A. How would define Machine Learning [1M]

A Machine Learning is the field of study that gives Computers, the ability to learn without being explicitly programmed.

—o—

B. What kind of Problems Can be Solved using Machine Learning [2M]

A
1. Identifying Spam.
2. Making Product Recommendations.
3. Customer Segmentation.
4. Image & Video Recognition.
5. Fraudulent Transactions.
6. Demand Forecasting.
7. Virtual Personal Assistant.
8. Sentiment Analysis.
9. Customer Service Automation.

—o—

C. Describe elements of Machine Learning [7M]

A It is 'six' elements of the machine learning

they are :-

1. Data 2. Task 3. Model 4. Loss Function

5. Learning Algorithm 6. Evaluation

1. Data :-

↳ Data Simply means information.

↳ All types and formats of information.

↳ There is enormous amount of data produced every second and this data Can be used to answer so many questions.

↳ There is text data as well as audio-video data, there is Structured data as well as unstructured data.

↳ one important thing to remember is that it doesn't matter in which format you get the data, ~~at~~ the

↳ At the end all the data needs to be encoded as numbers before feeding it to the Computers.

2. Task :-

Now, you have data.

↳ What task do I need to achieve?

↳ What task Can I achieve using the data that I have?

↳ Generally, Machine Learning tasks are divided into two parts that are 'Supervised Learning' & Unsupervised Learning.

↳ Supervised Learning is when you have Some input data as well as some output data corresponding to the input data.

↳ Then you can use Machine Learning to get the relationship between the input & output data and then use that relationship to predict output for the new input.

↳ Unsupervised Learning is when you have only input data.

↳ Supervised Learning is further divided into two sub sections as 'Regression' & 'Classification'

↳ Where Regression means you want to predict some continous value and Classification means you want a ~~distance~~ discrete answer such as yes or no.

↳ unsupervised learning is often used for Generation and clustering.

3. <u>Model</u> :-

↳ A model is nothing but a mathematical function which define relationship between i/p data & o/p data.

↳ A model can be a simple linear function such as "$y = mx + c$"

↳ There are many models available that are created over the years and each one has it's own pros and cons and each one is suitable for a unique task.

↳ For a human, it is almost impossible to come up.

with a function that defines relationship between the i/p & o/p, given that the input data has over thousand different features.

4. Loss Function:-

↳ Although, it's not possible to come up with a model just by looking at the given data, but let's assume

↳ we have come up with a model that you think correctly defines the relationship between input & output

↳ Three different values predicted by the models that are $f_1(x)$, $f_2(x)$ & $f_3(x)$ respectively, the true value 'y'

↳ We compute the difference between the true value and predicted value for all the data using a loss function.

↳ We get to know that the best model for our data is the one with the minimum value of loss function

5. Learning Algorithm:-

↳ We have a model with three parameters a, b & c and we need values of a, b and c at which the loss function's value is minimum.

↳ There are many optimization strategies available that we can use to find the best values of parameters.

Learning algorithm is one task that Computer does, other than that from data Collection to deciding task.

6. Evaluation:-

↳ let's say using the above five elements you've Created and ML model to detect the animal in the image.

↳ How do you know that your model of detecting right you have to test your model.

↳ That's what evaluation is, Testing your model by feeding it Some test data and then checking if it predicts the Correct output in all the Cases?

Note:- you evaluate your model using test data and not the training data.

— o —

2a) Differentiate Regression from classification.

Regression Algorithm	Classification Algorithm
(i) In Regression, the output variable must be of continuous nature or real value.	(i) In classification, the output variable must be a discrete value.
(ii) The task of the regression algorithm is to map the input value (x) with the continuous output variable (y).	(ii) The task of the classification algorithm is to map the input value (x) with the discreate output variable (y).
(iii) Regression Algorithms are used with continuous data.	(iii) Classification Algorithms are used with discreate data.
(iv) In Regression, we try to find the best fit line, which can predict the output more accurately.	(iv) In classification, we try to find the decision boundary, which can divide the data set into different classes.
(v) Regression algorithms can be used to solve the regression problems such as Weather predication, House price prediction, etc..	(v) Classification Algorithms can be used to solve classification problems such as identification of spam emails, Speech Recognition, identification of cancer cells, etc..
(vi) The Regression Algorithm can be further divided into Linear and Non-linear Regression.	(vi) The classification algorithms can be divided into Binary classifier and multi-class Classifier.
(vii) Types of Regression Algorithms are simple Linear Regression, multi-linear regression, polynomial Regression, support vector Regression, Decision Tree Regression, etc..	(vii) Types of classification Algorithms aree, Logistic Regression, Kernel SVM, K-Nearest Neighbours, Naïve Bayes, support vector Machines, Decision Tree classifications, etc..

2b) Differentiate an online learning system from Batch learning system?

Batch learning system	online learning system
(i) An Batch learning system handles large amounts of data which processed on a routine schedule.	(i) An online learning system handles transactions in real time and provides the output instantly.
(ii) processing occurs when the after the economic event occurs and recorded.	(ii) when the economic event takes place then the processing occurs.
(iii) In batch learning system fewer programming, hardware and training resources are required.	(iii) In online learning system more number of dedicated hardware resources, processing elements are required.
(iv) To avoid operational delays certain records are processed after the event.	(iv) Immediately all the records pertaining to event are processed.
(v) In this system input data is prepared before the execution.	(v) In this system data is prepared at time of execution as needed.
(vi) In this system the processing sequence is predictable.	(vi) In this system the processing sequence is unpredictable.
(vii) In this the programs and files cannot be shared.	(vii) In this the program and files can be shared.
(viii) In batch learning system, system recovery and restart is easy.	(viii) In online learning system, recovery and restart requires additional process.
(ix) This system uses tape storage	(ix) This system uses disk storage.
(x) Ex: Inventory query, website shopping transaction, e-Banking account withdrawal etc..	(x) Ex: month end tax calculation, data transformation, data analysis, data transformation etc..

2c) What is out-of-core learning?

Ans out-of-core (or "external memory") learning is a technique used to learn from data that cannot fit in a computer's main memory (RAM), but can easily fit into some data storage, such as local hard disk or web repository.

There are three ways to perform it in three steps:

a) streaming data

b) Extracting features

c) Training Model

3)a) What are the main Challenges in Machine Learning?

Main Challenges of machine learning

In Short, since your main task is to select a learning algorithm and train it on some data, the two things that can go wrong are "bad data and bad algorithm".

Examples of bad data:

1. Insufficient Quantity of Training data
2. Poor - Quality Data
3. Irrelevant Features
4. Nonrepresentative Training Data

Examples of bad algorithm:

5. Overfitting the Training Data
6. Underfitting the Training Data
7. Stepping Back

1. Insufficient Quantity of Training Data:

→ Data plays an important role in the machine learning languages.

→ Machine learning takes a lot of data for most machine learning algorithms to work properly.

→ Even for very simple problems you typically need thousands of examples, and for complex problems such as image or speech recognition you may need millions of examples.

→ Less amount of training data will produce inaccurate or biased predictions.

2. Poor- Quality Data :

→ Noisy data, incomplete data, inaccurate data and unclean data lead to accuracy in classification and low-quality results.

→ We don't want our algorithm to make inaccurate or faulty predictions

→ Hence the quality of data is essential to enhance the output.

→ Data quality can also be considered as a major problem while processing the machine learning algorithms.

→ Therefore, we need to ensure, that data preprocessing should be done in perfection.

3. Irrelevant Features :

→ Our model will only be capable of learning, if the training data contains enough relevant features and not too many irrelevant.

→ Feature Engineering is one of the key step to a successful machine learning model.

→ Feature Engineering involves the following steps :

a) Feature Selection : Selecting most useful features to train.

b) Feature Extraction : Combining existing features to produce a more useful one.

c) Creating new features by gathering new data.

4. Nonrepresentative Training Data :

→ In order to generalize well, it is crucial that your training data be representative of new cases you want to generalize to. This is true wether you use. instance-based learning or model-based learning.

→ If the sample is too small you will have sampling noise, but even very large samples can be nonrepresentative if the sample method is flawed. This is called sampling bias.

