- What is Bagging ? (decrease variance)

- An ensemble learning method that is commonly used to reduce variance within a noisy dataset. A random sample of data in a training set is selected with replacement.

- What is Boosting ? (decrease bias)

- Boosting creates a collection of predicters. It refers to a family of algorithms which converts weak learner to strong learner. It is an ensemble method for improving model predictions of given learning algorithm.

- What is pruning ?

- A data compression technique in ML that reduces the size of decision tree by removing sections of trees which are not required to classify instances

- What is pre-pruning ?

- Early stopping the growth of tree before it classify completely the training set.

- What is post-pruning ?

- Remove the sections of tree after it has completly grown.

- What is version space ?

- Intermediate space between general hypothesis space & specific hypothesis space.
  It is also hierarchial representation of knowledge.

- Disadvantages of Candidate Elimination Algo ?

- Fails for noisy | inconsistent data
- Uses partially learning, sometimes fails to predict right hypothesis for new training sample.

- Why we only consider particular rules for decision trees?

- A small change in data can cause a large change in the structure of decision tree causing unstability.

For Decision Trees, sometimes calculations can go for more complex compared to other algorithms.

Because of certain rules, decision trees requires less effort for data preparation during pre-processing.

- Applications of Machine Learning?

- Fraud detection
- Price prediction / Stock Markets Trading
- Products Recommendation
- Speech Recognition
- Self driving cars, Medical diagenosis

- What is Regression?

- A statistical method to predict continuous outcome based on one or more predicter variable values.

- What is Linear Regression?

- A supervised ML model in which it finds the best fit linear line between the independent and dependent variable.

- What is cost function?

- cost function (J) of Linear Regression is the Root Mean Squared error between predicted y value & true y value.

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

- What is hypothesis of Linear Regression?

θ_1 - Intercept

$h_\theta(x) = y = \theta_1 + \theta_2 \cdot x$ (θ_1, θ_2 are parameters that control the hypothesis y)

- What is gradient descent ?

- To update θ_1 & θ_2 values in order to minimize cost function (i.e. minimizing RMSE value) and achieving best fit line the model uses Gradient Descent.

Here, the idea is to start with random θ_1, θ_2 values and then iteratively update them, reaching minimum cost.

- Why Logistic Regression needed ?

- When our data has outliers, the best fit line predicted using linear regression may deviate and incorrectly predicts the output values.

- The output in linear regression is sometimes > 1 or < 0.

- What is sigmoid Function ?

- $F(z) = \frac{1}{1+e^{-z}}$; to prevent the output of linear regression (>1 or <0). This sigmoid function is an activation function that has range $(0, 1)$. Since we have to predict probability (lies between $(0, 1)$) this sigmoid Function is right choice.

- What is Bias ?

- The difference between what you expect to learn & truth.

- What is variance ?

- The difference between what you expect to learn & what you learnt.

small variance, high bias — underfitting (increase features)

high variance, small bias — overfitting (decision trees)

↓ Regularization

- What is Regularization ?

- Keep all features, but reduce parameter θ magnitude. It discourages learning more complex or flexible model, to prevent overfitting.

- Define Logistic Regression ?

- Regression used to fit a curve of data in which the dependent variable is Binary value or dichotomous

- What is Confusion Matrix ?

- A summarized table of no. of correct & incorrect predictions obtained ~~from~~ by a classifier or classification model. It is a performance measurement for ML algorithm.

Predicted values \ Actual values	1	0
1	TP	FP
0	FN	TN

TP - True Positive
FP - False Positive
FN - False Negative
TN - True Negative

TP - You predicted positive & it is True.

True Negative - You predicted negative & it is True.

False Negative - You predicted negative & it is False. (Type II error)

False Positive - You predicted positive & it is False. (Type I error)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (\text{as high as possible}) \qquad \text{Accuracy} = \frac{TP + TN}{\text{Total}} \quad (\text{high as possible})$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (\text{as high as possible})$$

- What is K-NN ? (An instance based Learner)

- A supervised ML algorithm that is used to solve both classification & regression problems.

