Note: Feel free to add questions, please ensure there is no redundancy and the questions are under the correct group.

Machine Learning

- **Mathematics**
  - Insightful Basics
    - How would you fit a curve to data?
      What assumptions would you make in the above process
    - What tricks could you use to normalise a distribution?
  - Ultra Basic Stuff
    - Tensors and vectors difference?
    - What is a matrix?
  - Basic Statistics
    - What is standard deviation?
    - What is the empirical relation between mean, median and mode?
    - What is descriptive statistics vs inferential statistics?
    - Give type of variables?
    - What is bessel's correction?
    - What are skewness and kurtosis and their effects?
    - Explain Law of large numbers
    - Correlations and covariance difference
    - Normalisation and standardization difference?
      And in which situation would you use which one?
  - Slightly more advanced statistics
    - How would one add two normal distributions
    - How to find multicollinearity between features in linear regression
    - Homoskedastic
      - If error residuals are symmetrical in nature, how would that affect linear regression?
  - Hypothesis testing
    - What is a hypothesis?
    - A test is administered annually. The test has a mean score of 150 and a standard deviation of 20. If Ravi's z-score is 1.50, what was his score on the test?
  - Probability and Bayes Theorem
  - P values
  - Types of distributions
    - What is Chebysheff's theorem?
    - Formula for a normal and multi-normal distribution?
  - Law of large numbers/ central limit theorem
    - What are the different types of Sampling methods that you have used?
  - Cosine Similarity, Correlations and covariance
    - Difference between correlation and covariance?
  - Eigen vectors and their utility [advanced]
  - Statistical Tests [advanced]

        - How to test multicollinearity?
    - Singular Value Decomposition [Advanced]
    - Principal Component Analysis [Advanced]

- **Metrics and model selection**
    - Bias - Variance [link] **link1**
        - Explain Bias - variance for KNN
        - Explain Bias - variance for linear regression
        - What is overfitting, why does it happen?
        - How to go about solving overfitting?
        - How would you go about reducing overfitting using deep neural networks?
        - Which types of algos are more likely to have overfitting?
        - What is underfitting, why does it happen and how to solve it?
        - What is generalization?
    - Model evaluation?
        - Explain the different ways you would evaluate a classification model?
            - What is a confusion matrix and how is it used?
            - What is precision?
            - What is recall?
            - When are they used and why, give example scenarios?
            - What is F1 score?
            - What is lift?
            - What is the expected value? And when would it be preferred as an evaluation metric?
            - What is leverage?
            - Difference between ROC curve and AUC curve?
            - Why do you use harmonic mean over normal mean for F1 score?
            - What is an F-beta score? Where can it be used ?
        - Regression
            - How to evaluate a model in case of regression?
        - Ranking
            - How to evaluate a model in case of ranking?
            - What happens to the confusion matrix if the threshold is changed?

    - Regression: MAE, MSE, r2 score, Adjusted r2 score
        - What is a statistical fit?
    - L1 and L2 regularization
        - Difference between L1 and L2 regularization, when to use what?
    - Cross validation
        - What are some issues with K fold CV?
        - What is K-fold cross validation?
    - Curse of Dimensionality?
    - Bayes Error/ Irreducible Error [link1] [advanced]

        - How can you estimate the max attainable accuracy/performance of your model?
        - Explain five different model evaluation techniques?
- **Unsupervised learning models**
    - Given data that has been clustered using two different clustering algorithms, how would you measure which one is better?

