Commonly Asked ML Interview Questions

1. What is the difference between Parametric and Non Parametric Algorithms?
2. Difference between convex and non-convex cost function; what does it mean when a cost function is non-convex?
3. How do you decide when to go for deep learning for a project?
4. Give an example of when False positive is more crucial than false negative and vice versa
5. Why is “Naive” Bayes naive?
6. Give an example where the median is a better measure than the mean
7. What do you mean by the unreasonable effectiveness of data?
8. Why KNN is known as a lazy learning technique?
9. What do you mean by semi supervised learning?
10. What is an OOB error and how is it useful?
11. In what scenario decision tree should be preferred over random forest?
12. Why Logistic Regression is called regression?
13. What is Online Machine Learning? How is it different from Offline machine learning? List some of it’s applications
14. What is No Free Lunch Theorem?
15. Imagine you are woking with a laptop of 2GB RAM, how would you process a dataset of 10GB?
16. What are the main differences between Structured and Unstructured Data?
17. What are the main points of difference between Bagging and Boosting?
18. What are the assumptions of linear regression?
19. How do you measure the accuracy of a Clustering Algorithm?
20. What is Matrix Factorization and where is it used in Machine Learning?
21. What is an Imbalanced Dataset and how can one deal with this problem?

22. How do you measure the accuracy of a recommendation engine?
23. What are some ways to make your model more robust to outliers?
24. How can you measure the performance of a dimensionality reduction algorithm on your dataset?
25. What is Data Leakage? List some ways using which you can overcome this problem.
26. What is Multicollinearity? How to detect it? List some techniques to overcome Multicollinearity.
27. List some ways using which you can reduce overfitting in a model.
28. What are the different types of bias in Machine Learning?
29. How do you approach a categorical feature with high cardinality?
30. Explain Pruning in Decision Trees and how it is done
31. What is ROC-AUC curve? List some of it's benefits.
32. What are kernels in SVM? Can you list some popular SVM kernels.
33. What is the difference between Gini Impurity and Entropy? Which one is better and why?
34. Why does L2 regularization give sparse coefficients?
35. List some ways using which you can improve a model's performance.
36. Can PCA be used to reduce the dimensionality of a highly nonlinear dataset?
37. What's the difference between probability and likelihood?
38. What cross-validation technique would you use on a time series data set.
39. Once a dataset's dimensionality has been reduced, is it possible to reverse the operation? If so, how? If not, why?
40. Why do we always need the intercept term in a regression model??
41. When Your Dataset Is Suffering From High Variance, How Would You Handle It?
42. Which Among These Is More Important Model Accuracy Or Model Performance?

43. What is active learning and where is it useful?
44. Why is Ridge Regression called Ridge?
45. State the differences between causality and correlation?
46. Does it make any sense to chain two different dimensionality reduction algorithms?
47. Is it possible to speed up training of a bagging ensemble by distributing it across multiple servers?
48. If a Decision Tree is underfitting the training set, is it a good idea to try scaling the input features?
49. Say you trained an SVM classifier with an RBF kernel. It seems to underfit the training set: should you increase or decrease γ (gamma)? What about C?
50. What is cross validation and it's types?
51. How do we interpret weights in linear models?
52. Which Gradient Descent algorithm (among those we discussed) will reach the vicinity of the optimal solution the fastest? Which will actually converge?
53. Why is it important to scale the inputs when using SVMs?
54. What is p value and why is it important?
55. What is OvR and OvO for multiclass classification and which machine learning algorithm supports this
56. How will you do feature selection using Lasso Regression?
57. What is the difference between loss function and cost function?
58. What are the common ways to handle missing data in a dataset?
59. What is the difference between standard scaler and minmax scaler? What you will do if there is a categorical variable?
60. What types of model tend to overfit?
61. What are some advantages and Disadvantages of regression models and tree based models.

62. What are some important hyperparameters for XGBOOST
63. Can you tell the complete life cycle of a data science project?
64. What are the properties of a good ML model?
65. What are the different evaluation metrices for a regression model?
66. What are the different evaluation metrices for a classification model?
67. Difference between R2 and adjusted R2? Why do you preffer adjusted r2?
68. List some of the drawbacks of a Linear model
69. What do you mean by Curse of Dimensionality?
70. What do you mean by Bias variance tradeoff?
71. Explain Kernel trick in SVM
72. What is the main difference between Machine Learning and Data Mining?
73. Why sometimes it is needed to scale or normalise features?
74. What is the difference between Type 1 and Type 2 error?
75. What is the difference between a Generative model vs a Discriminative model?
76. Why binary_crossentropy and categorical_crossentropy give different performances for the same problem?
77. Why does one hot encoding improve machine learning performance?
78. Considering the long list of machine learning algorithm, given a data set, how do you decide which one to use?
79. Differentiate between wide and tall data formats?
80. What is the difference between inductive machine learning and deductive machine learning?
81. How will you know which machine learning algorithm to choose for your classification problem?
82. What is the difference between Covariance and Correlation
83. How will you find the correlation between a categorical variable and a continuous variable ?

84. What are the differences between “Bayesian” and “Frequentist” approach for Machine Learning?
85. What is the difference between stochastic gradient descent (SGD) and gradient descent ?
86. What is the difference between Gaussian Mixture Model and K-Means Algorithm?
87. Is more data always better?
88. How can you determine which features are the most im- portant in your model?
89. Which hyper-parameter tuning strategies (in general) do you know?
90. How to select K for K-means?
91. Describe the differences between and use cases for box plots and histograms
92. How would you differentiate between Multilabel and MultiClass classification?
93. What is KL divergence, how would you define its usecase in ML?
94. Can you define the concept of Undersampling and Oversampling?
95. Considering a Long List of Machine Learning Algorithms, given a Data Set, How Do You Decide Which One to Use?
96. Explain the difference between Normalization and Standardization
97. List the most popular distribution curves along with scenarios where you will use them in an algorithm.
98. List all types of popular recommendation systems?
99. Which metrics can be used to measure correlation of categorical data?
100. Which type of sampling is better for a classification model and why?