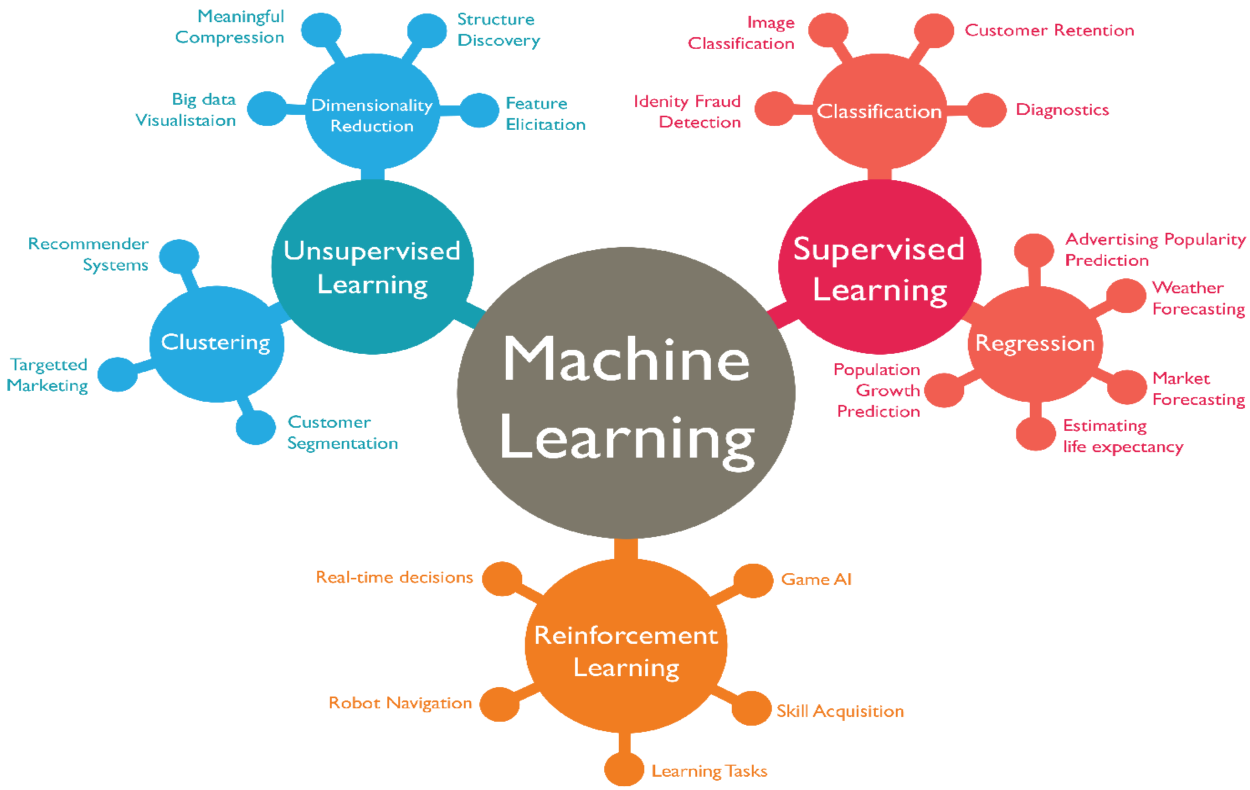
Interview Questions

Q1. What is Machine Learning?

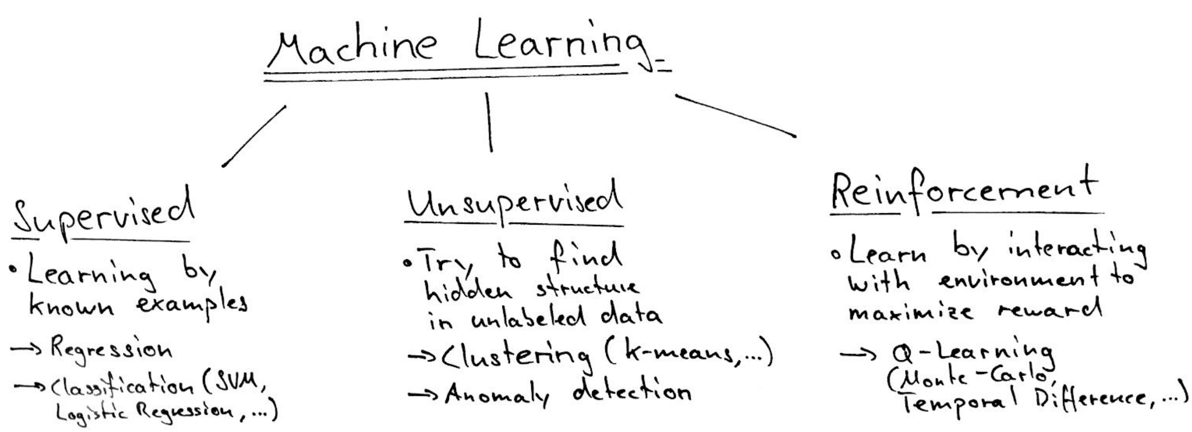
Instead of engineers "teaching" or programming computers to have what they need to carry out tasks, that perhaps computers could teach themselves – learn something without being explicitly programmed to do so. ML is a form of AI where based on more data, and they can change actions and response, which will make more efficient, adaptable and scalable. e.g., navigation apps and recommendation engines.
Classified into:-
1. Supervised
2. Unsupervised
3. Reinforcement learning

