

# ML NOTES

-BY CLOUDYML

Q 1] What are the different types of machine learning?

→ Supervised Learning

A model makes predictions or decision based on labeled data. Labeled data refers to set of data that given tags or label.

Unsupervised Learning

We don't have labeled data. A model can identify patterns, anomalies and relationship in the input data.

Reinforcement Learning

The model can learn based on the rewards it received for its previous actions.



Akash Raj  
(Data Scientist)

in 34K+  
31K+  
40K+



Q.2]

what is overfitting, and how can you avoid it?

→ It occurs when a model learns the training set, taking random mistakes in the testing data as concepts.

When a model is given the training data, it shows 100% accuracy - technically a slight loss. When we use the test data there may be error and low efficiency.

There are multiple ways of avoiding overfitting.

① Regularization : It involves a cost term for the features involved with objective functions.

② Making a Simple model. With lesser variables and parameters, the variance can be reduced.

③ Cross-validation methods like K-folds can also be used.



Akash Raj  
(Data Scientist)



34K+  
31K+  
40K+



Q.3

Explain the confusion matrix with respect to machine learning algorithms.

- It is a special table that is used to measure the performance of an algorithm.
- The confusion matrix has two parameters.
  - Actual Values
  - Predicted Values

Consider a confusion matrix [Binary Classification]

Actual Values

		Yes	No
Actual	Yes	12	3
	No	1	9

for actual values,

$$\text{Total Yes} = 12 + 1 = 13$$

$$\text{Total No} = 3 + 9 = 12$$

for predicted values,

$$\text{Total Yes} : 12 + 3 = 15$$

$$\text{Total No} : 1 + 9 = 10$$

$$\rightarrow \text{Total observations} = 12 + 3 + 1 + 9$$

$$= 25 //$$



Akash Raj  
(Data Scientist)



Q.4

what are the application of supervised machine learning of modern Business?

→ Email Spam Detection.

We train the model using historical data that consist of email categorized as spam or not spam.

→ Healthcare Diagnosis.

By providing images regarding a disease, a model can be trained to detect if a person is suffering from the disease or not.

→ Sentiment Analysis.

This refer to the process of using algorithm to documents and determine whether they are positive, neutral or negative.

→ Fraud Detection.

By training the model to identify suspicious patterns, we can detect instances of possible frauds.



Akash Raj  
(Data Scientist)

in 34K+  
31K+  
40K+



Q.5 How do you handle missing or corrupted data in a dataset?

→ One of the easiest way to handle missing or corrupted is to drop those rows or columns or replace them entirely with some other value.

There are two useful methods in Pandas:

- isnull() and dropna() will help to find the columns/rows with missing data and drop them.
- fillna() will replace the wrong values with a placeholder value.

Q.6 How can you choose a classifier based on a Training set Data size?

→ When the training set is small, a model that has high bias and low variance seems to work better because they are less to overfit.

for example, Ahve Bayes works best when the training set is large, models with low bias and high variance tend to perform better as work complex relationships.



Akash Raj  
(Data Scientist)



in 34K+  
31K+  
40K+



Q. 9

what are the three stages of Building a model in machine learning?

→ The Three Stages of building a ml model:

- Model Building

Choose a suitable algorithm for the model and train it according to the requirement.

- model testing.

Check the accuracy of model through the test data.

- Applying the model

make the required changes after testing and use the final model for real time projects.

Here, it's important to remember that once in a while, the model needs to be checked to make sure it's working correctly. It should be modified to make sure that is up-to-date.



Akash Raj  
(Data Scientist)



in 34K+

31K+

40K+



Q.8

What are unsupervised machine learning techniques.

→ There are two techniques used in unsupervised learning: Clustering and Association.

### Clustering

Clustering problems involve data to be divided into subsets. These subsets, also called clusters, contain data that are similar to each other. Different clusters reveal different details about objects unlike classification or regression.

### Association

In an association problem, we identify patterns of associations between different variables or items.

For example, an e-commerce website can suggest other items for you to buy based on the prior purchase that you have made, spending habits, items in your wishlist, others' purchase history and so on.



Akash Raj  
(Data Scientist)



in 34K+  
31K+  
40K+



**IF YOU WANNA MAKE YOUR CAREER IN**

# **DATA SCIENCE & ANALYTICS**

**Then Check Our Mega Combo Course**



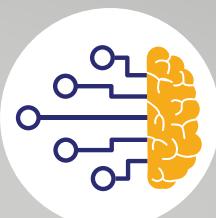
**Python**



**SQL**



**Statistics**



**Machine  
Learning**



**Deep  
Learning**



**DL Projects**



**AWS  
Sagemaker**



**Tableau**



**Excel**



**PowerBi**

**Click Here To Visit  
Our Website Now**