- **Supervised Learning Models**
    - Linear Regression
        - Assumptions of Linear Regression problem?[link]
        - Difference between Ridge, Lasso and elasticnet?
    - SVM [link1] [link2]
        - Loss function in SVM?
        - How to deal with non linearity?
        - Kernel Function?]
    - Decision Trees
        - Methods to improve Decision Trees?
        - Node splitting criteria?
        - What is entropy?
    - Logistic regression
        - How does a logistic regression model know what the coefficients are?
        - What is the loss function in logistic regression?
        - Why is sigmoid used in logistic regression?
        - What hapens if you use mean square instead of log odds?
    - Random Forests
        - Explain how it works?
    - Naive Bayes
        - Assumptions in Naive Bayes?
        - Multinomial vs Bernoulli?
    - Nearest Neighbours
        - Do we need feature scaling/normalization for KNN? [link]
    - MLP
    - Bayes Classifier [Advanced]
    - Kernel Density estimation [Advanced]
    - Ensemble Models
        - Difference between weak learners and strong learners?
        - Bagging vs Boosting?
        - What is blending?
        - What is adaboost?
        - What is xgboost?
        - Gradient boosting?
    - Case
        - Given two different datasets, one of size 5000 rows and another with a billion rows, on which would you use xgboost and on which random forest?

- **Loss Function and related things link**
  - Cross entropy
  - Hinge loss
  - Huber loss
  - MSE
    - Can we use MSE for logistic regression? link
  - Gradient Descent, SGD etc
    - What is the exploding gradient and the vanishing gradient? Link link
    - Difference between GD and SGD?
    - ADAGRAD, ADAM and momentum
    - In deep learning, why do we NOT take the derivative of the loss function and equate it to zero to find the minimum loss and the corresponding weights?

- **Deep Learning**
  - When should one typically **not** prefer deep learning models?
  - What are manifolds and how are they related to deep learning?
  - What is Batch Normalization?
  - Weight Initialization (Xavier, Standard Normal etc)
  - What happens if you initialize all weights to zero?
  - Activation function
    - Softmax vs Sigmoid
    - In a Neural Network, why does one need to have an activation function?
    - How sigmoid activation ensures non-linear decision boundary ? ie. Explain UAT through activation function.
    - What if we don't use an activation function in NN?
    - Tan h, relu and leaky relu properties, usages and comparisons and the types of data they may be used with. Link
  - Backpropagation & Regularization
    - Explain Back Propagation mathematically.
    - Why do gradient explode happen what are some regularization methods Link
    - Is Gradient clipping scalable ? If not, what are some effective ways?
    - What makes the concept of Drop out this significant?
    - Explain neuron co-dependence
    - Where should you use a Dropout layer in the dense layers ? In Top bunk or Bottom bunk ? Why there ?
  - Universal Approximator
    - Implement boolean gates using MLP Link
    - How would you go about designing a neural network?
  - CNN
    - What is the major difference between MLP and CNN models? OR Why did we move from MLP to CNNs? Link
    - Find the complexity of the CNN system?
    - What is the pooling layer?

        - Dropout?
        - Talk about positional invariance?
        - What is stride? How does it affect the output features size?
    - RNN
        - What activation function is generally used in RNNs and why?
        - How would different activation functions behave in RNNs? (Tanh vs sigmoid vs Relu vs Linear)
        - What kind of problems can RNN tackle which MLP fails to?
        - Backprop of RNN?
        - What is the utility of a bi-directional RNN?
    - LSTMs
        - Explain LSTMs in layman terms

- **NLP**
    - Bag of words
    - N gram
    - Named entity recognition
    - Topic modelling?
        - LDA
        - Probabilistic topic models
    - Text representation- Tf -idf
        - Relation between IDF and entropy
    - Word2vec
        - Why do we use a linear activation function in the hidden layer?
        - Skipgram vs CBOW?
        - How is w2v better than tfidf?
    - How would you build a spell checker for chatbot?
    - https://medium.com/modern-nlp/nlp-interview-questions-f062040f32f7
    - BERT