Machine-learning algorithms use statistics to find patterns in massive* amounts of data. And data, here, encompasses a lot of things—numbers, words, images, clicks, what have you. If it can be digitally stored, it can be fed into a machine-learning algorithm.

Machine learning is the process that powers many of the services we use today—recommendation systems like those on Netflix, YouTube, and Spotify; search engines like Google and Baidu; social-media feeds like Facebook and Twitter; voice assistants like Siri and Alexa. The list goes on.



Q2. What is the difference between Supervised learning, Unsupervised learning and Reinforcement learning?



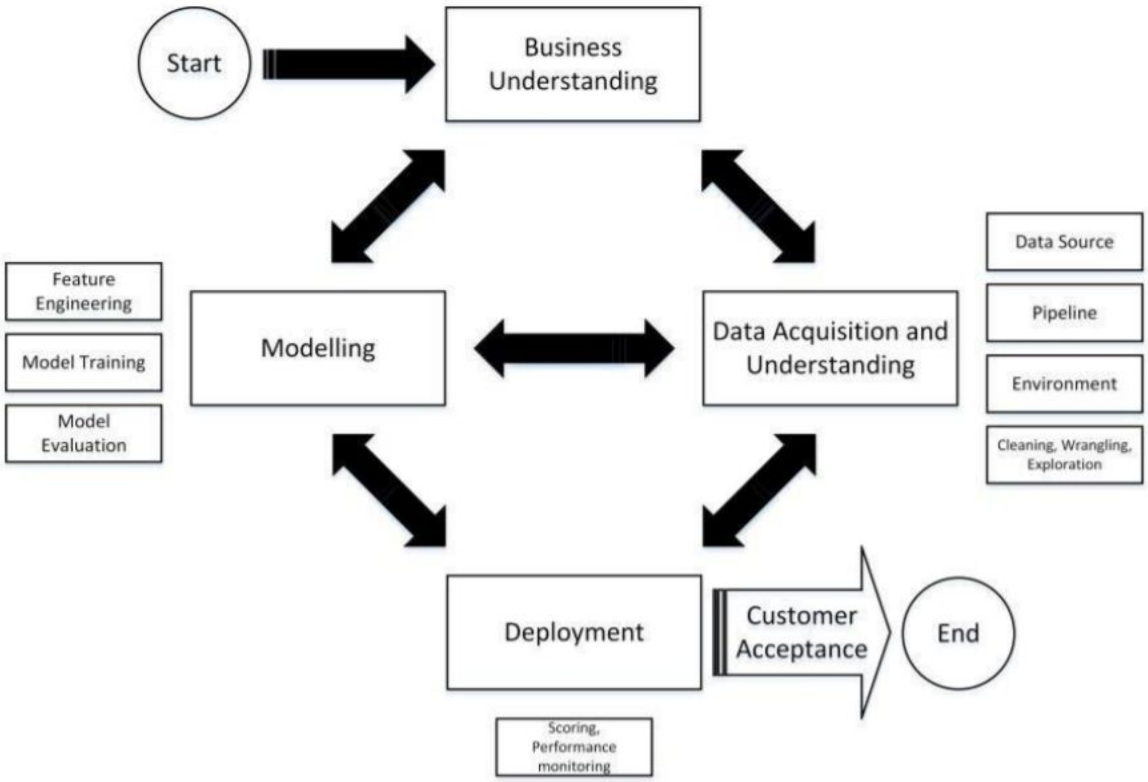
Supervised learning
In a supervised learning model, the algorithm learns on a labeled dataset, to generate reasonable predictions for the response to new data. (Forecasting outcome of new data)
• Regression

