

1. What is the optimization equation of GBDT ?
2. Write the formulation of hinge loss ?
3. What is the train time complexity of KNN ?
4. What is the Test time complexity of KNN in brute force ?
5. What is the test time complexity of KNN if we use kd-tree ?
6. How will you regularise the KNN model ?
7. Which of these model are preferable when we have low complexity power ?
 - a. SVM
 - b. KNN
 - c. Linear Regressions
 - d. XGboost
8. What is Laplace smoothing ?
9. How will you regularise your naive bayes model ?
10. Can we solve dimensionality reduction with SGD?
11. Which one will be doing more competitions GD or SGD ?
12. If A is a matrix of size (3,3) and B is a matrix of size (3,4) how many numbers of multiplications that can happen in the operations $A*B$?
13. What is the optimization equation of Logistic Regression ?
14. How will you calculate the $P(x/y=0)$ in case of gaussian naive baiyes ?
15. Write the code for proportional sampling.
16. What are hyperparameters in kernel svm ?
17. What are hyperparameters in SGD with hinge loss ?
18. Is hinge loss differentiable if not how we will modify it so that you apply SGD ?
19. Difference between ADAM vs RMSPROP ?
20. What is RMSPROP?
21. What is ADAM ?
22. What is the maximum and minimum values of gradient of the sigmoid function ?
23. What is RELU? Is it differentiable ?
24. What is F1 score ?
25. What is precission and recall ?
26. Name a few weight initialization techniques ?
27. Which of these will have more numbers of tunable parameters?
 - a. $7,7,512 \Rightarrow \text{flatern} \Rightarrow \text{Dense}(512)$
 - b. $(7,7,512) \Rightarrow \text{Conv}(512,(7,7))$
28. What is overfitting and underfitting ?
29. What do you do if a deep learning model is overfitting?
30. What is the batch Normalization layer ?
31. Write keras code to add a BN layer in an existing network ?
32. Number of tunable parameters in the BN layer.
33. What is convolution operation?
34. Number of parameters in a convolution neural network given in architecture
35. What are the inputs required to calculate the average f1 score ?
36. What macro average f1 score for 5 class classification problem.

37. How do you get probabilities for RF classifier outputs.
38. Is the Calibration classifier required to get probability values for logistic regression.?
39. How does kernel sum work in test time ?
40. What kind of base learners are preferable in random forest classifiers ?
41. How does bootstrapping work in RF classification.
42. Difference between one vs rest and one vs one.
43. Which one is better is one vs rest and one vs one.
44. What will happen if gamma increases in RBF kernel sum.
45. Explain linear regression.
46. What is difference between one hot encoding and a binary bow?
47. Kernel svm and linear svm (SGD classifier with hinge loss). Which has low latency and why.
48. Explain bayes theorem.
49. How to decrease the test time complexity of a logistic regression model.
50. What is the need for sigmoid function in logistic regression.
51. Why we need Calibration ?
52. What is MAP ? (mean average precision)
53. Why do we need gated mechanism in LSTM ?
54. What is stratified sampling ? Explain.
55. How do you compare two distributions ?
56. What will happen to train time of K means if data is very high dimension.
57. If you have 10mill records with 100dimension each for a clustering task. Which algorithm will you try first and why ?
58. What is matrix Factorization? Explain with an Example.
59. Which algorithm will give high time complexity if you have 10million records for a clustering task.
60. Difference between GD and SGD.
61. Which one will you choose GD or SGD? Why ?
62. Why do we need repetitive training of a model ?
63. How do you evaluate the model after productionization ?
64. What is need for laplace smoothing in N.B
65. Explain Gini impurity.
66. Explain entropy?
67. How to do multi-class classification with random forest ?
68. What is k-fold cross validation ?
69. What is need for CV ?
70. How do you do CV for a test classification problem using random search.
71. Assume We have very high dimension data. Which model will you try and which model will be better in a classification problem.
72. What is AUC?
73. Tell me one business case where recall is more important than precision.
74. Tell me one business case where precision is more important.
75. Can we use accuracy for very much imbalance data? If yes/no , why ?

