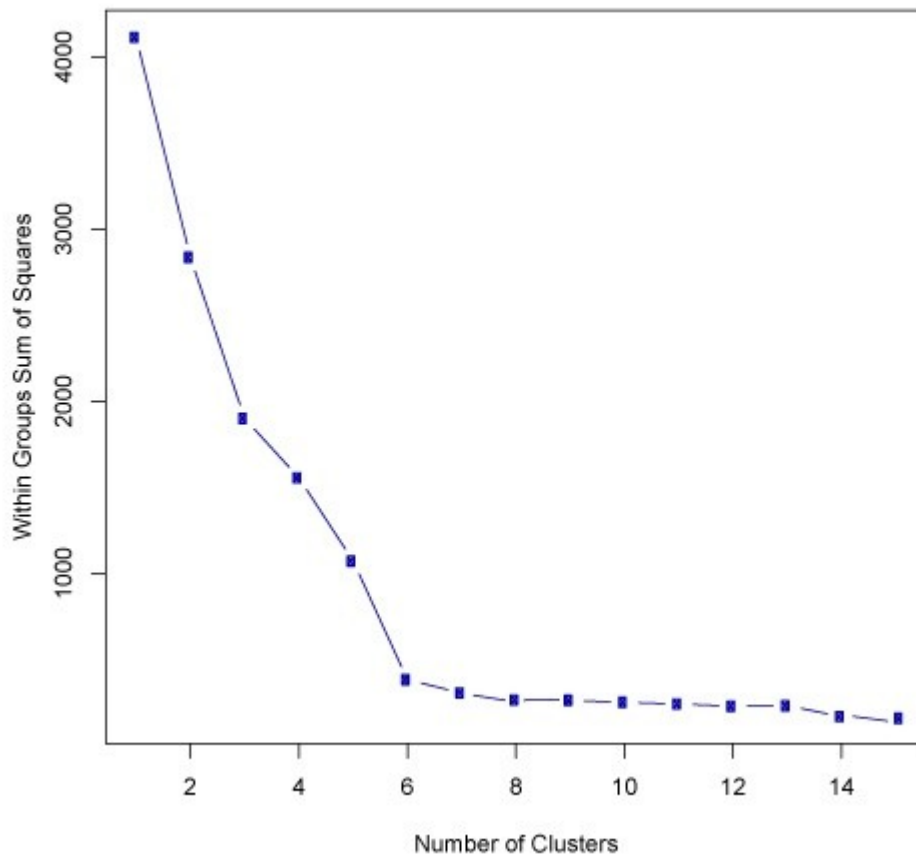


DSexam

DATA SCIENCE - PYTHON - EXAM 1

1. L2 regularization known as (1 point)
- Ridge Regression
 - Lasso Regression
 - Both of the above
 - None of the above
2. What is/are True for Supervised Learning (1 point)
- Label is mandatory
 - Label is not mandatory
 - Easy to execute program
 - Most difficult to execute, because we have to spend more time on Supervising.
3. Assume, you want to cluster 7 observations into 3 clusters using K-Means clustering algorithm. After first iteration clusters, C1, C2, C3 has following observations: (1 point)
- C1: {(2,2), (4,4), (6,6)}
- C2: {(0,4), (4,0)}
- C3: {(5,5), (9,9)}
- What will be the Manhattan distance for observation (9, 9) from cluster centroid C1. In second iteration.
- 10
 - $5\sqrt{2}$
 - $13\sqrt{2}$
 - None of these
4. (1 point)
- What should be the best choice for number of clusters based on the following results:



stion

- ☐ 5
- ☒ 6
- ☐ 14
- ☐ Greater than 14

5. How would you read data from the file using pandas by skipping the first three lines?

(1 point)

- ☐ `read_csv('email.csv', skip_rows=3)`
- ☒ `read_csv('email.csv', skiprows=3)`
- ☐ `read_csv('email.csv', skip=3)`
- ☐ None of these

6. Correlation of -0.80 means

(1 point)

- ☐ Very weak correlation
- ☐ Strong Correlation
- ☐ Weak Correlation and Negative

- Strong Positive Correlation
- Strong Negative Correlation

7. Suppose there is a dataset having variables with missing values of more than 30%, how will you deal with such a dataset? (1 point)

- ☐ Reject the file, and ask for Good data file with at least 90% quality
- Fill Missing Values with Mean
- Fill Missing Value with Median
- Drop Missing value Records.