• Classification

Unsupervised learning
An unsupervised model, in contrast, provides unlabelled data that the algorithm tries to make sense of by extracting features, co-occurrence and underlying patterns on its own. We use unsupervised learning for
• Clustering
• Anomaly detection
• Association
• Autoencoders

Reinforcement Learning
Reinforcement learning is less supervised and depends on the learning agent in determining the output solutions by arriving at different possible ways to achieve the best possible solution.

Q3. Describe the general architecture of Machine learning.



Business understanding:
Understand the give use case, and also, it's good to know more about the domain for which the use cases are built.

Data Acquisition and Understanding:
Data gathering from different sources and understanding the data. Cleaning the data, handling the missing data if any, data wrangling, and EDA(Exploratory data analysis).

Modeling:
Feature Engineering - scaling the data, feature selection - not all features are important. We use the backward elimination method, correlation factors, PCA and domain knowledge to select the features.
Model Training based on trial and error method or by experience, we select the algorithm and train with the selected features.
Model evaluation Accuracy of the model , confusion matrix and cross-validation.
If accuracy is not high, to achieve higher accuracy, we tune the model...either by changing the algorithm used or by feature selection or by gathering more data, etc.

Deployment -

Once the model has good accuracy, we deploy the model either in the cloud or Rasberry py or any other place. Once we deploy, we monitor the performance of the model.if its good...we go live with the model or reiterate the all process until our model performance is good.
It's not done yet!!!
What if, after a few days, our model performs badly because of new data. In that case, we do all the process again by collecting new data and redeploy the model.

Q4. What is Overfitting, and How Can You Avoid It?

Overfitting is a situation that occurs when a model learns the training set too well, taking up random fluctuations in the training data as concepts. These impact the model's ability to generalize and don't apply to new data.

When a model is given the training data, it shows 100 percent accuracy—technically a slight loss. But, when we use the test data, there may be an error and low efficiency. This condition is known as overfitting.

There are multiple ways of avoiding overfitting, such as:

Regularization. It involves a cost term for the features involved with the objective function
Making a simple model. With lesser variables and parameters, the variance can be reduced
Cross-validation methods like k-folds can also be used
If some model parameters are likely to cause overfitting, techniques for regularization like LASSO can be used that penalize these parameters

Q5. What Are the Three Stages of Building a Model in Machine Learning?

The three stages of building a machine learning model are:

- **Model Building**
  Choose a suitable algorithm for the model and train it according to the requirement
- **Model Testing**
  Check the accuracy of the 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 it is up-to-date.

Q6. What is feature engineering? How do you apply it in the process of modelling?

Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.

In layman terms, feature engineering means the development of new features that may help you understand and model the problem in a better way. Feature engineering is of two kinds — business driven and data-driven. Business-driven feature engineering revolves around the inclusion of features from a business point of view. The job here is to transform the business variables into features of the problem. In case of data-driven feature engineering, the features you add do not have any significant physical interpretation, but they help the model in the prediction of the target variable.

To apply feature engineering, one must be fully acquainted with the dataset. This involves knowing what the given data is, what it signifies, what the raw features are, etc. You must also have a crystal clear idea of the problem, such as what factors affect the target variable, what the physical interpretation of the variable is, etc.

Q7. What Are the Applications of Supervised Machine Learning in Modern Businesses?

Applications of supervised machine learning include:

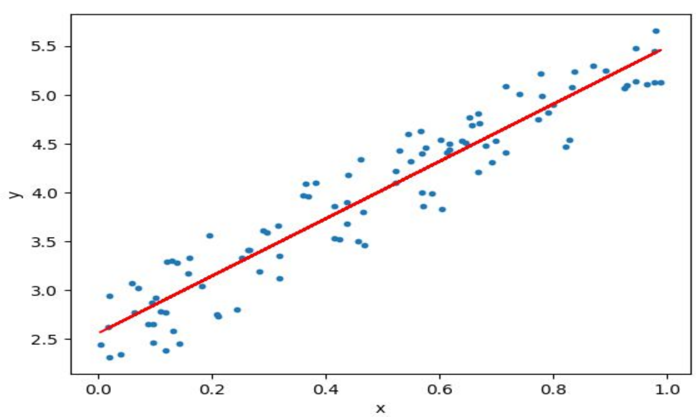
- Email Spam Detection
  Here we train the model using historical data that consists of emails categorized as spam or not spam. This labeled information is fed as input to the model.
- Healthcare Diagnosis
  By providing images regarding a disease, a model can be trained to detect if a person is suffering from the disease or not.

