ML-3

Foundations of Machine Learning

1. **Regression analysis** - Statistical method for modeling relationships between dependent and independent variables.
2. **Linear vs nonlinear regression** - Linear assumes straight-line relationship; nonlinear uses curves.
3. **Simple vs multiple linear regression** - Simple uses 1 predictor; multiple uses 2+ predictors.
4. **Regression evaluation metrics** - MSE, RMSE, R^2 score.
5. **Overfitting in regression** - Model fits training noise, performs poorly on new data.
6. **Logistic regression purpose** - Used for binary classification problems.
7. **Logistic vs linear regression** - Logistic predicts probabilities (0-1); linear predicts continuous values.
8. **Odds ratio** - Ratio of success to failure probability.
9. **Sigmoid function** - S-shaped curve mapping values to 0-1 range for probability.
10. **Logistic model evaluation** - Accuracy, precision, recall, ROC-AUC.

Decision Trees and SVM

11. **Decision tree** - Tree-like model making decisions via feature splits.
12. **Decision tree predictions** - Follow branches from root to leaf nodes.
13. **Entropy** - Measures impurity/disorder in data.
14. **Pruning** - Removing unnecessary branches to prevent overfitting.
15. **Missing values in trees** - Uses surrogate splits or imputation.
16. **SVM** - Finds optimal separating hyperplane between classes.
17. **Margin** - Distance between hyperplane and nearest points.
18. **Support vectors** - Critical data points defining the margin.
19. **Non-linear SVM** - Uses kernel tricks to handle complex data.
20. **SVM advantages** - Effective in high dimensions, memory efficient.

Naïve Bayes

21. **Naïve Bayes** - Probabilistic classifier using Bayes' theorem.
22. **"Naïve" assumption** - Features are conditionally independent.
23. **Handling features** - Different distributions for continuous vs categorical.
24. **Prior/posterior probabilities** - Prior is initial belief; posterior is updated belief.

25. **Laplace smoothing** - Prevents zero probabilities for unseen categories.

Advanced Concepts

26. **Naïve Bayes for regression** - No, only for classification.
27. **Missing values in NB** - Typically ignores missing features during prediction.
28. **NB applications** - Spam detection, sentiment analysis.
29. **Feature independence** - Core assumption that simplifies calculations.
30. **Many categories in NB** - Can cause sparse data issues.

Model Evaluation and Optimization

31. **Curse of dimensionality** - Performance degrades as features increase.
32. **Bias-variance tradeoff** - Balance underfitting vs overfitting.
33. **Cross-validation** - Technique to evaluate model generalizability.
34. **Parametric vs non-parametric** - Parametric has fixed parameters; non-parametric grows with data.
35. **Feature scaling** - Normalizing features to common scale.

Ensemble Methods

37. **Ensemble learning** - Combines multiple models (e.g., Random Forest).
38. **Bagging vs boosting** - Bagging trains parallel; boosting sequential.
39. **Generative vs discriminative** - Generative models data distribution; discriminative finds boundaries.

Optimization Techniques

40. **Batch vs SGD** - Batch uses all data; SGD uses samples.
41. **KNN** - Classifies based on nearest neighbors.
42. **KNN disadvantages** - Computationally expensive, sensitive to scale.
43. **One-hot encoding** - Converts categories to binary vectors.

Feature Engineering

44. **Feature selection** - Choosing most relevant features.
45. **Cross-entropy loss** - Measures performance in classification.
46. **Batch vs online learning** - Batch trains on all data; online updates incrementally.

Hyperparameter Tuning

47. **Grid search** - Exhaustive search over parameter combinations.
48. **Decision tree pros/cons** - Interpretable but prone to overfitting.
49. **L1 vs L2 regularization** - L1 creates sparsity; L2 shrinks coefficients.

Preprocessing

50. **Preprocessing techniques** - Scaling, encoding, imputation.
51. **Parametric examples** - Linear Regression; Non-parametric: KNN.
52. **Bias-variance** - Simple models have high bias; complex have high variance.

Ensemble Methods (Cont.)

53. **Random Forest pros/cons** - Reduces overfitting but less interpretable.
54. **Bagging vs boosting** - Bagging reduces variance; boosting reduces bias.
55. **Hyperparameter tuning** - Optimizes model performance.

Regularization

56. **Regularization vs feature selection** - Both reduce overfitting but differently.
57. **Lasso vs Ridge** - Lasso does feature selection; Ridge shrinks coefficients.

Evaluation (Cont.)

58. **Cross-validation purpose** - More reliable performance estimate.
59. **Regression metrics** - MSE, MAE, R^2.
60. **KNN predictions** - Majority vote of nearest neighbors.

Naïve Bayes (Cont.)

66. **NB with continuous features** - Uses Gaussian distribution.
67. **NB assumptions** - Feature independence, equally important features.
68. **Missing values in NB** - Omits missing features in calculations.
69. **NB applications** - Text classification, medical diagnosis.
70. **Generative vs discriminative** - NB is generative; SVM is discriminative.

Decision Boundaries

71. **NB decision boundary** - Can be linear or nonlinear depending on data.
72. **Multinomial vs Gaussian NB** - Multinomial for counts; Gaussian for continuous.
73. **Numerical stability** - Uses log probabilities to avoid underflow.

Smoothing Techniques

74. **Laplacian correction** - Type of smoothing for rare categories.
75. **NB for regression** - Not suitable, classification only.
76. **Conditional independence** - Core assumption of NB.
77. **Many categories** - Can use smoothing or feature hashing.
78. **NB drawbacks** - Strong independence assumption often violated.

79. **Smoothing purpose** - Handles unseen categories.
80. **Imbalanced data in NB** - Can perform poorly on minority class.