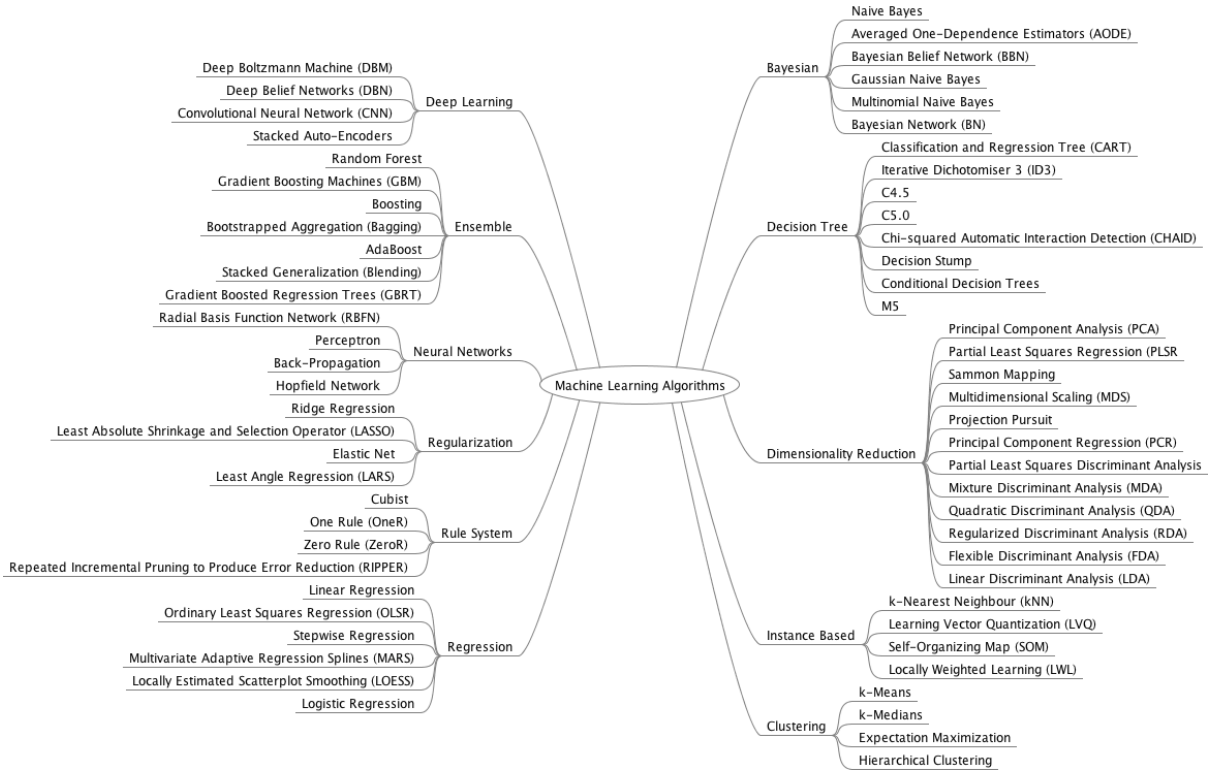
- **Generative Models**
    - Autoencoder
    - Variational Autoencoder [Advanced]
    - GANs [Advanced]
        - Tell me a peculiar difference between a normal error function and a GANs' error function
- **SQL**
- **Programming**
    - Incoming stream of numbers, how would you split them, so that they go into different machines and then rejoin them back again?
- **Other Analytical Questions**
    - What is co-occurance grouping?
    - What is profiling?
    - What is market basket analysis?
    - What is causal modelling?

    - In which kinds of cases would you use data reduction techniques and how would it affect analysis

- **Random Questions**
    - What's the kernel trick?
    - Discriminative vs generative?
    - Difference between regression and classification?
    - Supervised learning vs unsupervised learning?
    - What is inductive learning?
    - What is deductive learning?
    - What is the violin plot and what does it show us?
    - What are the different methods of visualizing data?
    - What is a recommendation system?
    - How does collaborative filtering work?
    - Parametric vs Non- Parametric modelling? [advanced]
    - Transfer Learning [advanced]

Cases

- How would you go about profiling wait times of customers who call into a call center?
    - Sub question - how would your answer change if the distribution of wait times vs proportions of calls is not normal? How could you go about normalising it?
- How would you go about finding latent dimensions of a particular data?
- What are the different ways to recommend movies to people, think in terms of link prediction?
- How would you target customers for search advertising?
- How would you target customers for display advertising? How are your customers different from search advertising ones and what groups of customers do you have?
- A hurricane hit and a Walmart executive decided it's an opportunity. How would you use your data science skills to narrow down on the precise opportunity?


Machine Learning Algorithms
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression