MACHINE LEARNING

ML FUNDAMENTALS

1. What is Machine Learning? How is it different from traditional programming?

2. What is the difference between AI, ML, and Deep Learning?

3. What is supervised learning? Give examples.

4. What is unsupervised learning? Give use cases.

5. What is semi-supervised learning?

6. What is reinforcement learning? Explain the basic idea.

7. What is the difference between classification and regression?

8. What are labelled data and unlabelled data?

9. What are training data, validation data, and test data?

10. What is a model? How is it different from an algorithm?

ML Workflow

11. What is the typical workflow of an ML project?

12. Why is problem definition important?

13. What is the importance of data collection and data quality?

14. What is feature engineering? Why is it important?

15. How do you perform model selection?

16. What is hyperparameter tuning?

17. How do you evaluate a model?

18. What does model deployment mean?

19. Why is model monitoring necessary?

20. What is an ML pipeline?

21. What is bias in ML?

22. What is variance?

23. Explain the bias-variance trade-off in detail.

24. What does a high bias model mean?

25. What does a high variance model mean?

26. What is underfitting? When does it occur?

27. What is overfitting? How do you detect it?

28. What are methods to prevent overfitting?

29. What is a good fit?

30. Draw and explain the graph of bias-variance trade-off.

DATA PREPROCESSING

31. What are the different methods to handle missing values?

32. When should you use mean/median/mode imputation?

33. What are forward fill and backward fill?

34. When is it appropriate to drop missing data?

35. What is MICE (Multiple Imputation)?

36. What are outliers? How do you detect them?

37. What are the methods to handle outliers?

38. How do you remove outliers using the IQR method?

39. Explain the Z-score method for outlier detection.

40. Should outliers always be removed? Discuss.

41. Why is feature scaling important?

42. What is normalization? What is the formula?

43. What is standardization? State the formula.

44. What is the difference between Min-Max scaling and Z-score standardization?

45. When is scaling necessary? Which algorithms require it?

46. Why don't tree-based models need scaling?

47. What is robust scaling?

48. How do you apply feature scaling on train and test data?

49. Why do we convert categorical variables to numerical?

50. What is label encoding? When should you use it?

51. What is one-hot encoding? Give an example.

52. What is the dummy variable trap?

53. When should you use label encoding vs one-hot encoding?

54. What is ordinal encoding?

55. What is target encoding / mean encoding?

FEATURE ENGINEERING & SELECTION

56. What is feature engineering? Explain its importance.

57. How does domain knowledge help in feature engineering?

58. What is feature extraction?

59. What is feature construction? Give an example.

60. What are polynomial features?

61. What are interaction features?

62. What is binning / discretization?

63. When do you use log transformation?

64. How do you extract useful features from date-time data?

65. How do you create features from text data? (TF-IDF, Bag of Words)

66. Why is feature selection important?

67. What are the problems with too many features?

68. What are filter methods? Give an example.

69. What are wrapper methods?

70. What are embedded methods?

71. How do you perform feature selection using a correlation matrix?

72. What is the variance threshold method?

73. What is Recursive Feature Elimination (RFE)?

74. How do you determine feature importance from tree-based models?

75. What is the difference between dimensionality reduction and feature selection?

MODEL EVALUATION

76. What is a confusion matrix? Explain with an example.

77. What are True Positive, True Negative, False Positive, and False Negative?

78. What is accuracy? State the formula.

79. When can accuracy be misleading?

80. What is precision? Explain the formula and intuition.

81. What is recall (Sensitivity)?

82. What is the trade-off between precision and recall?

83. What is F1-score? When is it used?

84. What is F-beta score?

85. What is specificity?

86. What is the ROC curve? How do you plot it?

87. What is AUC (Area Under Curve)? How do you interpret it?

88. What is the perfect score for ROC-AUC?

89. What is the Precision-Recall curve?

90. What is the best metric for imbalanced datasets?

91. What is the difference between macro average and micro average (multi-class)?

92. What is log loss / cross-entropy loss?

93. What is Cohen's Kappa score?

94. What is MAE (Mean Absolute Error)?

95. What is MSE (Mean Squared Error)?

96. What is RMSE (Root Mean Squared Error)?

97. What is the difference between MAE and RMSE?

98. What is R-squared (R^2)? How do you interpret it?

99. What is adjusted R-squared? How is it different from R^2?

100. What is MAPE (Mean Absolute Percentage Error)?

CROSS-VALIDATION

101. Why do we perform train-test split?

102. What is a typical train-test split ratio?

103. What is the purpose of a validation set?

104. What is holdout validation?

105. What is K-Fold cross-validation? Explain in detail.

106. What is the typical value of K in K-fold?

107. What is Stratified K-Fold? When should you use it?

108. What is Leave-One-Out Cross-Validation (LOOCV)?

109. How do you perform cross-validation for time series data?

110. What is the advantage of cross-validation over simple train-test split?

LINEAR REGRESSION

111. What is linear regression? What is the equation?

112. What is the difference between simple linear regression and multiple linear regression?

113. What are dependent variable and independent variables?

114. What are the assumptions of linear regression?

115. What is the linearity assumption?

116. What is homoscedasticity?

117. What is independence of errors?

118. What is normality of residuals?

119. Explain the no multicollinearity assumption.

120. What are residuals? How do you analyse them?

121. What is the Ordinary Least Squares (OLS) method?

122. What is the cost function / loss function in linear regression?

123. What is gradient descent? How does it work?

124. What is learning rate? What happens if it's too high or too low?

125. How do you interpret coefficients?

126. What does the intercept mean?

127. What does the p-value of coefficients tell you?

128. How do you detect multicollinearity? What is VIF?

129. How do you solve the multicollinearity problem?

130. When should you not use linear regression?

LOGISTIC REGRESSION

131. What is logistic regression? Why is it used for classification?

132. What is the sigmoid function / logistic function?

133. Draw the equation and graph of the sigmoid function.

134. What is the output of logistic regression? (Probability)

135. What is the decision boundary?

136. What are log odds?

137. How is logistic regression related to linear regression?

138. What are the assumptions of logistic regression?

139. How do you set the threshold in binary classification?

140. How do you use logistic regression for multi-class classification?

141. What is the One-vs-Rest (OvR) strategy?

142. What is the One-vs-One (OvO) strategy?

143. What is the SoftMax function? For multi-class.

144. What is the cost function of logistic regression? (Log loss)

145. When does logistic regression fail?

REGULARIZATION

146. What is regularization? Why do we use it?

147. How does regularization control overfitting?

148. What is L1 regularization (Lasso)?

149. What is L2 regularization (Ridge)?

150. What is the difference between Lasso and Ridge?

151. How does Lasso perform feature selection?

152. How does Ridge regression shrink coefficients?

153. What is Elastic Net?

154. When do you use Elastic Net?

155. What is the lambda (λ) / alpha (α) parameter in regularization?

156. What happens if lambda value is very high?

157. What does lambda = 0 mean?

158. How do you tune the regularization parameter?

159. When is regularized regression better than normal regression?

160. What is the relationship between regularization and feature selection?

DECISION TREES

161. What is a decision tree? What is its structure?

162. What is root node, internal nodes, and leaf nodes?

163. How does a decision tree work for both classification and regression?

164. What is splitting? How do you choose the best split?

165. What is Information Gain?

166. What is entropy? State the formula.

167. How do you calculate entropy?

168. What is Gini Impurity?

169. What is the difference between Gini Index and Entropy?

170. What is the splitting criterion in a classification tree?

171. What is the splitting criterion in a regression tree? (Variance reduction)

172. What is tree depth?

173. What is pruning? Why do we do it?

174. What is the difference between pre-pruning and post-pruning?

175. What is the role of the max_depth hyperparameter?

176. What is min_samples_split?

177. Explain the min_samples_leaf parameter.

178. What are the advantages of decision trees?

179. What are the disadvantages of decision trees?

180. Why are decision trees prone to overfitting?

181. Can decision trees capture non-linear relationships?

182. Do decision trees need feature scaling?

183. Are decision trees affected by outliers?

184. Is it easy or difficult to interpret decision trees?

185. How do you extract feature importance from a decision tree?

ENSEMBLE METHODS

