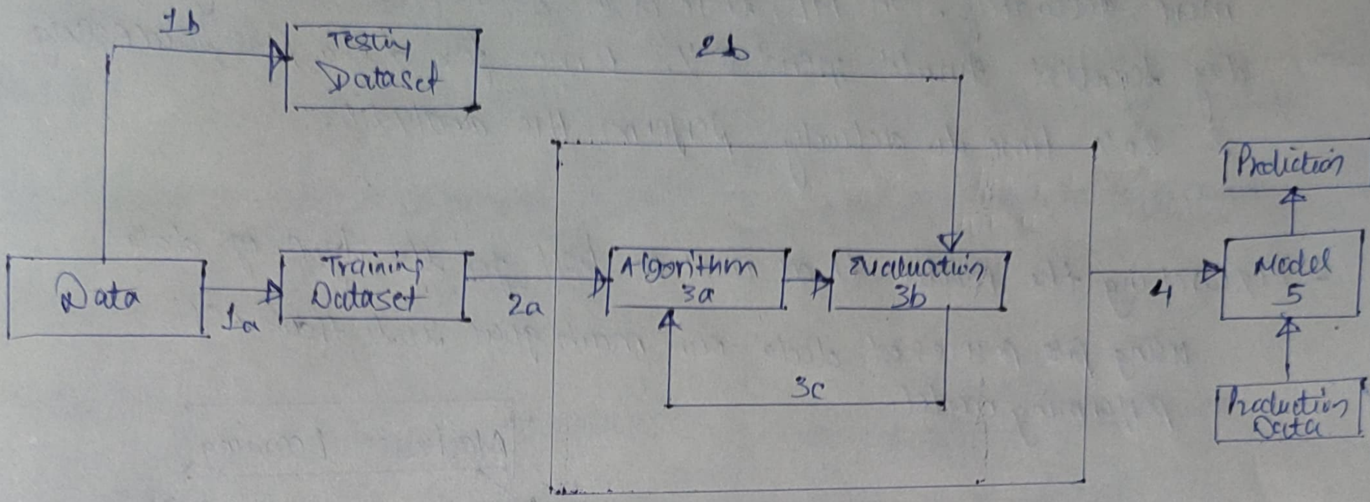
G PATEL
301029

ASSIGNMENT-01

Q-1 Describe ML process flow with appropriate diagram.



Fig(1):- overview of ML process flow.

ML workflow in 3 stages:-

1. Gathering data
2. Data Pre-processing
3. Researching the model that will be best for the type of data.
4. Training & testing the model.
5. Evaluation.

- Gathering of data :-
  - It depends on the type of project we desire to make, if we want to make an ML project that uses real-time data, then we can build an IOT system that using different sensor data. The data set can be collected from various sources such as file, database, sensor and many other.

P.T.O
Please turn over

CHIRAG PATEL ASSIGNMENT-01

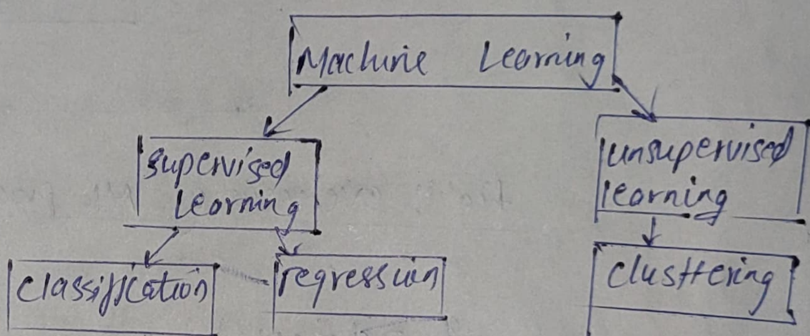
18630102g

- Data pre-processing :-

It helps in building machine learning models more accurately. In ML, there is an 80/20 rule. Every day scientist should spend 80% time for data pre processing & 20% time to actually perform the analysis.

- Researching the model that will be best for the type of data.

using pre-processed data our main goal is to train best performing model.



- Training & testing the model on data:-

we split the model in 3 section which are training data, validation data and testing data.

