒ True

☐ False

Which among sensitivity, accuracy and specificity is the highest for the model below? * 2 points

Actual/Predicted	Not Churn	Churn
Not Churn	80	40
Churn	30	50

☐ Accuracy

☐ Specificity

☒ Sensitivity



Computer the accuracy of matrix below: *

2 points

Actual/Predicted	Not Churn	Churn
Not Churn	80	30
Churn	20	70

- ☐ 70%
- ☐ 65%
- ☐ 80%
- ☐ 75%

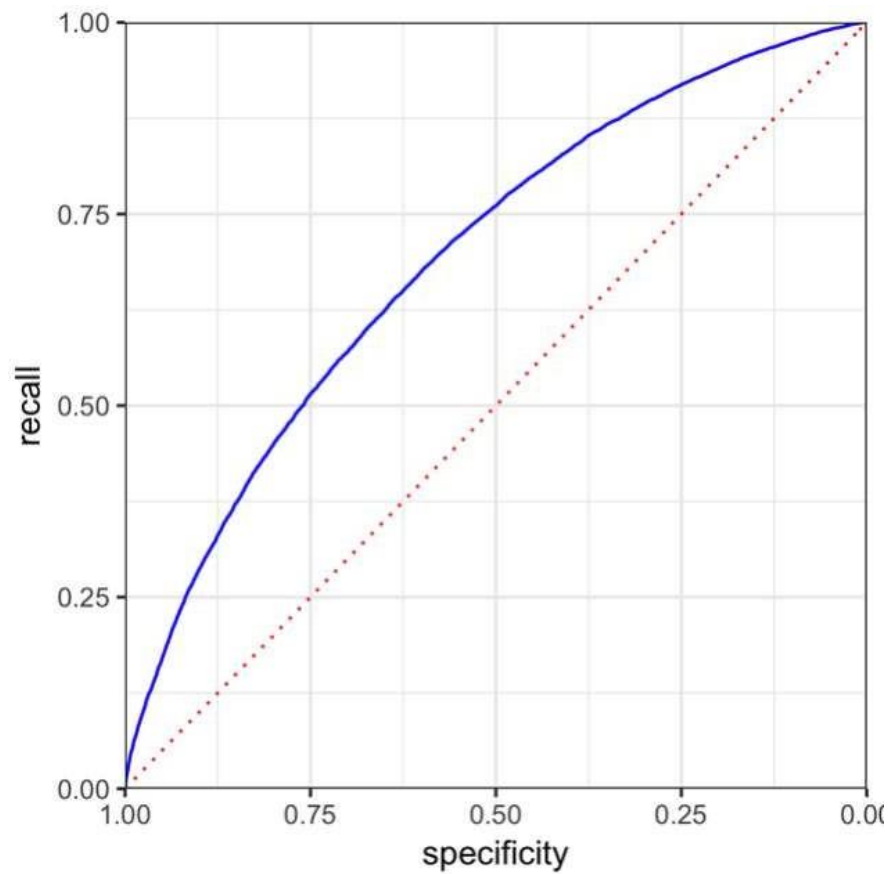
Which option is correct about the ROC curve? *

1 point

- ☐ You can see that there is a tradeoff between recall and specificity
- ☐ The ROC curve plots sensitivity (recall) on the y-axis against specificity on the x-axis
- ☐ The ideal classifier would classify the 1s without misclassifying more 0s as 1s
- ☒ The ROC curves are especially useful in evaluating data with highly unbalanced outcomes
- ☐ All of the above



The plot below is a ROC curve of a model. What is the recall in this case if we want a classifier with a specificity of at least 50%? * 1 point

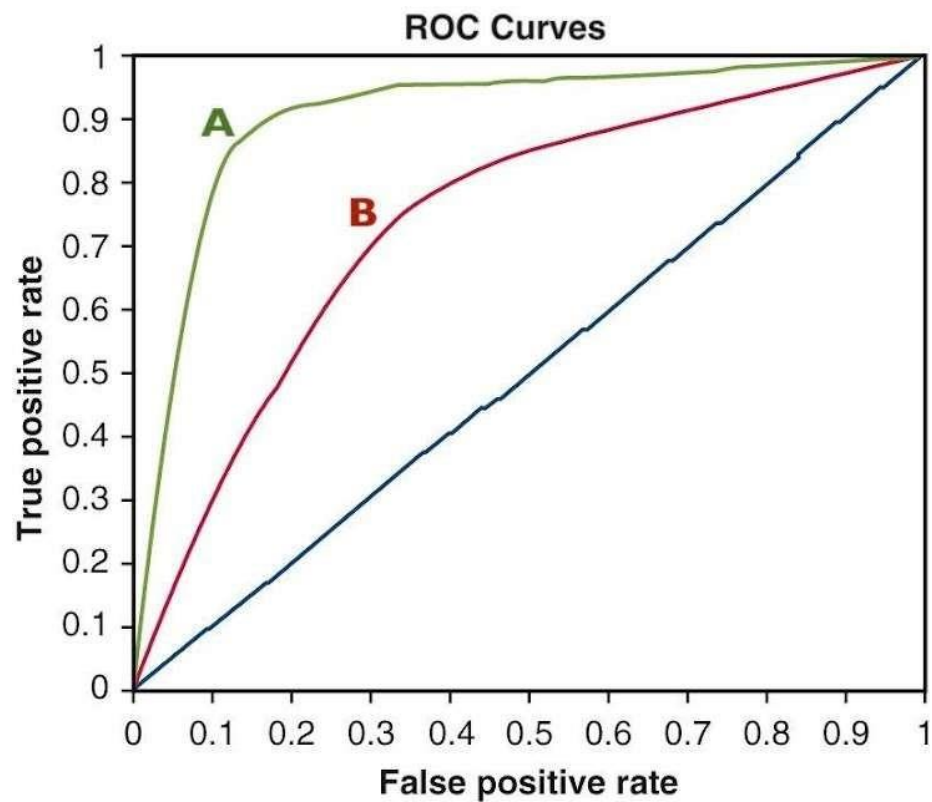


- ☐ 50%
- ☐ 25%
- ☐ 100%
- ☐ 75%
- ☐ 0%



According to the ROC curve and AUC below, which answer is correct in terms of an effective classifier?

