**Linear Algebra**

1. Define Point/Vector (2-D, 3-D, n-D)?

2. How to calculate Dot product and angle between 2 vectors?

3. Define Projection, unit vector?

4. Equation of a line (2-D), plane(3-D) and hyperplane (n-D)?

5. Distance of a point from a plane/hyperplane, half-spaces?

6. Equation of a circle (2-D), sphere (3-D) and hypersphere (n-D)?

7. Equation of an ellipse (2-D), ellipsoid (3-D) and hyper ellipsoid (n-D)?

8. Square, Rectangle, Hyper-cube and Hyper-cuboid?

**Probability And Statistics**

1. What is Random variables: discrete and continuous?

2. Define Outliers (or) extreme points?

3. What is PDF?

4. What is CDF?

5. explain about 1-std-dev, 2-std-dev, 3-std-dev range?

6. What is Symmetric distribution, Skewness and Kurtosis?

7. How to do Standard normal variate (z) and standardization?

8. What is Kernel density estimation?

9. Importance of Sampling distribution & Central Limit theorem.?

10. Importance of Q-Q Plot: Is a given random variable Gaussian distributed?

11. What is Uniform Distribution and random number generators?

12. What Discrete and Continuous Uniform distributions?

13. How to randomly sample data points?

14. Explain about Bernoulli and Binomial distribution?

15. What is Log-normal and power law distribution?

16. What is Power law & Pareto distributions: PDF, examples

17. Explain about Box-Cox/Power transform?

18. What is Co-variance?

19. Importance of Pearson Correlation Coefficient?

20. Importance Spearman Rank Correlation Coefficient?

21. Correlation vs Causation?

22. What is Confidence Interval?

23. Confidence Interval vs Point estimate?

24. Explain about Hypothesis testing?

25. Define Hypothesis Testing methodology, Null-hypothesis, test-statistic, p-value?

26. How to do K-S Test for similarity of two distributions?

**Dimensionality Reduction**

1. What is dimensionality reduction?

2. Explain Principal Component Analysis?

3. Importance of PCA?

4. Limitations of PCA?

5. What is t-SNE?

6. What is Crowding problem?

7. How to apply t-SNE and interpret its output?

**Performance Measurement Models:**

1. What is Accuracy?

2. Explain about Confusion matrix, TPR, FPR, FNR, TNR?

3. What do you understand about Precision & recall, F1-score? How would you use it?

4. What is the ROC Curve and what is AUC (a.k.a. AUROC)

5. What is Log-loss and how it helps to improve performance?

6. Explain about R-Squared/ Coefficient of determination?

7. Explain about Median absolute deviation (MAD)?

8. Define Distribution of errors?

**Classification algorithms in various situations:**

1. What is Imbalanced and balanced dataset?

2. Define multi-class classification?

3. Explain Impact of Outliers?

4. What is Local Outlier Factor?

5. What is k-distance (A), N(A)?

6. Define reachability-distance (A, B)?

7. What is Local-reachability-density(A)?

8. Define LOF(A)?

9. Impact of Scale & Column standardization?

10. What is Interpretability?

11. Handling categorical and numerical features?

12. Handling missing values by imputation?

13. Bias-Variance tradeoff?

**K-NN (K Nearest Neighbour)**

1. Explain about K-Nearest Neighbors?

2. Failure cases of KNN?

3. Define Distance measures: Euclidean(L2), Manhattan(L1), Minkowski, Hamming?

4. What is Cosine Distance & Cosine Similarity?

5. How to measure the effectiveness of k-NN?

6. Limitations of KNN?

7. How to handle Overfitting and Underfitting in KNN?

8. Need for Cross validation?

9. What is K-fold cross validation?

10. What is Time based splitting?

11. Explain k-NN for regression?

12. Weighted k-NN?

13. How to build a KD-tree?

14. Find nearest neighbors using KD-tree?

15. What is Locality sensitive Hashing (LSH)?

16. Hashing vs LSH?

17. LSH for cosine similarity?

18. LSH for Euclidean distance?

**Naive Bayes:**

1. What is Conditional probability?

2. Define Independent vs Mutually exclusive events?

3. Explain Bayes Theorem with example?

4. How to apply Naive Bayes on Text data?

5. What is Laplace/Additive Smoothing?

6. Explain Log-probabilities for numerical stability?

7. In Naive bayes how to handle Bias and Variance tradeoff?

8. What Imbalanced data?

9. What is Outliers and how to handle outliers?

10. How to handle Missing values?

11. How to Handle Numerical features (Gaussian NB)?

12. Define Multiclass classification?

**Logistic Regression and Linear Regression:**

1. Explain about Logistic regression?

2. What is Sigmoid function & Squashing?

3. Explain about Optimization problem in logistic regression.

4. Importance of Weight vector in logistic regression.

5. L2 Regularization: Overfitting and Underfitting.

6. L1 regularization and sparsity.

7. What is Probabilistic Interpretation: Gaussian Naive Bayes?

8. Explain about Hyperparameter search: Grid Search and Random Search?

9. What is Column Standardization?

10. Explain about Collinearity of features?

11. Find Train & Run time space and time complexity of Logistic regression?

**Support Vector Machine:**

1. Explain About SVM?

2. What is Hinge Loss?

3. Dual form of SVM formulation?

4. What is Kernel trick?

5. What is Polynomial kernel?

6. What is RBF-Kernel?

7. Explain about Domain specific Kernels?

8. Find Train and run time complexities for SVM?

9. Explain about SVM Regression?

**Decision Trees:**

1. How to Build a decision Tree?

2. What is Entropy?

3. What is information Gain?

4. What is Gini Impurity?

5. How to Construct a DT?

6. Importance of Splitting numerical features?

7. How to handle Overfitting and Underfitting in DT?

8. What are Train and Run time complexity for DT?

9. How to implement Regression using Decision Trees?

**Ensemble Models:**

1. What are ensembles?

2. What is Bootstrapped Aggregation (Bagging)?

3. Explain about Random Forest and their construction?

4. Explain about Boosting?

5. What are Residuals, Loss functions and gradients?

6. Explain about Gradient Boosting?

7. What is Regularization by Shrinkage?

8. Explain about XGBoost?

9. Explain about AdaBoost?

10. How do you implement Stacking models?

11. Explain about cascading classifiers?

**Clustering:**

1. What is K-means? How can you select K for K-means?

2. How is KNN different from k-means clustering?

3. Explain about Hierarchical clustering?

4. Limitations of Hierarchical clustering?

5. Time complexity of Hierarchical clustering?

6. Explain about DBSCAN?

7. Advantages and Limitations of DBSCAN?

**Recommender Systems and Matrix Factorization:**

1. Explain about Content based and Collaborative Filtering?

2. What is PCA, SVD?

3. How is KNN different from k-means clustering?

4. Explain about Hierarchical clustering?

5. Limitations of Hierarchical clustering?

6. Time complexity of Hierarchical clustering?

7. Explain about DBSCAN?

8. Advantages and Limitations of DBSCAN?

9. What is NMF?

10. How to do MF for Collaborative filtering?

11. How to do MF for feature engineering?

12. Explain relation between Clustering And MF?

13. What is Hyperparameter tuning?

14. Explain about Cold Start problem?

15. How to solve Word Vectors using MF?

16. Explain about Eigenfaces?