- Sentiment Analysis
  This refers to the process of using algorithms to mine documents and determine whether they’re positive, neutral, or negative in sentiment.
- Fraud Detection
  Training the model to identify suspicious patterns, we can detect instances of possible fraud.

Q8. What is Linear Regression?

Linear Regression tends to establish a relationship between a dependent variable(Y) and one or more independent variable(X) by finding the best fit of the straight line.

The equation for the Linear model is **Y = mX+c**, where m is the slope and c is the intercept



In simple terms, linear regression is a method of finding the best straight line fitting to the given data, i.e. finding the best linear relationship between the independent and dependent variables.

In technical terms, linear regression is a machine learning algorithm that finds the best linear-fit relationship on any given data, between independent and dependent variables. It is mostly done by the Sum of Squared Residuals Method.

Q9. What is L1 Regularization (L1 = lasso) ?

The main objective of creating a model(training data) is making sure it fits the data properly and reduces the loss. Sometimes the model that is trained will fit the data but it may fail and give a poor performance during analyzing of data (test data). This leads to overfitting. Regularization came to overcome overfitting.

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds “Absolute value of magnitude” of coefficient, as penalty term to the loss function.

$$Cost\ (W) = RSS(W) + \ \lambda * (sum\ of\ absolute\ value\ of\ weights)$$

$$= \sum_{i=1}^{N} \left\{ y_i \ - \sum_{j=0}^{M} w_j\ x_{ij} \right\}^2 + \lambda \sum_{j=0}^{M} |w_j|$$

Lasso shrinks the less important feature’s coefficient to zero; thus, removing some features altogether. So, this works well for feature selection in case we have a huge number of features.

Methods like Cross-validation, Stepwise Regression are there to handle overfitting and perform feature selection work well with a small set of features. These techniques are good when we are dealing with a large set of features.
Along with shrinking coefficients, the lasso performs feature selection, as well. Because some of the coefficients become exactly zero, which is equivalent to the particular feature being excluded from the model.

Q10. What is L2 (=Ridge) Regression?

Overfitting happens when the model learns signal as well as noise in the training data and wouldn’t perform well on new/unseen data on which model wasn’t trained on.
To avoid overfitting your model on training data like cross-validation sampling, reducing the number of features, pruning, regularization, etc. So to avoid overfitting, we perform Regularization.

The Regression model that uses L2 regularization is called Ridge Regression.

The formula for Ridge Regression:-

$$Cost\ function = Loss + \frac{\lambda}{2m} * \sum \|w\|^2$$

Regularization adds the penalty as model complexity increases. The regularization parameter (lambda) penalizes all the parameters except intercept so that the model generalizes the data and won't overfit.
Ridge regression adds "squared magnitude of the coefficient" as penalty term to the loss function. Here the box part in the above image represents the L2 regularization element/term.
Lambda is a hyperparameter.
If lambda is zero, then it is equivalent to OLS. But if lambda is very large, then it will add too much weight, and it will lead to under-fitting.
Ridge regularization forces the weights to be small but does not make them zero and does not give the sparse solution.
Ridge is not robust to outliers as square terms blow up the error differences of the outliers, and the regularization term tries to fix it by penalizing the weights
Ridge regression performs better when all the input features influence the output, and all with weights are of roughly equal size.
L2 regularization can learn complex data patterns.

Q11. What is R square(where to use and where not)?

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.
The definition of R-squared is the percentage of the response variable variation that is explained by a linear model.

R-squared = Explained variation / Total variation

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

R-squared is always between 0 and 100%.

0% indicates that the model explains none of the variability of the response data around its mean.
100% indicates that the model explains all the variability of the response data around its mean.
In general, the higher the R-squared, the better the model fits your data.

There is a problem with the R-Square. The problem arises when we ask this question to ourselves. **Is it good to help as many independent variables as possible?**
The answer is No because we understood that each independent variable should have a meaningful impact. But, even** if we add independent variables which are not meaningful**, will it improve R-Square value?