76. Difference between micro average F1 and macro average F1 for a 3 class classification.
77. Difference between AUC and accuracy ?
78. How do we calculate AUC for a multiclass classification.
79. Test the complexity of Kernel sum ?
80. Can we use TSNE for dimensionality reduction i.e convert the data n to d dimension.
81. What is pearson correlation coefficient ?
82. Training time complexity of naive bayes ?
83. Numbers of tunable parameters in maxpooling layer ?
84. 100,50) -> Embeddylayer (36) -> output shape ?
85. Number of tunable parameters in embedding layer (36, vocab size = 75)
86. Relation between KNN and kernel sum ?
87. Which is faster
 - a. SVC(C=1). Fit(x,y)
 - b. SGD(Log=hinge).fit(x,y)
88. Explain about KS test ?
89. What is KL divergence ?
90. How QQ plot works ?
91. What is the need of confidence interval ?
92. How do you find the outliers in the given data set ?
93. Can you name a few sorting algorithms and their complexity ?
94. What is the time complexity of "a in list ()" ?
95. What is the time complexity of "a in set ()" ?
96. What is percentile ?
97. What is IQR ?
98. How do you calculate the length of the string that is available in the data frame column ?
99. Can you explain the dict.get() function ?
100. Is list is hash table ?
101. Is tuple is hash table ?
102. What is parameter sharing in deep learning?
103. What will be the alpha value for non support vectors.
104. What will be the effect of increasing alpha values in multinomial NB ?
105. What is recurrent equation of RNN output function ?
106. What is the minimum and maximum value of tanh ?
107. How many thresholds we need to check for a real valued features in DT ?
108. How do you compute the feature importance in DT ?
109. How do you compute the feature importance in SVM ?
110. Prove that L1 will given sparsity in the weight vector ?
111. What are L1,L2 regularizers ?
112. What is elastic net ?
113. What are the assumption of NB ?
114. What are the assumptions of KNN ?
115. What are the assumptions of linear regression ?
116. Write the optimization equation of linear regression ?

117. What is time complexity of building KD tree ?
118. What is the time complexity to check if a number is prime or not ?
119. Angle between two vectors (2,3,4) (5,7,8).
120. Angle between the weigh vector of $2x+3y+5=0$ and the vector(7,8).
121. Distance between (7,9) and the line $7x+4y-120=0$.
122. Distance between the lines $4x+5y+15=0$, $4x+5y-17=0$.
123. Which of this hyperplane will classify these two class points
a. P: (2,3), (-3,4) N: (-5,7), (-5,-9)
b. $4x+5y+7=0$, $-3y+3x+9=0$
124. Which of the vector pairs perpendicular to each other
a. (3,4,5) (-3,-4,5)
b. (7,4,6) (-4,-7,-12)
125. How dropout works ?
126. Explain the back propagation mechanism in dropout layers ?
127. Explain the loss function used in auto encoders assuming the network accepts images ?
128. Numbers of tunable parameters in dropout layer ?
129. When F1 score will be zero? And why ?
130. What is the need of dimensionality reduction.
131. What happens if we do not normalize our dataset before performing classification using KNN algorithm.
132. What is standard normal variate ?
133. What is the significance of covariance and correlation and in what cases can we not use correlation.
134. How do we calculate the distance of a point to a plane.
135. When should we choose PCA over t-sne.
136. How is my model performing if
a. Train error and cross validation errors are high.
b. Train error is low and cross validation error is high.
c. Both train error and cross validation error are low.
137. How relevant / irrelevant is time based epitting of data in terms of weather forecasting ?
138. How is weighted knn algorithm better simple knn algorithm.
139. What is the key idea behind using a kdtree.
140. What is the relationship between specificity and false positive rate.
141. What is the relationship between sensitivity,recall,true positive rate and false negative rate?
142. What is the alternative to using euclidean distance in Knn when working with high dimensional data ?
143. What are the challenges with time based splitting? How to check whether the train / test split will work or not for given distribution of data ?
144. How does outliers effect the performance of a model and name a few techniques to overcome those effects.

145. What is reachability distance
146. What is the local reachability density ?
147. What is the need of feature selection ?
148. What is the need of encoding categorical or ordinal features.
149. What is the intuition behind bias-variance tradeoff ?
150. **Can we use algorithm for real time classification of emails.**
151. What does it mean by precision of a model equal to zero is it possible to have precision equal to 0.
152. What does it mean by $FPR = TPR = 1$ of a model.
153. What does $AUC = 0.5$ signifies.
154. When should we use log loss, AUC score and F1 score.
155. What performance metric should use to evaluate a model that see a very less no. of positive data points as compared to -ve data points.
156. What performance metric does t-sne use to optimize its probabilistic function.
157. What happens in laplace smoothing in my smoothing factor ' α ' is too large.
158. When to use cosine similarity over euclidean distance.
159. What is fit, transform and fit transform in terms of BOW,tf-idf,word2vector.
160. How do we quantify uncertainty in probability class labels when using KNN model for classifications.
161. How do we identify whether the distribution of my train and test is similar or not.
162. What does it mean by embedding high dimensional data points to a lower dimension ? what are the advantages and disadvantages of it.
163. What is the crowding problem w.r.t t-sne.
164. What is the need of using log probabilities instead of normal probabilities in naive bayes.
165. What do you mean by hard margin SVM ?
166. What is kernel function in svm ?
167. Why do we call an svm a maximum margin classifier ?
168. Is svm affected by outliers ?
169. What is locality sensitive hashing ?
170. What is sigmoid function? What is its range ?
171. Instead of sigmoid function can we use any other function in LR?
172. Why is accuracy not a good measure for classification problem ?
173. How to deal with multiclass classification problem using logistic regression ?
174. Can linear regression be used for classification purpose ?
175. What is the use of ROC curve ?
176. When EDA should be performed, before or after splitting data? Why ?
177. How k-nn++ is different from k-means clustering ?
178. Where ensemble techniques might be useful ?
179. What is feature forward selection ?
180. What is feature backward selection ?
181. What is type 1 & type 2 error ?
182. What is multicollinearity ?

