Sakshi Dhananjay Shete
Roll no - 24061
Branch - CSE

51/100

Rajdhani
DATE / /

Class Assessment-3

Section A - Python Basics

Q.1 What is difference a list & tuple in python? provide an example of when you would use each.

→

List	Tuple
1) List is ~~mu~~mutable	1) Tuple is immutable
2) List iteration is slower & is time consuming	2) Tuple iteration is faster
3) List consumes ~~more~~ memory	3) Tuples consumes less memory
4) List operation are more errors prone	4) Tuples operation are safe
5) List provides many in-built methods	5) Tuples have less in-built methods
6) eg- list_num=[1,2,3,4] print(list_num)	6) eg- tup_num=(1,2,3,4) print(tup_num)

Q.2 Write a python function to calculate factorial of 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 Explain concept of list comprehension in python. provide a example how it can be used to create a list.

→ List comprehension

1) when you want to create a new list based on value of an existing list
2) E.g Based on list of fruits, you want a new list, containing only fruits with letter "a" in 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 (new list)
```

Q.4 Briefly explain purpose of following python libraries : Numpy, pandas & mataplotlib

→ a) i) Numpy :- Numpy can be used to perform a wide variety of mathematical operation a 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 purpose

2) Pandas is used for creating heterogeneous two-dimensional data objects

c i) Matplotlib - It is a comprehensive library for creating static, animated & interactive visualization in python
2) It provide an objects-oriented API for embedding plots into appln using general purpose GUI tool kits like tkinter

Section B - Machine Learning Basics

Q.1) Define supervise learning & unsupervised learning provide an example of each

→ A] supervised learning
1) In supervised learning, the machine is trained on a set of labeled data, which means that input data is paired with desired o/p.
2) The machine then learns to provide o/p for new input data.

B] Unsupervised learning
1) In this machine is trained on set of unlabeled data which means that input data is not paired with desired o/p
2) It used for tasks such as clustering dimensionality reduction & namely detection.

Rajdhani
DATE / /

Q.2 Explain bias-variance tradeoff in context of machine learning models How does it impact model performance?

→ 1) In machine learning as you try to maintain an component of error the other component tends to inc & vice versa

2) Finally righ 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 algorithm is unable to capture relevent rel^n bet^n factors & target output

4) A high bias model typically includes more assumption about target function or end results.

Q.3 Describe steps involved in machine learning pipeline.

→ 1) The steps in 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 well-organized format which could be suitable to apply for further steps.

3) Data ~~Ingestion~~ validation - In this step it required before training a new model. If focuses on

Rajdhani
DATE / /

Statistics of new data
4) Data pre-processing - The pre-processing step involves preparing the raw data & making it suitable for ML model. The process includes different sub-steps.
5) Model Training & Tuning - In this step the model is trained to take input & predict an o/p with highest possible accuracy
6) Model Analysis - After model training we need to determine the optimal set of parameters by using loss of accuracy metrics
7) Mode versioning - The model versioning steps keeps track of which model, set of hyper parameters & dataset have been selected as next version to be deployed
8) Model Deployment: After training & analyzing model, it's time to deploy the model An ML model can be deployed in 3 ways which are
- using model server
- In browser
- On edge device

4) A) Cross-validation
1) Cross-validation is technique for evaluating ML models by training several ML models on subsets of available input data & evaluating them on complementary subset of data
2) It is important in ML because it use to detect overfitting ie failing to generalize pattern

Rajdhani shelf
DATE / /

3) It used in ML to evaluate performance of 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 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 & then k-1 folds are used for training model.

Regression	Classification
1) In this problem statement the target variables are continuous	1) In this problem statement the target variables are discrete.
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 Algorithm
3) Input data are independent variables & continuous dependent variable	3) Input data are independent variable. & 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

6. Eg of regression alg. are linear regression, Polymial Regression, Ridge regression

5) Eg use cases are spam detection, image recognition

6) Eg of classification alg are Logistic regression decision trees, Random forest, SVM, K-NN etc.

Section C: Statistics & probability

1) Measures of central tendency help you find middle, or average of dataset.
2) The 3 most common measures of central tendency are mode, median & mean
3) Mode - The most frequent value
4) Median - The middle no in an ordered dataset
5) Mean - The sum of all values divided by total number of values
6) In addition to central tendency, the variability & distribution of your dataset is important to understand when performing descriptive statistics.

Q.3 1) The P value is defined as probability under assumption of no or no difference of obtaining a result equal to or more extream than what was actually observed.

2) The P stands for probability & measures how likely it is that any observed diff bet^n group is due to chance
3) In statistical hypothesis testing, P-value of can be defined as measure of probability that real-valued test statistic is at least as extreme as value actually obtained.
4) A P-value indicates probability of getting an effect no less than that actually observed in sample data
5) P-values are used in statistical hypothesis testing to determine whether to reject null hypothesis
6) The smaller P-value the stronger likelihood that you should reject null hypothesis

Q.4 1) Understanding the diff bet^n correlation & causation is essential in data science & statistics
2) Corelation refers to statistical rel^n ship bet^n two variables
3) While causation is rel^{n}ship bet^n cause & effect
4) While corelation 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 corelation bet^n ice cream sales & crime rates bcoz you notice that they both seem to rise & fall together

Rajdhani
DATE / /

Section D: Advanced Topics

Q.1) 1) Overfitting is an undesirable machine learning behaviour that occurs when the ML model gives accurate prediction for training data but not for new data.

2) When data scientist use ML models for making prediction, they first train model on known dataset.

3) Overfitting is modeling error which occurs when fun is too closely fit to a limited set of data pts.

4) In this case the ML model learns the detail & noise in training data such that it negatively affects performance of market on test data

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

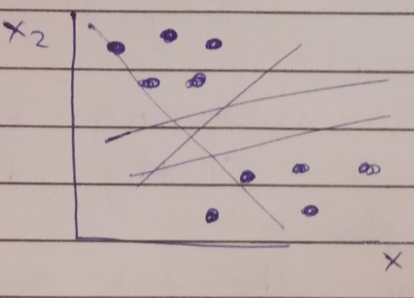
Q.2 1) SVM - SVM is ML algorithm that uses supervised learning models to solve complex classification, regression & outliers detection problem by performing optimal data transformation that determine boundaries betn data pts based on predefined classes labels or O/P

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

3) The main objective of svm algo is to find optimal hyperplane in N-dimensional

space that can separate data pts in diff classes in 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



Q.3 Deep learning

1) Deep learning, an other hand is subset of machine learning that uses neural network with multiple layers to analyze patterns & relationships in data.

2) DL models can recognize complex patterns in pictures, text, sounds & other data to produce accurate insights & prediction

3) ML & DL are both types of AI

In short ML is AI that can automatically adopt with minimal human interface.

4) DL is subset of ML that uses artificial neural networks to mimic learning process of human brain

5) SegNet is DL architecture applied to solve image segmentation problem

6) It consist of sequence of processing layers followed by corresponding set of decoders for a

Rajdhani
DATE / /

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

Rajdhani
DATE / /

improve model performance.