5. Overfitting of Training Data :

→ Overfitting occurs when the machine learning model trained with a massive amount of data that negatively affect its performance.
Eg. Trying to fit in oversized jeans.

→ This is one of most common issue faced by the machine learning professionals and data scientists.

→ The algorithm is trained with noisy and biased data which will affect its overall performance.

→ To overcome this issue :

a. Reduce the noise

b. Increase training data in a data set.

c. Reduce model complexity by selecting fewer parameters.

d. Reduce the no. of attributes in training data.

6. Underfitting the Training Data :

→ Underfitting occurs when the data is unable to establish an accurate relationship between input and output Variables.

Eg: Trying to fit in undersized jeans.

→ It generally happens when we have limited data in the dataset and we try to build the model.

→ To overcome the issue :

a. Maximize the training data
b. Increase model complexity
c. Remove noise from data.
d. Add more feature to the data
e. Reduce Regular Parameters.

7. Stepping Back :

→ In machine learning, there are so many concepts however, we went through these concepts that may feel like little lost, so let's step back.

3) b) If your model performs great on training data but generalizes poorly to new instances, what is happening? Can you name three possible Solutions?

If a model performs well on the training data but poorly on new instances, it's likely that the model is overfitting to the training data. Overfitting occurs when a model becomes too complex and fits the noise in the data rather than the underlying pattern.

Here are three possible Solutions to address overfitting:

1. Regularization : Regularization techniques can help prevent overfitting by adding a penalty to the loss function of the model for large parameter values. This can be achieved by adding a L1 or L2 penalty to the cost function, which reduces the weight of certain features or limits the magnitude of the weights in the model.

2. Increasing Data : Increasing the size of the training dataset can help reduce overfitting. This is because the more data the model is exposed to, the better it can capture the underlying pattern and generalize to new instances. This can be achieved by collecting more data, data augmentation or Synthetic data generation.

3. Early Stopping : Early stopping is a technique that monitors the validation error during training and stops training when the validation error starts to increase. This prevents the model from overfitting to the training data and can improve its ability to generalize to new instances.

Cross-validation :- It is simplest form in a round validation where ~~leave~~ leave one sample as in-time validation and rest for training the model. But for keeping lower variance a higher fold cross validation is Preferred.

Pruning :- It is used extensively while building CART Models. It simply removes the nodes which add little Predictive power ~~of~~ for the Problem.

4. (a). What are the purposes of Training set, ~~A~~ validation set and Test set in the context of machine learning?

The original data is split into subsets like training set, validation set and test set.

Training set :- The training set is used to fit or train the model. These data points are used to learn the parameters of the model.

(1). This is the biggest of all sets in terms of size. The training set includes the features and well as labels in the case of supervised learning.

(2). In the case of unsupervised learning, it can simply be the feature sets.

(3). These labels are used in the training phase to get the training accuracy score.

(4). The training set is usually taken as 70% of the original dataset but can be changed per the use case or available data.

Validation Set :-

(1). The validation set is used to provide an unbiased evaluation of the model fit during hyperparameter tuning of the model.

(2). It is the set of examples that are used to change learning process parameters.

(3). Optimal values of hyper parameters are tested against the model trained using the training set.

(4). In machine learning we generally need to test multiple models with different hyper parameters and check which model gives the best result. This process is carried out with the help of validation set.