Yes, this is the basic problem with R-Square. How many junk independent variables or important independent variable or impactful independent variable you add to your model, the R-Squared value will always increase. It will never decrease with the addition of a newly independent variable, whether it could be an impactful, non-impactful, or bad variable, so we need another way to measure equivalent RSquare, which penalizes our model with any junk independent variable.

So, we calculate the Adjusted R-Square with a better adjustment in the formula of generic R-square.

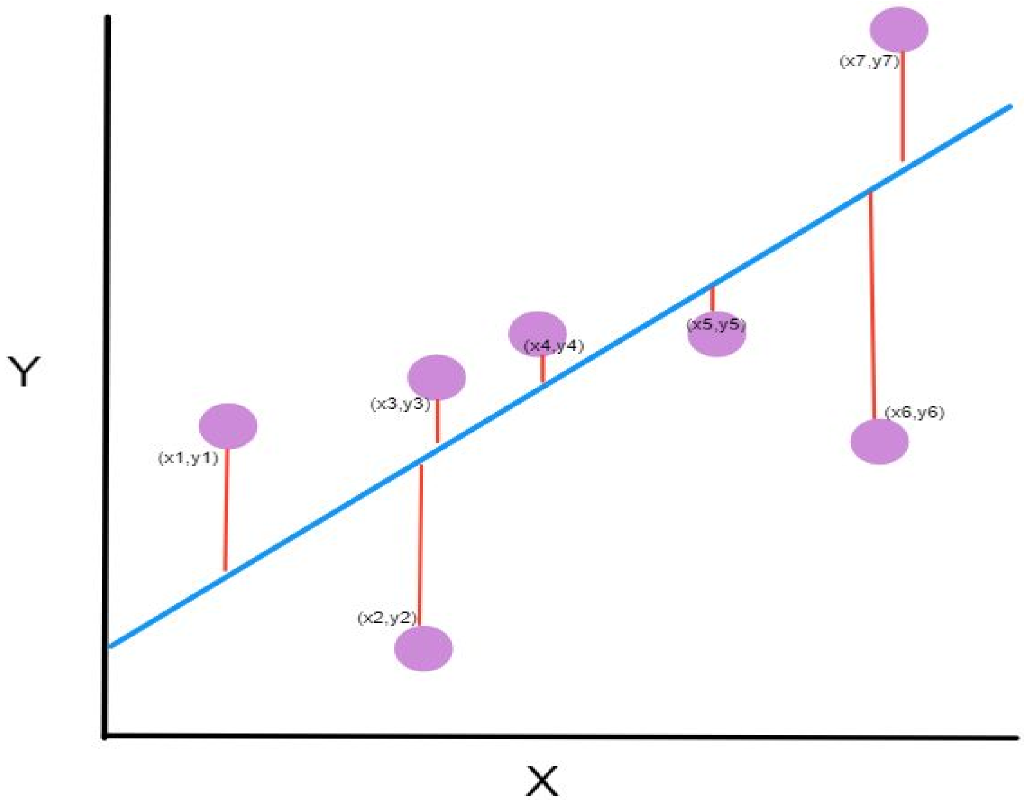
$$R^2_{\text{adjusted}} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

where
R^2 = sample R-square
p = Number of predictors
N = Total sample size.

Q12. What is Mean Square Error?

The mean squared error tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the “errors”) and squaring them.

Giving an intuition



The line equation is y=Mx+B. We want to find M (slope) and B (y-intercept) that minimizes the squared error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

$\hat{Y}_i$ = predicted values

Q13. How is hypothesis testing used in linear regression?

Hypothesis testing can be carried out in linear regression for the following purposes:

To check whether a predictor is significant for the prediction of the target variable. Two common methods for this are —

By the use of p-values:
If the p-value of a variable is greater than a certain limit (usually 0.05), the variable is insignificant in the prediction of the target variable.

By checking the values of the regression coefficient:
If the value of the regression coefficient corresponding to a predictor is zero, that variable is insignificant in the prediction of the target variable and has no linear relationship with it.
To check whether the calculated regression coefficients are good estimators of the actual coefficients.

Q14. What is the generalized linear model?

The generalized linear model is the derivative of the ordinary linear regression model. GLM is more flexible in terms of residuals and can be used where linear regression does not seem appropriate. GLM allows the distribution of residuals to be other than a normal distribution. It generalizes the linear regression by allowing the linear model to link to the target variable using the linking function. Model estimation is done using the method of maximum likelihood estimation.