

Intro To ML - AMMI

RAHEEM, OLAMILEKAN WALIYU

February 2021

1 What is the main key difference between supervised and unsupervised machine learning?

Supervised Learning is an algorithm where we specify input and output during learning process. Supervised learning is mostly Regression and Classification while Unsupervised learning has inputs of untagged data; finds some correlation and pattern in it. Unsupervised learning is mostly Clusters and Association.

2 What is Linear Regression?

Linear Regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. Linear Regression is usually used for predictive analysis.

3 List all assumptions for data to be met before starting with linear regression?

Before adopting linear regression, the data must be observed to have the features below:

- No Multi-collinearity: There is little or no multi-collinearity in the data. Multi-collinearity happens when the independent variables are highly correlated with each other.
- Independence: Observations are independent of each other.
- No auto-correlation: single column data values are not related to each other.
- Linear Relationship: The relationship between the independent and dependent variables should be linear.

- Multivariate Normal: All the variables should be multivariate normal. For all the variables to be multivariate normal each variable separately has to be univariate normal means a bell shaped curve. And any subset of variables should also be multivariate normal.

4 Differentiate between Regression and Classification?

The most significant difference between regression and classification is that while regression helps predict a continuous quantity, classification predicts discrete class labels.

5 Explain the difference between Normalization and Standardization

Normalization simply shrinks the range of the data such that the range is fixed between 0 and 1 (or -1 to 1 if there are negative values). Standardization means putting all the feature data into one standard mean and standard deviation. It is the process of rescaling the features so that they'll have the properties of a Gaussian distribution.

6 Why is logistic regression a type of classification technique and not a regression? Name the function it is derived from?

Logistic Regression is derived from cost function, it is a classification technique because it generates discrete value using regression as a classifier.

7 Explain the phrase “Curse of Dimensionality”

Curse of Dimensionality describes the phenomenon where the feature space becomes increasingly sparse for an increasing number of dimensions of a fixed-size training dataset.

“As the number of features or dimensions grows, the amount of data we need to generalize accurately grows exponentially.” Charles Isbell, Professor and Senior Associate Dean, School of Interactive Computing, Georgia Tech

8 What's a Fourier transform?

Fourier Transform takes a time-based pattern, measures every possible cycle, and returns the overall "cycle recipe" (the amplitude, offset, rotation speed for every cycle that was found). Fourier Transform transforms data from the time (or spatial) domain into the frequency domain

9 What is a confusion matrix and why do you need it?

The confusion matrix visualizes the accuracy of a classifier by comparing the actual and predicted classes. The binary confusion matrix is composed of squares:

- True Positive (TP): Predicted values correctly predicted as actual positive.
- False Positive (FP): Predicted values incorrectly predicted an actual positive. i.e., Negative values predicted as positive.
- False Negative (FN): Positive values predicted as negative
- True Negative (TN): Predicted values correctly predicted as an actual negative

It is needed because the terminologies can be confusing and needed to compute accuracy.

10 What is the Principle Component Analysis?

Principal Component Analysis is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. The objective is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

11 What is the difference between regularization and normalisation?

Normalization adjusts the data while regularization adjusts the prediction function.

12 List the most popular distribution curves along with scenarios where you will use them in an algorithm.

- Normal Distribution: is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean for instance, the weight of newborn babies.
- Binomial Distribution: used in a possibility of only two options.
- Exponential Distribution: is memory-less (has no history or previous occurrence) and concerned about when an event would occur
- Bernoulli Distribution: would be used in a classification prediction.
- Uniform Distribution: used when there is a constant probability.
- Poisson Distribution: when an event has occurred in the past and wants to predict again. It requires a history of previous occurrence

13 Can you mention some advantages and disadvantages of decision trees?

Advantages:

- Less data cleaning is required
- Easy to prepare
- Easy to read and interpret

Disadvantages:

- Less effective in predicting continuous variable.
- Very unstable; a slight change in data may influence the structure

14 What is the exploding gradient problem while using back propagation technique?

Exploding gradients are a problem where large error gradients accumulate and result in very large updates to neural network model weights during training. This has the effect of the model being unstable and unable to learn from the training data.

15 What is the difference between stochastic gradient descent (SGD) and gradient descent (GD)?

Stochastic Gradient Descent uses a random instance of training data at each step and then computes the gradient making it much faster as there is much fewer data to manipulate at a single time, unlike Batch Gradient Descent which involves calculations over the full training set at each step as a result, it is slower especially on very large training data.

16 When does regularization come into play in Machine Learning?

Regularization is a technique used for tuning the function by adding an additional penalty term in the error function. The additional term controls the excessively fluctuating function such that the coefficients don't take extreme values. This technique of keeping a check or reducing the value of error coefficients are called shrinkage methods or weight decay in case of neural networks. It is used to avoid overfitting and underfitting.

17 What is overfitting?

A model is overfitted when it is so specific to the original data that trying to apply it to data collected in the future would result in problematic or erroneous outcomes and therefore less-than-optimal decisions. Overfitting happens when a machine learning model has become too attuned to the data on which it was trained and therefore loses its applicability to any other dataset.

18 Explain One-hot encoding and Label Encoding. How do they affect the dimensionality of the given dataset?

One-hot encoding is representing the categorical variables as binary vectors while Label Encoding is changing of form of the labels into the numeric form to change it into a form that can be read by the machine. Machine learning algorithms can thereafter determine in a correct way as to how these labels must be managed. It is a crucial pre-processing measure during the integrated dataset in supervised learning. Using one-hot encoding increases the dimensionality of the data set while Label encoding does not affect the dimensionality of the data set.

19 There are many machine learning algorithms till now. If given a data set, how can one determine which algorithm to be used for that?

Choice of machine learning algorithm depends majorly on the type of data given. Futhermore, We can use exploratory data analysis to understand the purpose of using the dataset to come up with the best fit algorithm.

20 How do you select important variables while working on a data set?

There are a few method of selecting important variables in a dataset:

- Top features can be chosen based on information gained from data set.
- Lasso Regression
- Forward, Backward, and Stepwise selection
- Random Forest

21 What is the difference between deep learning and machine learning?

Deep learning is a subfield of machine learning and both fall under the broad category of Artificial Intelligence. Machine learning uses algorithms to parse data, learn from that data, and make informed decisions based on what it has learned. Deep learning structures algorithms in layers to create an "artificial neural network" that can learn and make intelligent decisions on its own.

22 What are the different types of Learning/ Training models in ML?

The types of learning models in ML are:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

23 Explain the terms Artificial Intelligence (AI), Machine Learning (ML and Deep Learning?

Artificial Intelligence studies ways to build intelligent programs and machines that can creatively solve problems, which has always been considered a human prerogative.

Machine learning is a subset of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. In ML, there are different algorithms that help to solve problems.

Deep Learning is a subset of machine learning, which uses the neural networks to analyze different factors with a structure that is similar to the human neural system.

24 How are covariance and correlation different from one another?

Covariance signifies the direction of the linear relationship between the two variables while Correlation analysis is a method of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables. Covariance is when two variables vary with each other, whereas Correlation is when the change in one variable results in the change in another variable.

25 State the differences between causality and correlation?

Causality means that two events appear at the same time or one after the other. These two variables not only appear together, the existence of one causes the other to manifest.

Correlation refers to the degree of association between two random variables. Correlation between two data sets is the amount to which they resemble one another.