Testing Set :-

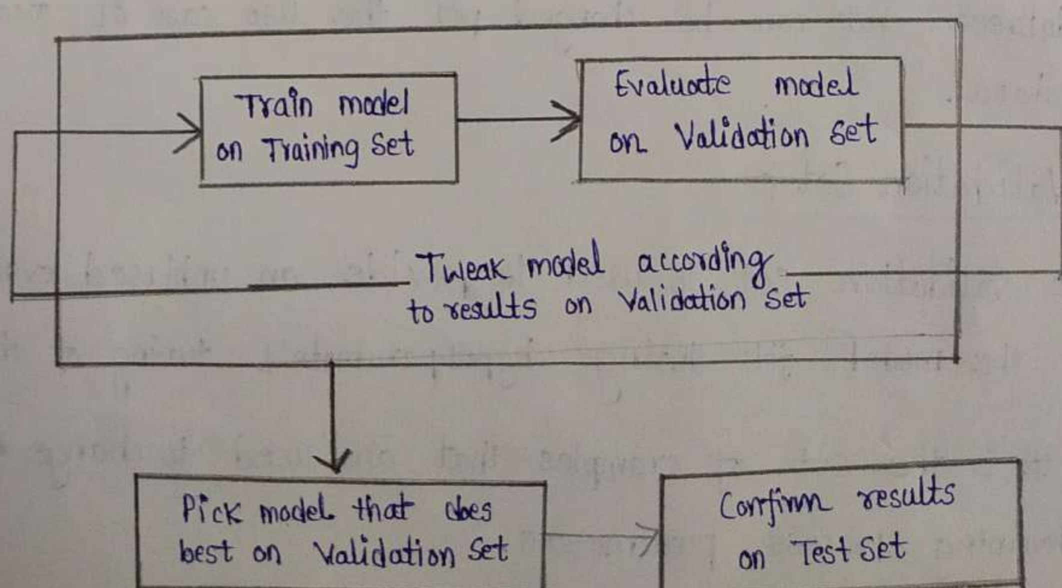
(1). Once we have the model trained with the training set and the hyperparameter tuned using the validation set, we need to test whether the model can generalize well on unseen data.

(2). To accomplish this, a test set is used. Here we can check and compare the training and test accuracies.

(3). To ensure that the model is not overfitting or underfitting, test accuracies are highly useful.

(4). If there is a large difference in train and test accuracies, overfitting might have occured.

Training data | validation | test



4 (b). Write about stochastic Gradient Descent Optimization method along with its merits and de-merits.

(1). Stochastic gradient descent (SGD) uses only a single example (a batch size of 1) per iteration.

(2). Given enough iterations, SGD works but is very noisy. The term "stochastic" indicates that the one example comprising each batch is choosen at random.

(3). SGD is a variant of the gradient descent algorithm used for optimizing machine learning models. In this variant, only one random training example is used to calculate the gradient and update the parameters at each iteration.