186. What is ensemble learning?

187. Why are ensemble methods better than single models?

188. What is bagging? What is the full form?

189. What is boosting?

190. What is the main difference between bagging and boosting?

191. What is stacking?

192. What is a voting classifier? (Hard voting vs Soft voting)

193. What is Random Forest? How does it work?

194. What is "random" in Random Forest?

195. What is bootstrap sampling in Random Forest?

196. What is feature randomness in Random Forest?

197. What are the advantages of Random Forest?

198. How does Random Forest handle overfitting?

199. What does the n_estimators hyperparameter mean in Random Forest?

200. Random Forest vs single Decision Tree - which is better and why?

BOOSTING

201. What is boosting in machine learning?

202. How is boosting different from bagging?

203. What is a weak learner?

204. What is a strong learner?

205. Explain the general boosting algorithm workflow.

206. Why is boosting called "sequential"?

207. What is the difference between boosting and stacking?

208. What types of problems can boosting solve?

209. Why does boosting reduce bias?

210. Can boosting overfit? Why?

211. What are the main advantages of boosting?

212. What are the main disadvantages of boosting?

213. When should you use boosting over Random Forest?

214. When should you NOT use boosting?

215. How does boosting handle imbalanced datasets?

216. What is the loss function in boosting?

217. What are residuals in boosting?

218. Explain the concept of learning rate in boosting.

219. What is the relationship between number of estimators and learning rate?

220. What is shrinkage in boosting?

221. What is early stopping in boosting?

222. How do you choose the number of boosting iterations?

223. What is the base learner in boosting algorithms?

224. How deep should the trees be in boosting?

225. What is the typical boosting workflow?

226. How does AdaBoost calculate the weak learner weight (alpha)?

227. What loss function does AdaBoost use?

228. Why does AdaBoost focus on misclassified samples?

229. What are decision stumps in AdaBoost?

230. Can AdaBoost use other base learners besides trees?

231. How does AdaBoost handle multi-class classification?

232. What is SAMME in AdaBoost?

233. What are the key hyperparameters in AdaBoost?

234. What is the difference between SAMME and SAMME.R?

235. How do you tune AdaBoost hyperparameters?

236. What is the optimal learning rate for AdaBoost?

237. How many estimators should you use in AdaBoost?

238. Is AdaBoost sensitive to outliers?

239. How does AdaBoost handle missing values?

240. Can AdaBoost be parallelized?

241. What is the computational complexity of AdaBoost?

242. How do you prevent overfitting in AdaBoost?

243. What is Gradient Boosting?

244. How is Gradient Boosting different from AdaBoost?

245. Explain the Gradient Boosting algorithm.

246. What are pseudo-residuals in Gradient Boosting?

247. Why is it called "Gradient" Boosting?

248. What loss functions can Gradient Boosting use?

249. How does Gradient Boosting handle regression?

250. How does Gradient Boosting handle classification?

251. What is the learning rate (shrinkage) in Gradient Boosting?

252. What are the key hyperparameters in Gradient Boosting?

253. What is subsample in Gradient Boosting?

254. What is the difference between max_features and subsample?

255. How do you tune Gradient Boosting hyperparameters?

256. What is early stopping in Gradient Boosting?

257. What is a good starting point for Gradient Boosting hyperparameters?

258. What is Stochastic Gradient Boosting?

259. How does regularization work in Gradient Boosting?

260. What is the trade-off between learning_rate and n_estimators?

261. How do you interpret feature importance in Gradient Boosting?

262. What is the difference between Gradient Boosting and Random Forest?

263. Can Gradient Boosting handle categorical features?

264. How does Gradient Boosting handle missing values?

265. What are the computational requirements of Gradient Boosting?

266. How do you diagnose overfitting in Gradient Boosting?

MODEL EVALUATION METRICS

CLASSIFICATION METRICS

1. What is a confusion matrix? Draw and explain a 2x2 matrix.

2. What is True Positive (TP)? Give a real-world example.

3. What is True Negative (TN)?

4. What is False Positive (FP)? Why is it called Type I error?

5. What is False Negative (FN)? Explain Type II error.

6. How many metrics can you derive from a confusion matrix?

7. How does a confusion matrix look for multi-class classification? (3x3, 4x4)

8. What is accuracy? State the formula.

9. Formula to calculate accuracy: (TP + TN) / (TP + TN + FP + FN)

10. When can accuracy be misleading?

