Machine Learning Assignment 2 Answers

This notebook contains comprehensive answers to the questions from the ml_assignment-2.pdf.

1. What is regression analysis?

Regression analysis is a statistical method to model and analyze the relationship between a dependent variable and one or more independent variables.

2. Difference between linear and nonlinear regression:

Linear regression models a linear relationship between variables; nonlinear regression models more complex, nonlinear relationships.

3. Difference between simple linear regression and multiple linear regression:

Simple linear regression uses one independent variable; multiple linear regression uses two or more independent variables.

4. How is the performance of a regression model typically evaluated?

Using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared.

5. What is overfitting in regression models?

When the model fits the training data too closely, capturing noise and failing to generalize.

6. What is logistic regression used for?

For binary classification problems, predicting probabilities of class membership.

7. How does logistic regression differ from linear regression?

Logistic regression predicts probabilities using a sigmoid function; linear regression predicts continuous values.

8. Explain the concept of odds ratio in logistic regression:

Odds ratio measures the change in odds of the outcome for a one-unit change in the predictor.

9. What is the sigmoid function in logistic regression?

A function that maps any real-valued number into the (0,1) interval, representing probability.

10. How is the performance of a logistic regression model evaluated?

Using metrics like accuracy, precision, recall, F1-score, ROC-AUC.

11. What is a decision tree?

A tree-like model used for classification and regression that splits data based on feature values.

12. How does a decision tree make predictions?

By traversing from root to leaf nodes based on feature conditions.

13. What is entropy in decision trees?

A measure of impurity or disorder used to decide splits.

14. What is pruning in decision trees?

Removing parts of the tree to reduce overfitting.

15. How do decision trees handle missing values?

By surrogate splits or assigning the most common value.

16. What is a support vector machine (SVM)?

A classifier that finds the hyperplane maximizing the margin between classes.

17. Explain the concept of margin in SVM:

The distance between the separating hyperplane and the nearest data points.

18. What are support vectors in SVM?

Data points closest to the hyperplane that influence its position.

19. How does SVM handle non-linearly separable data?

Using kernel functions to map data into higher dimensions.

20. Advantages of SVM over other classifiers:

Effective in high-dimensional spaces, robust to overfitting.

21. What is the Naïve Bayes algorithm?

A probabilistic classifier based on Bayes theorem with independence assumptions.

22. Why is it called "Naïve" Bayes?

Because it assumes feature independence, which is often not true.

23. How does Naïve Bayes handle continuous and categorical features?

Continuous features are modeled with distributions (e.g., Gaussian); categorical features with frequency counts.

24. Explain prior and posterior probabilities in Naïve Bayes:

Prior is initial belief about class; posterior is updated belief after seeing data.

25. What is Laplace smoothing and why is it used?

A technique to handle zero probabilities by adding a small value.

26. Can Naïve Bayes be used for regression tasks?

No, it is primarily for classification.

27. How do you handle missing values in Naïve Bayes?

By ignoring missing features or imputing values.

28. Common applications of Naïve Bayes:

Spam filtering, text classification.

29. Explain feature independence assumption in Naïve Bayes:

Features are assumed to be independent given the class.

30. How does Naïve Bayes handle categorical features with many categories?

By grouping rare categories or using smoothing.

31. What is the curse of dimensionality?

High-dimensional data causes sparsity and degrades model performance.

32. Explain bias-variance tradeoff:

Balance between underfitting (high bias) and overfitting (high variance).

33. What is cross-validation and why is it used?

A technique to evaluate model generalization by partitioning data.

34. Difference between parametric and non-parametric algorithms:

Parametric assume fixed form (e.g., linear regression); non-parametric do not (e.g., KNN).

35. What is feature scaling and why important?

Rescaling features to a common scale to improve model performance.

36. What is regularization and why used?

Technique to prevent overfitting by adding penalty terms.

37. Explain ensemble learning and give example:

Combining multiple models to improve performance; e.g., Random Forest.