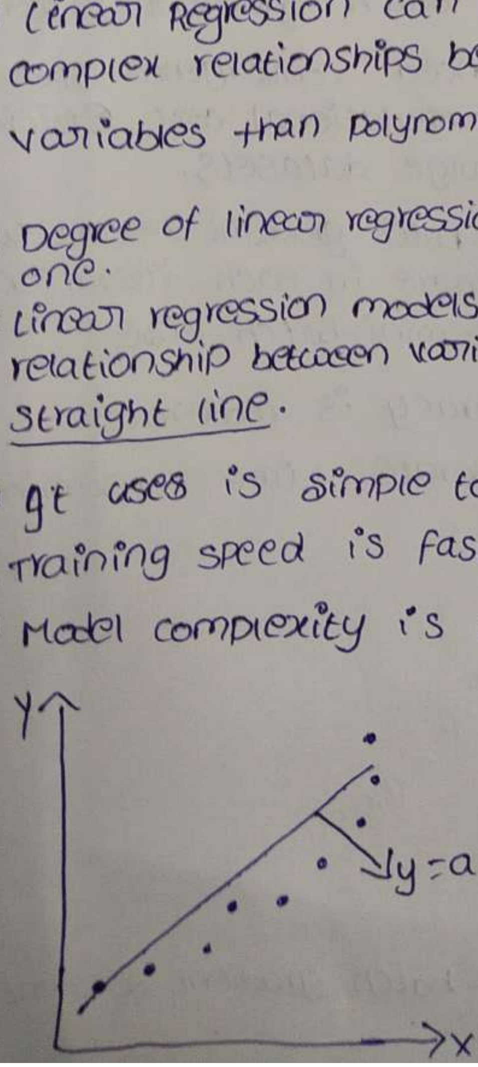
Merits and de-merits of SGD :-

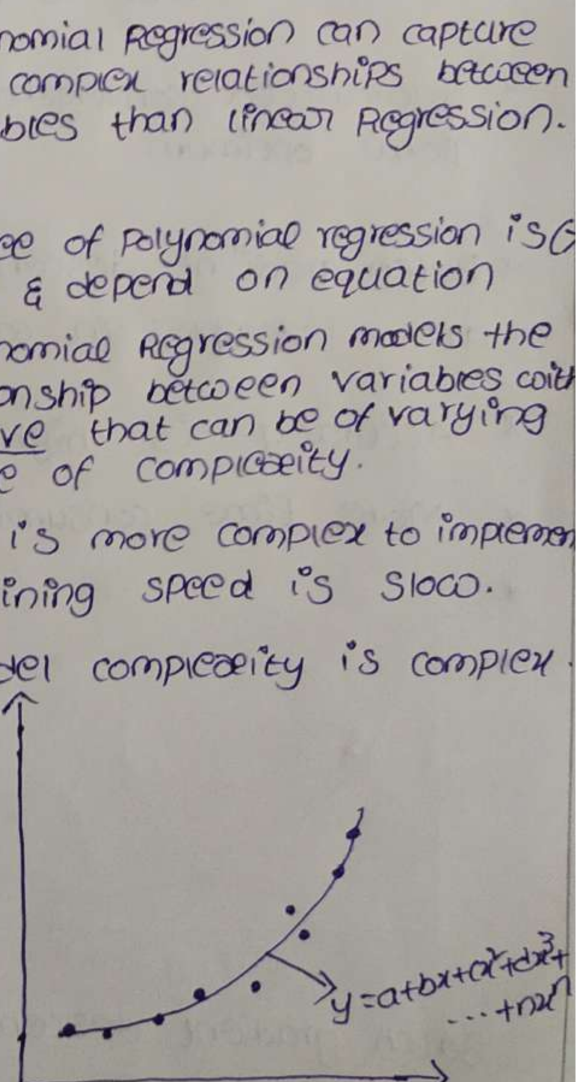
Merits	De-merits
1. Speed :- SGD is faster than other variants of gradient descent such as Batch gradient descent and mini-Batch gradient descent.	(1). Noisy updates :- The updates in SGD are noisy and have a high variance, which can optimization less stable.
2. Memory Efficiency :- It is memory efficient and can handle large datasets that cannot fit into memory.	(2). Slow Convergence :- SGD may require more iterations to converge to the minimum since it updates parameter for each training set one at a time.

3. Avoidance of local minima :- Due to noisy updates in SGD, it has the ability to escape from local minima and converge to a global minimum.	Less Accurate :- Due to noisy updates, SGD may not converge to the exact global minimum and can result in a suboptimal solution.

56. Compare Polynomial Regression and Linear Regression?

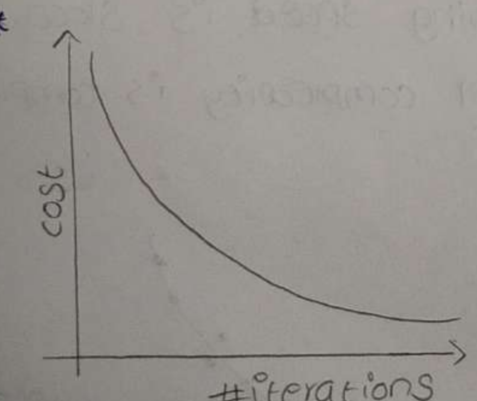
Linear Regression	Polynomial Regression
* Linear Regression model shows linear relationship between dependent(y) and one or more independent(y) variables	* Polynomial Regression model shows the relationship between dependent(y) and independent variable (x) as n^{th} degree polynomial.
* The eqn of the form of $y = a + bx$ where y is dependent variable, x is independent variable, a is intercept, b is slope of the line	* The eqn of the form of $y = a + bx + cx^2 + dx^3 + \dots + nx^n$, y is dependent variable, x is independent variable, n is degree of polynomial.
* Linear Regression assumes a linear relationship between variables.	* Polynomial Regression assumes a non-linear relationship between variables
* Linear Regression can capture less complex relationships between variables than polynomial Regression.	* Polynomial Regression can capture more complex relationships between variables than linear Regression.
* Degree of linear regression is one.	* Degree of Polynomial regression is high & depend on equation
* Linear regression models the relationship between variables with straight line.	* Polynomial Regression models the relationship between variables with a curve that can be of varying degree of complexity.
* It uses is simple to implement.	* It is more complex to implement
* Training speed is fast	* Training speed is slow.
* Model complexity is simple.	* Model complexity is complex.