11. Why is accuracy problematic in imbalanced datasets? Give an example.

12. Is 95% accuracy good or bad? Why does context matter?

13. What is balanced accuracy? How is it different from normal accuracy?

14. What is precision? Explain the formula and intuition.

15. Precision formula: TP / (TP + FP)

16. When is precision important? Give a use case (spam detection).

17. What is recall (Sensitivity/True Positive Rate)?

18. Recall formula: TP / (TP + FN)

19. When is recall important? Give a use case (cancer detection).

20. What is the trade-off between precision and recall? Explain in detail.

21. What is F1-score? Why do we use it?

22. F1-score formula: 2 × (Precision × Recall) / (Precision + Recall)

23. Why is F1-score the harmonic mean, not arithmetic mean?

24. What is F-beta score? What is the role of the beta parameter?

25. When should you use F2-score and F0.5-score?

26. What is specificity (True Negative Rate)? State the formula.

27. Specificity formula: TN / (TN + FP)

28. What is the difference between sensitivity and specificity?

29. What is False Positive Rate (FPR)? Formula: FP / (FP + TN)

30. What is False Negative Rate (FNR)?

31. What is the ROC curve? What is the full form (Receiver Operating Characteristic)?

32. How do you plot the ROC curve? What is on the X-axis and Y-axis?

33. What does the diagonal line in the ROC curve mean?

34. How does the ROC curve of a perfect classifier look?

35. What is AUC (Area Under Curve)?

36. What is the range of AUC? (0 to 1)

37. What does AUC = 0.5 mean?

38. What does AUC = 1.0 mean?

39. Is AUC = 0.7 good or bad? How do you interpret it?

40. Is ROC-AUC appropriate for imbalanced datasets?

41. What is the Precision-Recall curve?

42. When is the PR curve better than the ROC curve?

43. For imbalanced datasets, should you use PR curve or ROC curve?

44. What is Average Precision (AP)?

45. How do you calculate the AUC of the PR curve?

46. What is Log Loss / Cross-Entropy Loss?

47. How do you calculate log loss? State the formula.

48. What is the range of log loss? Is lower better or higher?

49. What is Cohen's Kappa score? When is it used?

50. What is Matthews Correlation Coefficient (MCC)?

51. What is MAE (Mean Absolute Error)? State the formula.

52. MAE formula: $(1/n) \times \Sigma|y - \hat{y}|$

53. What is the intuition behind MAE? How do you interpret it?

54. What is MSE (Mean Squared Error)?

55. MSE formula: $(1/n) \times \Sigma(y - \hat{y})^2$

56. What is the main difference between MAE and MSE?

57. Why is MSE sensitive to outliers?

58. When should you use MAE and when MSE?

59. What is RMSE (Root Mean Squared Error)?

60. RMSE formula: $\sqrt{MSE}$

61. What is R-squared (R^2)? Explain the coefficient of determination.

62. R^2 formula: 1 - (SS_res / SS_tot)

63. What is the range of R^2? (-∞ to 1)

64. What does R^2 = 0.8 mean?

65. When can R^2 be negative?

66. What is Adjusted R-squared?

67. Why is Adjusted R^2 better than normal R^2?

68. How does R^2 change when you add features?

69. How does Adjusted R^2 behave when you add irrelevant features?

70. When should you use R^2 and when Adjusted R^2?

71. What is MAPE (Mean Absolute Percentage Error)?

72. MAPE formula: (100/n) × Σ|(y - ŷ)/y|

73. What are the limitations of MAPE? (y = 0 problem)

74. What is RMSLE (Root Mean Squared Logarithmic Error)?

75. When do you use RMSLE? (When percentage errors are important)

MULTI-CLASS & MULTILABEL METRICS

76. How does a multi-class confusion matrix look?

77. What is macro average? Explain the formula.

78. What is micro average?

79. What is weighted average?

80. When should you use macro vs micro vs weighted?

81. What is multi-class log loss?

82. How do you calculate metrics in One-vs-Rest strategy?

83. What is the practical difference between macro F1 and weighted F1?

84. What is multilabel classification? How is it different from multi-class?

85. What is Hamming loss in multilabel classification?

86. What is Jaccard score / Jaccard similarity?

87. What is exact match ratio?

88. What is subset accuracy in multilabel problems?

89. What is Mean Average Precision (MAP)? (For ranking problems)

90. What is NDCG (Normalized Discounted Cumulative Gain)?

CLUSTERING METRICS

91. What is Silhouette score? What is its range? (-1 to 1)

92. How do you interpret the Silhouette score?

93. What is the Davies-Bouldin Index?

94. What is the Calinski-Harabasz Index (Variance Ratio Criterion)?

95. What is Inertia / Within-cluster sum of squares?

External Evaluation Metrics

96. What is Adjusted Rand Index (ARI)?

97. What is Normalized Mutual Information (NMI)?

98. What is the Fowlkes-Mallows Index?

99. What are Homogeneity, Completeness, and V-measure?

100. What is Purity score in clustering?