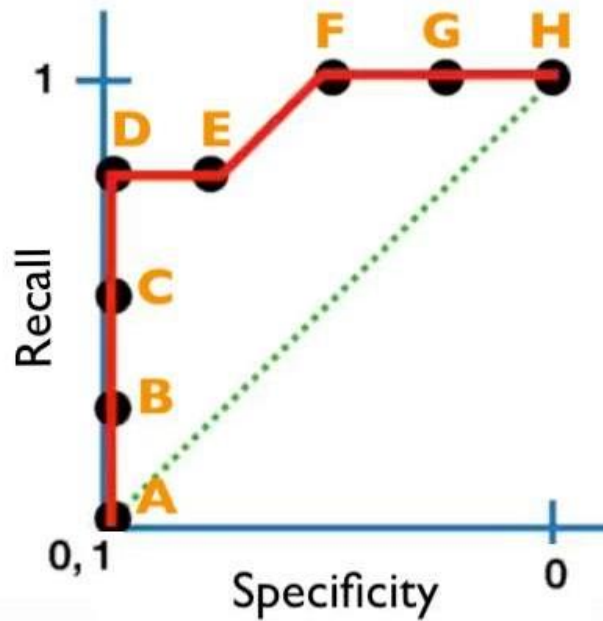
* 1 point



- ☐ A=B
- ☐ A<B
- ☐ A»B
- ☐ None of the above



Costco is using a Computer Vision application to detect whether customers * 1 point wear face masks or not. You as a Data Scientist, got a ROC curve and AUC of the classifier. Our goal is to get as high as possible on successfully detecting customers who do not wear face masks (True Positive) and get as low as possible on False Negative. According to the goal, which threshold on the ROC curve should we take?



- ☐ A
- ☐ B
- ☐ C
- ☐ D
- ☐ E
- ☐ F
- ☐ G
- ☐ H

What will be output of following code ? *

2 points

```
import pandas as pd
data = [['Anuj',21],['Rama',25],['Kapil',22]]
df = pd.DataFrame(data,columns=['Name','Age'])
print (df)
```

	Name	Age
0	Anuj	21
1	Rama	25
2	Kapil	22

☐ Option1

☐ Option2

☐ Option3

☐ Option4

The execution of following code will result in: *

1 point

```
import pandas as pd
s =pd.Series([1,2,3,4,5],index= ['a','b','c','d','e'])
print(s['f'])
```

- ☐ KeyError
- ☐ IndexError
- ☐ ValueError
- ☐ None of the above mentioned



Dataframe object is value mutable. *

1 point

- ☐ True
- ☐ False

Suppose that you are given two lists: *

1 point

a = [1,2,3]

b = [4,5,6]

Your task is to create a list which contains all the elements of a and b in a single dimension.

Output:

a = [1,2,3,4,5,6]

Which of the following functions will you use?

- ☐ a.append(b)
- ☐ a.extend(b)
- ☐ any one of the above
- ☐ none of the above

What will be output for the following code? *

1 point

```
import pandas as pd
import numpy as np
s = pd.Series(np.random.randn(4))
print s.ndim
```

- ☐ 0
- ☐ 1
- ☐ 2
- ☐ 3



Which of the following is not a method to create a chart using pyplot? *

1 point

- ☐ pie()
- ☒ histg()
- ☐ plot()
- ☐ barh()

To import pyplot module we can write *

1 point

- ☐ mport matplotlib.pyplot
- ☐ Import matplotlib.pyplot as plt
- ☐ Import pyplot as pit
- ☒ Both (A) and (B)

By default, Plot() function plots a *

1 point

- ☐ Bar chart
- ☒ Linechart
- ☐ Pie chart
- ☐ Horizontal barchart



What is the output of the following code? *

1 point

```
def outer_fun(a, b):  
    def inner_fun(c, d):  
        return c + d  
    return inner_fun(a, b)  
  
res = outer_fun(5, 10)  
print(res)
```

- ☐ Syntax Error
- ☐ is
- ☐ (5, 0)

What is the output of the add() function call *

1 point

```
def add(a, b):  
    return a+5, b+5  
  
result = add(3, 2)  
print(result)
```

- ☐ is
- ☐ 8
- ☐ (8, 7)
- ☐ Syntax Error



What is the output of the following function call *

1 point

```
def outer_fun(a, b):  
    def inner_fun(c, d):  
        return c + d  
  
    return inner_fun(a, b)  
    return a  
  
result = outer_fun(5, 10)  
print(result)
```

☐ S

15

☐ (15,5)

Syntax Error

In dataset loading, The variables of data are called its? *

1 point

☐ Response

Features

Target

☐ Vector

Su mit

[Clear form](#)

Never submit passwords through Google Forms.

This form was created inside of University of Central Missouri. [Report Abuse](#)

Google Forms