38. Difference between bagging and boosting:

Bagging builds models independently; boosting builds sequentially focusing on errors.

39. Difference between generative and discriminative models:

Generative models model joint distribution; discriminative model conditional distribution.

40. Explain batch gradient descent and stochastic gradient descent:

Batch uses all data per update; stochastic uses one sample per update.

41. What is K-nearest neighbors (KNN) and how it works:

Classifies based on majority class among k nearest neighbors.

42. Disadvantages of KNN:

Computationally expensive, sensitive to irrelevant features.

43. Explain one-hot encoding and its use:

Converts categorical variables into binary vectors.

44. What is feature selection and why important:

Selecting relevant features to improve model accuracy and reduce complexity.

45. Explain cross-entropy loss and its use:

Loss function for classification measuring difference between predicted and true labels.

46. Difference between batch learning and online learning:

Batch learns from entire dataset; online learns incrementally.

47. Explain grid search and its use:

Systematic hyperparameter tuning method.

48. Advantages and disadvantages of decision trees:

Easy to interpret but prone to overfitting.

49. Difference between L1 and L2 regularization:

L1 promotes sparsity; L2 penalizes large weights.

50. Common preprocessing techniques:

Normalization, encoding, missing value imputation.

51. Difference between parametric and non-parametric algorithms with examples:

Parametric: Linear regression; Non-parametric: KNN.

52. Bias-variance tradeoff and model complexity:

Complex models have low bias, high variance; simple models vice versa.

53. Advantages and disadvantages of ensemble methods like random forests:

Improved accuracy but less interpretable.

54. Difference between bagging and boosting:

Bagging reduces variance; boosting reduces bias.

55. Purpose of hyperparameter tuning:

Optimize model performance.

56. Difference between regularization and feature selection:

Regularization penalizes complexity; feature selection removes features.

57. How Lasso (L1) differs from Ridge (L2) regularization:

Lasso can zero out coefficients; Ridge shrinks coefficients.

58. Explain cross-validation and why used:

Repeated model evaluation to ensure generalization.

59. Common evaluation metrics for regression:

MSE, RMSE, MAE, R-squared.

60. How KNN makes predictions:

By majority vote of nearest neighbors.

61. Curse of dimensionality and its effect:

High dimensions cause data sparsity, reducing model effectiveness.

62. Importance of feature scaling:

Ensures features contribute equally.

63. How Naïve Bayes handles categorical features:

Uses frequency counts and smoothing.

64. Explain prior and posterior probabilities in Naïve Bayes:

Prior is initial class probability; posterior updated after data.

65. What is Laplace smoothing and why used:

Avoid zero probabilities in categorical data.

66. Can Naïve Bayes handle continuous features:

Yes, using Gaussian distribution.

67. Assumptions of Naïve Bayes:

Feature independence.

68. How Naïve Bayes handles missing values:

Ignores or imputes missing features.

69. Common applications of Naïve Bayes:

Text classification, spam detection.

70. Difference between generative and discriminative models:

Generative models model data distribution; discriminative models model decision boundary.

71. Decision boundary of Naïve Bayes for binary classification:

Typically linear or quadratic depending on distribution.

72. Difference between multinomial and Gaussian Naïve Bayes:

Multinomial for discrete counts; Gaussian for continuous data.

73. How Naïve Bayes handles numerical instability:

Using log probabilities.

74. What is Laplacian correction and when used:

Same as Laplace smoothing, used to handle zero counts.

75. Can Naïve Bayes be used for regression:

No.

76. Explain conditional independence assumption in Naïve Bayes:

Features are independent given the class.

77. How Naïve Bayes handles categorical features with many categories:

Grouping or smoothing.

78. Drawbacks of Naïve Bayes:

Strong independence assumption, poor with correlated features.

79. Explain smoothing in Naïve Bayes:

Technique to handle zero probabilities.

80. How Naïve Bayes handles imbalanced datasets:

Using class priors or resampling.