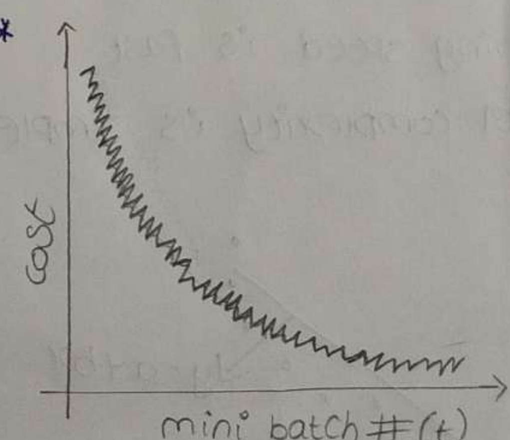


5a) Compare Batch gradient descent optimization algorithm and Mini-Batch gradient descent optimization algorithm?

Batch Gradient Descent	Mini-Batch Gradient Descent
* The entire Dataset samples in each gradient step	* The subset of the dataset samples in each gradient step.
* It requires 1 update ~~for~~ per epoch	* It requires $\frac{N}{\text{size of mini-Batch}}$ updates per epoch.
* Entire dataset for updation	* Subset of data for updation.
* cost function reduces smoothly	* Smoother cost function.
* computation cost is very high	* computation cost is lesser than Batch Gradient Descent.
* Guaranteed convergence to global optimum	* Balanced convergence speed and computational cost, efficient for large datasets.
* Batch gradient descent use all m examples in each iteration	* Mini-batch gradient descent use b example in each iteration b=mini-batch size
* Accuracy is high	* Accuracy is moderate
* More time consuming	* Moderate time consuming.
*	*



Batch gradient descent



* Mini-batch gradient descent.

6. Write about commonly used Regression loss Functions

The loss function measures how well a given machine learning model fits the specific data set. The average error of n-samples in the whole data is called Cost Function and the error for individual data points is the loss Function.

The various types of regression loss functions to estimate the performance of the machine which uses different types of algorithms are as follows.

1. Mean Absolute Error
2. Mean Squared Error
3. Root Mean Squared Error
4. R-Squared
5. Adjusted R-Squared
6. Mean Bias Error
7. Mean Percentage Error
8. Mean Absolute Percentage Error.

1. Mean Absolute Error / L_1 loss :-

Mean Absolute Error takes the average sum of the absolute differences b/w the actual and the predicted values. For a data point x_i and its predicted values y_i, n being the total number of data points in the dataset, the MAE is defined as

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

2. Mean Squared Error / L_2 loss :-

Mean Squared Error is the average of the squared difference b/w the actual and the predicted values. For a data points y_i and its predicted values $\hat{y}_i$, where n is the total no. of data points in the dataset, the MSE is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

3. Root Mean Squared Error :-

This is calculated by applying square root function on Mean Squared Error.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

4. Adjusted R-Square :-

It represents the proportion of variation in your data, explained by the model. The r-squared value ranges from 0 to 1, where 0 means no relationship and 1 means 100% related. This corresponds to the overall quality of the model. The higher the adjusted R^2, the better the model.

5. Mean Bias Error :-

Mean Bias Error is the Exact difference b/w the predicted value and the actual value without any math function like absolute or square root applidd to it. "The major limitation that MBE has is positive and negative Errors have a chance of cancelling out". That is why this is rarely used and less popular function.

$$MBE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}$$

7. Mean Percentage Error :-

This is calculated as follows.

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{y_i}$$

8. Mean Absolute Percentage Error :-

This is calculated as follows

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|(y_i - \hat{y}_i)|}{y_i}$$