8. Which of the following method is used for finding optimal of cluster in K-Mean algorithm? (1 point)

- Elbow method
- Manhattan method
- Euclidean method
- All of the above

9. How can you avoid overfitting your model? (1 point)

- Keep the model simple
- cross-validation techniques,
- regularization techniques
- All the above
- None of the above

10. What does the ROC Curve represent ? (1 point)

- Graphical representation of the contrast between false-positive rates and true positive rates at different thresholds.
- Graphical representation of the contrast between True-positive rates and true positive rates at different thresholds.
- graphical representation of the contrast between false-positive rates and false positive rates at different thresholds.
- None of the above

11. `import numpy as np` (1 point)

```
arr = np.array([[[1, 2, 3], [4, 5, 6]], [[1, 2, 3], [4, 5, 6]]])  
print(arr)
```

This creates _____ dimensions of array?

- ☐ 1
- ☐ 2
- ☒ 3
- ☐ 4

12. `def fun(x):` (1 point)
`x[0] = 5`
`return x`

`g = [10,11,12]`

`print(fun(g), g)`

- ☒ [5, 11, 12] [5, 11, 12]
- ☐ [5, 11, 12] [10, 11, 12]
- ☐ [10, 11, 12] [10, 11, 12]
- ☐ [10, 11, 12] [5, 11, 12]

13. After studying the behavior of a population, you have identified four specific individual types that (1 point)
are valuable to your study. You would like to find all users who are most similar to each
individual type. Which algorithm is most appropriate for this study?

Choose the correct option:

- ☒ K-means clustering
- ☐ Linear regression
- ☐ Association rules
- ☐ Decision trees

14. What is pass in Python? (1 point)

- ☒ no-operation Python statement
- ☒ place holder in compound statement, where nothing has to be written
- ☐ Passive Data split
- ☐ All the above

15. Which of the below method depicts hierarchical data in nested format? (1 point)

- ☐ Scatter plots
- ☐ Population pyramids
- ☐ Area charts

- Treemaps

16. what is the purpose of ZIP function (1 point)
- We can zip all files
 - we can zip two or more lists
 - we can use for string concatenation
 - None of the above
17. What are the differences between correlation and covariance? (1 point)
- correlation is dimensionless whereas covariance is represented in units
 - $\text{correlation}(X, Y) = \text{covariance}(X, Y) / (\text{covariance}(X) \text{ covariance}(Y))$
 - No difference between them
 - correlation is only for vector analysis
18. ARIMA technique used in (1 point)
- Auto linear Regression
 - Auto Responsive Integrated machine Access
 - Time series
 - Kernel found in Support Vector Machine
19. Seaborn module is for ? (1 point)
- Support Vector
 - Ocean Summary Analysis
 - Graphics
 - Matrix Calculation
20. How do you reverse a string in Python? (1 point)
- Stringname = 'python'
- Stringname[::-1]
 - Stringname[::1]
 - Stringname[::2]
 - None of the above

21. How would you import a decision tree classifier in sklearn? (1 point)
- ☐ from sklearn.decision_tree import DecisionTreeClassifier
 - ☐ from sklearn.ensemble import DecisionTreeClassifier
 - ☒ from sklearn.tree import DecisionTreeClassifier
 - ☐ None of the above
22. write a program in Python, that prints the numbers ranging from one to 50 (1 point)
- ☒ range(1,51)
 - ☐ range(1,50)
 - ☐ range(1,49)
 - ☐ None of the above
23. (1 point)
- Given the information that the demand is 100 in October 2020, 150 in November 2020, 350 during December 2020 and 400 during January 2021. Calculate a 3-month simple moving average for February 2021.
- ☒ 300
 - ☐ 250
 - ☐ 400
 - ☐ Cannot calculate using this information
24. from numpy import random
x = random.rand()
print(x) (1 point)
- this creates _____ random numbers?
- ☒ Between all values of 0 and 1
 - ☐ Random number of any value
 - ☐ Both above
 - ☐ none of the above
25. Imbalanced Data means? (1 point)
- ☒ distributed unequally across different categories
 - ☐ distributed equally across different categories
 - ☐ Both of the above

☐ none of the above

26. `import time` (1 point)
`str = '21/01/2017'`
`datetime_value = time.strptime(str,date_format)`

To convert the above string, what should be written in place of `date_format`?

- ☐ “%d/%m/%y”
- ☐ “%D/%M/%Y”
- ☐ “%d/%M/%y”
- ☒ “%d/%m/%Y”

27. Commonly used module for Graphical Outputs in Python (1 point)

- Matplotlib
- seaborn
- Plotly
- pandas

Explanation:

Pandas also have graphic module for Eg:Bar Plot

28. What would be the formula representation of this problem in terms of x and y variables: “The price of 2 pens and 1 pencil as 10 units”? (1 point)

- ☐ $2x + y = 200$
- ☐ $2x+2y=10$
- ☒ $2x + y = 10$
- ☐ $x + 2y = 10$

29. Standard Deviation means, square Root of Variance (1 point)

- ☒ True
- ☐ False

30. What is TRUE for Cross-Validation? (1 point)

- Cross-Validation is a Statistical technique used for improving a model’s performance
- Entire Dataset will be used in Training
- Entire Data set used in Testing
- ☐ 70% Records for Training, 30 % Records for Testing

31. Purpose of RANGE function

(1 point)

- ☒ returns a sequence of numbers
- ☐ returns a random sequence of numbers
- ☐ returns length of list
- ☐ returns length of Large Range String

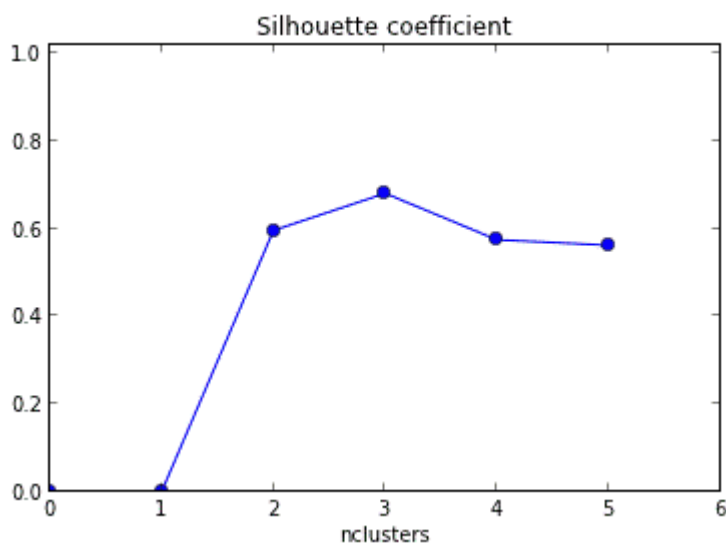
32. PCA technique is for

(1 point)

- ☐ Increase more dimensionality
- ☒ Reduce Dimensionality
- ☐ Both above
- ☐ None of the above

33. What should be the best choice of no. of clusters based on the following results:

(1 point)



- ☐ 1
- ☐ 2
- ☒ 3
- ☐ 4

34. Feature scaling is an important step before applying K-Mean algorithm. What is reason behind this?

(1 point)

- ☒ In distance calculation it will give the same weights for all features
- ☐ You always get the same clusters. If you use or don't use feature scaling
- ☐ In Manhattan distance it is an important step but in Euclidian it is not
- ☐

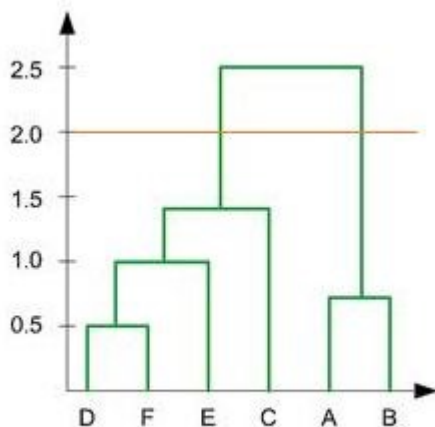
None of these

35. What is a confusion matrix?

(1 point)

