

Let
100

Class Assessment - 3

Section A : Python Basic

Q. 1)

List

Tuples

- | | |
|--|---|
| <ul style="list-style-type: none"> 1. List are mutable 2. Iterations are time-consuming 3. Lists consume more memory. 4. List have several built-in methods. | <ul style="list-style-type: none"> 1. Tuples are immutable 2. Iterations are comparatively faster. 3. Tuple consumes less than the list 4. A tuple does not have many built-in methods because of immutability. |
|--|---|

Q. 2)

Python funⁿ to calculate Factorial of a given number.

```

import math
def fact(n):
    return (math.factorial (n))
num = int (input ("Enter the number :"))
f = fact (num)
print ("Factorial of ", num, "is", f)
  
```

O/P = Enter the number : 6

Factorial of 6 is 720

Q. 3)

List Comprehension

- 1) List comprehension offers a shorter syntax

when you want to create a new list based on the values of an existing list.

2) Eg. Based on a list of fruits, you want a new list, containing only the fruits with the letter "a" in the name.

3) with list comprehension you can do all with only one line of code:

```
fruits = ["apple", "banana",
```

```
"cherry", "kiwi", "mango"]
```

```
newlist = [x for x in fruits if "a"  
          in x]
```

```
print(newlist)
```

c. 4) a) NumPy : NumPy can be used to perform a wide variety of mathematical operations on array.

2) NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.

b) i) Pandas : Pandas is most commonly used for data wrangling & data manipulation purposes.

2) Pandas is used for creating heterogenous, two-dimensional data objects.

c) Matplotlib : matplotlib is a comprehensive

library for creating static, animated & interactive visualizations in python.

2) It provides an object-oriented API for embedding plots into appn using general purpose GUI toolkits like Tkinter.

Section B: Machine Learning, Basics

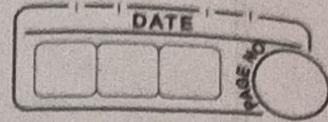
Q.1)

A) Supervised learning:

- 1) In supervised learning, the machine is trained on a set of labeled data, which means that the input data is paired with the desired o/p.
- 2) The machine then learns to predict the o/p for new input data.
- 3) Supervised learning is often used for tasks such as classification, regression, & object detection.

B) Unsupervised learning:

- 1) In this machine is trained on a set of unlabeled data, which means that the input data is not paired with the desired o/p.
- 2) The machine then learns to find patterns & relationships in the data.
- 3) It used for tasks such as clustering, dimensionality reduction & anomaly detection.



Q. 2)

- 1) In machine learning, as you try to minimize one component of the error (eg bias), the other component (eg variance) tends to increase & vice versa,
- 2) Finding the right balance of bias & variance is key to creating an effective & accurate model.
- 3) A high level of bias can lead to under-fitting, which occurs when the algorithm is unable to capture relevant relations b/w features & target output.
- 4) A high-bias model typically includes more assumptions about the target funⁿ or end result.

Q. 3)

- 1) The steps in a machine learning pipeline include Data ingestion, Data validation, Data pre-processing, Model training & tuning, Model analysis, Model versioning, Model deployment, feedback loop.
- 2) Data Ingestion : In this step, the data is processed into a well-organized format, which could be suitable to apply for further steps.
- 3) Data Validation : It is required to perform before training a new model. It focuses on statistics of new data.
- 4) Data Pre-processing : The pre-processing step involves preparing the raw data & making it suitable for the ML model. The process

includes different sub-steps, such as data cleaning, feature scaling, etc.

5) Model Training & Tuning : In this step, the model is trained to take the input & predicts an OLP with highest possible accuracy.

6) Model Analysis : After model training, we need to determine the optimal set of parameters by using the loss of accuracy metrics.

7) Model versioning : The model versioning steps keeps track of which model, set of hyper-parameters & datasets have been selected as the next version to be deployed.

8) Model Deployment : After training & analyzing the model, it's time to deploy the model. An ML model can be deployed in three ways, which are:

- using the model server
- In a browser
- on edge device

Q. 4)

A) Cross-validation :

1) Cross-validation is a technique for evaluating ML models by training several ML models on subsets of the available input data & evaluating them on the complementary subset of the data.

2) It is used important in ML because it helps to detect overfitting, i.e. failing to generalize a pattern.

3) It's used in ML to evaluate the performance of a model on unseen data.

4) Eg. K-fold cross-validation.

In this technique, the whole dataset is partitioned in k parts of equal size & each partition is called a fold.

It's known as K -fold since there are k parts where k can be any integer - 3, 4, 5, etc. One fold is used for validation & other $k-1$ folds are used for training the model.

Regression

Classification

1. In this problem statement, the target variables are continuous.

1. In this problem statement, the target variables are ^{discrete} continuous.

2. Problems like House Price Prediction, Rainfall Prediction like problems are solved using regression Algorithms.

2. Problems like Spam Email classification, Disease prediction like problems are solved using Classification Algo.

3. Input data are independent variables of continuous dependent variable

3. Input Data are Independent variables & categorical dependent variable

4. Output is continuous numerical values.

4. Output is categorical labels.

5. Eg. use cases are stock price prediction, house price prediction.

5. Eg. use cases are spam detection, image recognition.

6. Eg. of regression algo. are Linear Regression, Polynomial Regression, Ridge Regression etc.

6. Eg. of classification algo. are: Logistic Regression, Decision trees, Random forest, SVM, K-NN etc.

Selection C: Statistics & Probability.