183. How is eigenvector different from other general vectors ?
184. What is eigenvalue & eigenvectors ?
185. What is A/B testing
186. How to split data which has temporal nature.
187. What is response encoding of categorical features ?
188. What is the binning of continuous random variables.
189. Regularization parameter in dual form of SVM ?
190. What is the difference between sigmoid and softmax ?
191. For a binary classification which among the following cannot be the last layer ?
 - a. sigmoid(1)
 - b. sigmoid(2)
 - c. softmax(1)
 - d. softmax(2)
192. What is P-value in hypothesis testing ?
193. How to check if a particular sample follows a distribution or not ?
194. What is the difference between covariance and correlation ?
195. On what basis would you choose agglomerative clustering over k means clustering and vice versa ?
196. What is the metric that we use to evaluate unsupervised models.
197. What is the difference between model parameters and hyper parameters ?
198. Number of parameters in LSTM is $4m(m+n+1)$. How many number of parameters do we have in GRU ?
199. What is box cox transform? When can it be used ?
200. In what format should the data be sent to embedding layer?
201. What does trainable = true/false mean in embedding layer ?
202. What happens when we set return sequence = true in LSTM ?
203. Why are RNN'S and CNN'S called weight shareable layers ?
204. What happens during the fit and transform of following modules ?
 - a. Standard scaler
 - b. Count vectorizer
 - c. PCA
205. Can we use t-sne for transforming test data ? if not why ?
206. Find the sum of diagonals in the numpy array ?
207. Write the code to get the count of row for each category in the dataframes.
208. Difference between categorical cross entropy and binary cross entropy.
209. When you use w2v for test factorization, and we each sentence is having different words how can you forward data into models ?
210. What is tf idf weighted w2v ?
211. How to you use weighted distance in content based recommendation ?
212. What is the time complexity of SVD decomposition ?
213. What is the difference between content based recommendation and collaborative recommendation ?
214. Why do you think inertia actually works in choosing elbow point in clustering ?

215. What is gradient clipping ?
216. Which of these layers will be a better option as a last layer in multilabel classification ?
 - a. Sigmoid
 - b. Softmax
217. Is there a relation or similarity between LSTM and RESNET ?
218. What are the values returned by np.histogram()
219. What is PDF, can we calculate PDF for discrete distribution ?
220. Can the range of CDF be (0.5 - 1.5).
221. Number of parameters in the following network :
 - a. Number of neurons = 4
 - b. Problem = binary classification
 - c. no: of FC = 2
 - d. Neurons in 1st FC = 5
 - e. Neurons in 2nd FC = 3
222. How do we interpret alpha in dual form of sum? What is the relation between C and Alpha?
223. How does back propagation work in case of LSTM?
224. What is the difference between supervised and unsupervised models?
225. What is the derivative of this fraction $1/(1+e^{\sin x})$.
226. What will be the output of $a = [1\ 2\ 3\ 10], [4\ 5\ 6\ 11], [7\ 8\ 9\ 12]$ $a[:, :-1]$
227. What is the output of this $a = [1\ 5\ 9], [2\ 6\ 10], [3\ 7\ 11], [4\ 8\ 12]$ $a[:, -2, :]$
228. What will be the output of
 - a. `a= dict()`
 - b. `a[('a','b')]` = 0
 - c. `a[(a,b)]` = 1
 - d. `print(a)`
229. What will be the output of
 - a. `a = [1 2 3],[4 5 6],[7 8 9]`
 - b. `np.mean(a,axis=1)`
230. What will be the output of
 - a. `a =[3 4 5],[6 7 8],[9 10 11]`
 - b. `b = [1 2 3],[4 5 6],[7 8 9]`
 - c. `np.stack((a,b), axis= 0)`
231. What is "local outlier factor"?
232. How RANSAC works?
233. What are jaccard & Cosine Similarities
234. What are assumptions of Pearson correlation ?
235. Differences between Pearson and Spearman correlation?
236. What is the train time complexity of DBSCAN?
237. Explain the procedure of "prediction in hierarchical clustering"
238. Relation between knn and kernel SVM
239. Proof of "convergence of kmeans"

240. What is the optimal value of minpoints for the data (1000, 50)