

Job Guidance Program 2024

assimilate.ai

Screening Test

Full Marks – 20 Marks

Time allocated – 1.5 hrs.

Good luck 😊

(2M)

Q.1 What is a model Learning Rate? We should choose a high learning rate for getting quick results on training data. Comment on the statement.

(2M)

Q.2 Explain One-Hot Encoding. Does the dimensionality of the dataset increase, decrease or remains constant after applying this technique?

(4M)

Q.3 A financial institution has just hired you to build a system which will decide what car insurance package to offer to different clients. The information recorded about the clients are -gender, age group (under 20, 20-35, 35-55 and over 55), credit rating, occupation, number of accidents in the last year, car make, model, year and the city where the person lives.

The financial institution has a database of roughly 20000 current clients. Explain step by step how you would use a logistic regression approach to build the system.

Your answer should contain:

- Type of problem you are dealing with
- How would you cater the categorical data e.g. occupation, car make, model, city etc using them as features in your logistic regression model
- Other relevant details

(2M)

Q.4 Write the steps in designing an Unsupervised Machine Learning model. How do we change the penalty parameter in case we have a condition of underfitting using regularization technique?

(1M)

Q.5 What would be the expected output of the follow code snippet?

```
import numpy as np
from sklearn.cluster import KMeans

X = np.array([[1, 2], [1, 4], [1, 0], [10, 2], [10, 4],
              [10, 0]])

kmeans = KMeans(n_clusters=2)

kmeans.fit(X)
```

(2M)

Q.6 For the given set of data points, identify the clusters and construct a dendrogram based on agglomerative clustering technique.

Datapoints	x	y
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

Consider that you have been hired as a Data Scientist for Zomato, the Indian multinational restaurant aggregator and food delivery company, assume the following scenario and visualize the questions that follow by using test_zomato.csv dataset:

(4 M)

Q.7 Out of the first 10 restaurants,

1. Which city or cities have the most expensive restaurants?
2. Which city or cities have the cheapest restaurants?
3. Which city or cities have the highest rated restaurants?
4. Which restaurant or restaurants do not provide table booking facility?

(3 M)

Q.8 For Indian restaurants in between index numbers 1400 to 1405(excluding),

1. Which restaurant has less than 4/5 rating?
2. Which restaurant is the most expensive?
3. Which restaurant is the cheapest?

Created by:

Aashutosh K Mishra

Sr. Data Science Consultant, Deloitte USI

Advisor to Govt. of India

-----THE END-----