- Evaluation :- Its an integral part of model development process, helps to find the best model that represent our data & how well the chosen model will work in future.

Ques-02: Differtiate -

A F E S T D D D

Analysis method, Format, Ease to search, Storage, Tool, Database, Data model, Data nature

(a)	Structured Data	Unstructured Data
Formats	Several formats	A huge variety of formats
Data model	Pre-defined/not flexible	Not pre-defined / flexible
Storages	Data warehouses	Data lakes
Databases	SQL Relational Databases	NoSQL Non-relational databases
Ease of Search	Easy to search	Difficult to search
Data nature	Quantitive	Qualitative
Analysis method	• classification • Regression • Data clustering	• Data stacking • Data mining
Tools and technologies	• RDBMS • CRM • OLAP • OLTP	• NoSQL DBMS • AI-driven tools • Data storage architecture • Data visulization tools.
(b)	Online Machine learning	Offline Machine learning
Complexity	More complex	Less complex
Computational power	More Computational power is required	fewer Computational power is required

C C P A

Complexity, Computer pow, User input, Appln

P.T.O
Please turn over

CHIRAG PATEL
19630102 9

ASSIGNMENT-04

- Use in Production	Harder to implement & control because the production model changes in real-time according to its data feed.	Easier to implement because offline learning provides engineers with more time to perfect the model before deployment.
- Application	used where new data patterns are constantly required (eg. whether prediction tools.)	used where data application data patterns remain constant and don't have sudden concept drifts (eg. image classification)

Ques-03 Define bias-variance trade-off?

The goal of any supervised machine learning algorithm is to achieve low bias and low variance. In return the algorithm should achieve good prediction performance.

The parameterization of ML is often a battle to balance out bias & variance

example of configuring the bias-variance trade-off for specific algorithm.

- The K-nearest algorithm has low bias & high variance. but the trade-off can be changed by increasing the value of K which increases the no. of neighbours that contribute the prediction and in return. increases the bias of the model.

P.T.O
Please. turn over

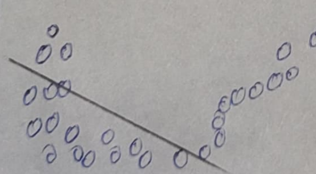
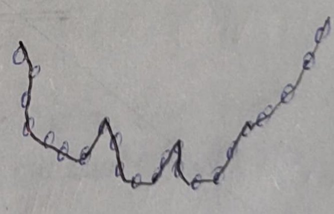
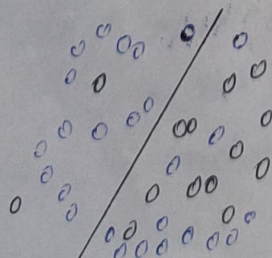
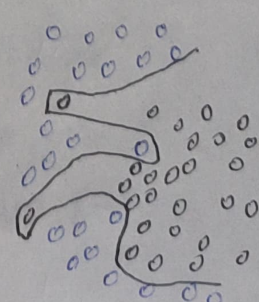
There is no escaping the relationship between bias and variance in ML.

$$\boxed{\text{Bias} \propto \frac{1}{\text{variance}}}$$

There is a trade-off at play between these two concerns and the algorithms you choose and the way you choose to configure them are finding different balances this trade-off for your problem.

Symptom Regresn^ clas^ Deeple Possible Remd.
S R C D/ P/

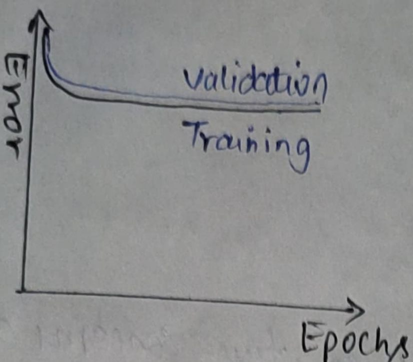
Ques-04 Differentiate:

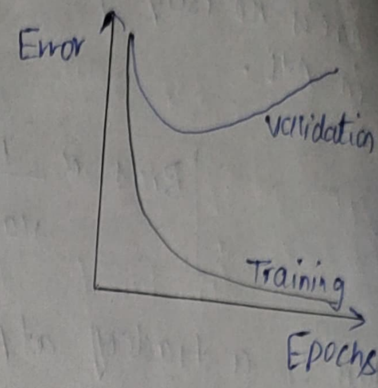
	Underfitting	Overfitting
Symptoms	• high training error • training error close to test error • high bias	• very low training error • training error much lower than test error • high variance
Regression illustration		
classification illustration		

P.T.O
Please turn over

196301029

Deep. Learning illustration





Possible remedies

- Complexity model
- Add more feature
- Train longer

- Perform regulonization
- Get more data

(10 marks) extra topics

→ cross validation

End

Machine learning

Cross-Validation :

- It's a technique ~~to~~ in which we train our model using the subset of the data-set and then evaluate using the complementary subset of data.

- There are three steps involved in Cross Validation

1. Reserve some portion of sample data-set.
2. Using the rest data-~~rest~~ set to train the model.
3. Test the model using the reserve portion of data set.

- Why we use cross validation
  - to test stability of model
  - we can't ~~find~~ fit our model on the training dataset.
  - we need a particular sample of dataset, which is not part of training dataset

Method used for Cross Validation

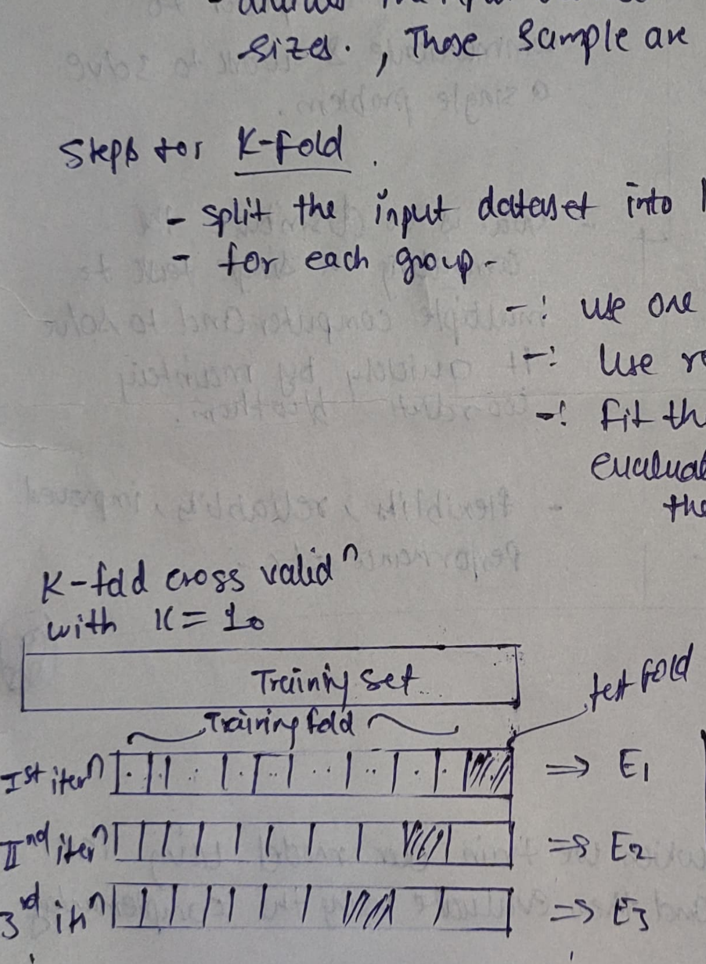
① K-fold cross-validation

- divides the input dataset in K-groups of sample of equal sizes., Those sample are called folds.

Steps for K-fold

- split the input dataset into K groups.
- for each group -
  - : use one group as reserve or test data set
  - : use remaining group as training dataset.
  - : fit the model on the training set & evaluate the performance of model using the test set.

K-fold cross valid^n with K = 10



$$E = \frac{1}{10}\sum_{i=1}^{10} E_i$$

② Stratified K-fold Cross validation → Best approach to deal w bias & variance

→ works on stratification concept → rearranging the data so that each fold or group is a good representative of complete dataset.

→ similar to K-fold with some little changes.