- 1) Measures of central tendency help you find the middle, or the average, of a dataset.
- 2) The 3 most common measures of central tendency are the mode, median & mean.
- 3) Mode is the most frequent value.
- 4) Median is the middle number in an ordered dataset.
- 5) Mean is the sum of all values divided by the total number of values.
- 6) In addition to central tendency, the variability of distribution of your dataset is important to understand when performing descriptive statistics.

Q. 3)

- 1) The P value is defined as the probability under the assumption of no effect or no difference (null hypothesis), of obtaining a result equal to or more extreme than what was actually observed;
- 2) The P stands for probability & measure how likely it is that any observed diff. betw group is due to chance.
- 3) In statistical hypothesis testing, P-value can be defined as the measure of probability that real-valued test statistic is at least as extreme as the value actually obtained.
- 4) A P-value indicates the probability of getting an effect no less than that actually observed in the sample data.
- 5) P-values are used in statistical hypothesis testing to determine whether to reject the null hypothesis.
- 6) The smaller P-value, the stronger the likelihood that you should reject null hypothesis.

Q. 4)

- 7) Theoretically the diff. betw the two types of relationships are easy to identify - an action or occurrence can cause another.
(eg. smoking causes

Q.4)

- 1) Understanding the diff. betn correlation & causation is essential in data science & statistics.
- 2) Correlation refers to the statistical relationship betn two variables.
- 3) while causation is the relationship betn cause & effect.
- 4) while correlation can help identify patterns it does not imply causation.
- 5) Correlation is when two things happen together, while causation is when one thing causes another thing to happen.
- 6) For eg., you might say that there is a correlation betn ice cream sales & crime rates because you notice that they both seem to rise & fall together.

~~Q.5)~~ JM Stati Section D: Advanced Topics

Q.1)

- 1) Overfitting is an undesirable machine learning behavior that occurs when the ML model gives accurate predictions for training data but not for new data.
- 2) when data scientists use ML models for making predictions, they first train the model on a known data set.
- 3) Overfitting is a modeling error which occurs when func. is too closely fit to a limited set of data pts.

4) In this case, the ML model learns the details of noise in the training data such that it negatively affects the performance of model on test data.

5) It can be mitigated by using diff techniques such as train with more data, Data augmentation, feature selection, cross-validation, regularization etc.

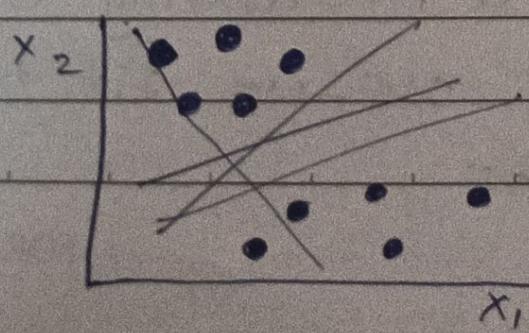
2)

1) SVM & SVR is a ML algorithm that uses supervised learning models to solve complex classification, regression & outlier detection problems by performing optimal data transformations that determine boundaries betn data pts based on predefined classes, labels or OIP.

2) SVM is a supervised ML algo used for both classification & regression.

3) The main objective of SVM algo is to find the optimal hyperplane in N-dimensional space that can separate data pts. in diff. classes in the feature space.

4) Let's consider two independent variable x_1, x_2 , & one dependent variable which is either a black circle or white circle.



- 5) SVM kernel functions : is to take data as I/P & transform it into the required form.
- 6) Diff. SVM algo. use diff. types of kernel funn. these funn. can be different types.
- 7) For eg. linear, nonlinear, polynomial, radial basis funn & sigmoid.

- (Q.3)
- 1) **Deep Learning :** It is a method in AI that teaches computers to process data in a way that is inspired by the human brain.
 - 2) DL models can recognize complex patterns in pictures, text, sounds & other data to produce accurate insights & predictions.
 - 3) ML & DL are both types of AI. In short ML is AI that can automatically adapt with minimal human interference.
 - 4) DL is a subset of ML that uses artificial neural networks to mimic the learning process of the human brain.
 - 5) SegNet is a DL architecture applied to solve image segmentation problem.
 - 6) It consists of sequence of processing layers (encoders) followed by a corresponding set of decoders for a pixelwise classification.
 - 7) A CNN is a DL architecture designed for image analysis & recognition.
 - 8) It employs specialized layers to automatically learn features from images, capturing patterns of increasing complexity.

- Q. 4) 1) Feature scaling is the process of normalizing the range of features in a dataset.
- 2) Real-world datasets often contain features that are varying in degrees of magnitude, range & units.
- 3) Therefore, in order for ML models to interpret these features on the same scale, we need to perform feature scaling.
- 4) It preserves the relationship b/w the minimum & maximum values of each feature, which can be important for some algorithms.
- 5) It also improves the convergence & stability of some machine learning algo., particularly those that use gradient-based optimization.
- 6) By scaling features to a similar range, the alg. can take steps more uniformly across diff. dimensions, speeding up the learning process.
- 7) When using gradient descent-based optimization algo. Feature scaling can help speed up convergence & improve model performance.