

Section 1: Supervised Learning – Solved Questions

Q1. What assumption does Naive Bayes make about features?

Answer: Features are independent of each other.

Explanation: Naive Bayes applies the conditional independence assumption. This is why it works well in text classification but can fail when features are correlated.

Q2. Which of the following is true about Logistic Regression for classification?

Answer: It uses the sigmoid function to output probabilities between 0 and 1.

Explanation: Logistic regression models the log-odds of the probability. It is suitable for binary classification and can be extended to multi-class using 'OvR' or 'softmax'.

Q3. Decision Trees are prone to which problem?

Answer: Low bias, High variance.

Explanation: Trees can perfectly fit training data, leading to overfitting. Bagging/Random Forest reduces variance by combining multiple trees.

Q4. Which metrics are used in Decision Trees to measure split quality?

Answer: Entropy, Gini Impurity.

Explanation: Entropy measures information gain, Gini measures impurity. Both are standard for classification trees.

Q5. Hard margin SVM works best when:

Answer: Data is perfectly linearly separable.

Explanation: Hard margin SVM requires no misclassifications. It fails if data has noise or outliers → soft margin is used instead.

Q6. Increasing the number of neurons in hidden layers will always lead to better model performance. True/False?

Answer: False. Optimal neurons require experimentation.

Explanation: Too many neurons cause overfitting and long training. Too few lead to underfitting. Hyperparameter tuning is necessary.

Q7. GaussianNB performance suffers when features are dependent. Why?

Answer: Because it assumes conditional independence.

Explanation: Naive Bayes calculates probabilities as product of individual feature likelihoods. Dependency breaks this assumption and reduces accuracy.