- ☒ Binary Classifier
- ☐ Cluster
- ☐ Confusion of data
- ☐ All the above

36. In the figure below, if you draw a horizontal line on y-axis for $y=2$. What will be the number of clusters formed? (1 point)



- ☐ 1
- ☒ 2
- ☐ 3
- ☐ 4

37. Which of the following can act as possible termination conditions in K-Means?

(1 point)

1. For a fixed number of iterations.
2. Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
3. Centroids do not change between successive iterations.
4. Terminate when RSS falls below a threshold.

- ☐ 1, 3 and 4
- ☐ 1, 2 and 3
- ☐ 1, 2 and 4
- ☒ All of the above

38. Difference Between List and Tuple

(1 point)

- ☐ Both are same, except size of members in List is unlimited
- ☒ List is mutable and Tuple is immutable
- ☐ Tuple can be modified dynamically by code
- ☒ We cannot insert, update in Tuple through code

39. `import numpy as np`

(1 point)

```
arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12])  
newarr = arr.reshape(X, Y)
```

If we want 3 columns and 4 rows, then X and Y will be

- ☒ 4,3
- ☐ 3,4
- ☐ Both A and B
- ☐ None of the Above

Explanation:

Rows By Column

40. Consider a case where you know the probability of finding at least one shooting star in a 15-minute interval is 30%. Evaluate the probability of finding at least one shooting star in a one-hour duration?

(1 point)

- ☐ 75%
- ☒ 85%
- ☐ 95%
- ☐ 55%

Explanation:

We know that,

Probability of finding atleast 1 shooting star in 15 min = $P(\text{sighting in 15min}) = 30\% = 0.3$

Hence, Probability of not sighting any shooting star in 15 min = $1 - P(\text{sighting in 15min}) = 1 - 0.3 = 0.7$

Probability of not finding shooting star in 1 hour = $0.7^4 = 0.1372$

Probability of finding atleast 1 shooting star in 1 hour = $1 - 0.1372 = 0.8628$

41. p-value of statistical test is 0.61 means

(1 point)

- ☐ Very Good Result and significant
- ☒ Very Good result for Non-significant
- ☐ Both above
- ☐ None of the above

42. sklearn._____.GradientBoostingClassifier , Fill the blank (1 point)

- ☒ ensemble
- ☐ randomForest
- ☐ Boosting
- ☐ Bagging

43. What are the model parameters that are used to build ML models using iterative methods under model-based learning methods? (1 point)

- ☐ mini-batches
- ☐ elasticity
- ☒ hyperparameters
- ☐ optimizers

44. Keras is for (1 point)

- ☐ K- Nearest Neighbour Model
- ☐ K- means Algorithm
- ☒ Deep Learning Module
- ☐ Logistic Regression Modified

45. Random Forest useful in? (1 point)

- ☐ Classification
- ☐ Regression
- ☒ Both Classification and Regression
- ☐ None of the above

46. How you can convert a number to a string? (1 point)

- ☐ num()
- ☒ str()
- ☐ numstr()
- ☐ strnum()

47. KMeans algorithm is (1 point)
- ☐ Supervised
 - ☒ Unsupervised
 - ☐ Both supervised and unsupervised
 - ☐ None of the above
48. Regularisation? Why is it useful? (1 point)
- ☒ Prevent overfitting
 - ☐ Increase Overfitting
 - ☐ No such useful
 - ☐ Prevent underfitting
49. How will you treat missing values during data analysis? (1 point)
- ☒ Drop the missing Records
 - ☒ Fill missing Values
 - ☐ Do not fill missing values
 - ☐ Do not drop missing values
50. `b = "Python Coding Fun"`
`print(b[-1])` (1 point)
- ☐ P
 - ☒ N
 - ☐ Y
 - ☐ U