It works by finding the distances between a query instance & all the data (training examples) nearest to query & then takes K-nearest examples & votes for most frequent output.

- Why is KNN called lazy Algorithm ?

- Because it doesnot learn a discriminate function from training data rather memorizes the training data & stores it. to classify new training example.

- What is difference between KNN & weighted-KNN ?

- Weighted KNN is modified version of KNN. While taking majority vote, the nearest neighbors vary widely in their distance & ~~closeby~~ closest neighbors more reliably indicate the class of object

- What is Support Vector Machine ?

- The objective of SVM algorithm is to find a hyperplane in an N-dimensional space (N - the no. of features) that distinctly classifies the data points.

- What is hyperplane ?

- In an n-dimensional has a flat, n-1 dimensional subset of that space that divides the space into two disconnected parts. Called hyperplane

- What is margin ?

- The distance between the hyperplanes drawn to differentiate the different classes.

- What is dimension of Hyperplane ?

– Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes.

The dimension of hyperplane depends upon the no. of features. If the no. of input features is 2, then hyperplane is just a line. If the no of input features is 3, then the hyperplane becomes 2-D plane.

- What are support vectors ?

– The data points that are closeer to the hyperplane and influence the position & orientation of hyperplane. Using these, we can maximize the margin of the classifier.

- Explain Large Margin Intuition ?

– We take the output of the linear function and if that output is greater than 1, identified is in other class. The reinforcement range [-1, 1] acts as margin (also called threshold values)

- what is hinge Loss ?

– The Loss function that helps maximize the margin is hinge loss.

- What is Kernal - trick ?

– A method of using linear classifier to solve a non - linear problem by mapping non - linear data into higher dimensional space. ~~without~~

- Why do we need Kernel trick ?

- If the data is not linearly seperable in 2-dimensional space, to build a linear classifier, we have to transform our data into 3D space while dealing with 2-D data.

- Types of Kernals ?

1) Linear Kernel -
2) Gaussian Kernel -
3) Polynomial Kernel -
4) String Kernel -
5) Chi-Square Kernel -

- What is Linear Kernel ?

- Used when data is linearly seperable, i.e. it can be seperated using single line.

- What is Gaussian Kernel ?

- Gaussian is one such Kernel that gives good linear seperation in higher dimension for many non linear problems.

- Logistic Regression (vs) SVM ?

- n = no. of features, m = no. of training examples

1. n is large than m : Use logistic regression or SVM without linear Kernel

2. n is small, m is intermediate : Use SVM with Gaussian Kernel

3. n is small, m is large : Create/add more features, then use logistic regression or linear SVM.

- SVM multiclass classification ?

- Use one (vs) all method. Train K SVMs, one to distinguish $y = i$ from the rest, get $\theta^{(1)}, \theta^{(2)} \ldots \theta^{(k)}$

  Then pick class i with the largest $[\theta^{(i)T} \cdot x]$

- SVM Regression ?

- Support Vector Regression is supervised learning algorithm that is used to predict discrete values. The basic idea behind SVR is to find best fit line.

  In SVR, the best fit line is the hyperplane that has the maximum no. of points. Here, SVR tries to fit the best line within the distance between hyperplane & boundary line.

- Disadvantages of SVM ?

- 1) Not suitable for large data.
  2) Doesn't fit for data with noise (target class are overlapping).
  3) no. of features >>> no. of training samples, SVM fails
  4) Long training time on datasets.
  5) choosing a "good" kernel is not easy.

- What is Radial Basis ?

$$K(x_1, x_2) = e^{\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)}$$

{ Find a non linear classifier or regression line.

σ - variance of hyperparameter.

$\|x_1 - x_2\|$ - Euclidean distance between x_1, x_2 points.