ASSIGNMENT-2

1. The logistic function (sigmoid) converts the output of a linear model to probabilities between 0 and 1. It is used in logistic regression to predict the probability of the target class.

2. Decision trees commonly use information gain or gini impurity to determine the best split. These metrics measure how well a split separates the classes.

3. Entropy measures the disorder/impurity in a dataset. Information gain measures the decrease in entropy after a split, indicating more homogeneous subsets.

4. Random forests use bagging to train each tree on a random subset of data. They also randomize the features considered for splits to decorrelate the trees. This improves overall accuracy.

5. KNN classification commonly uses Euclidean distance. A smaller k-value weights distance more heavily, making the algorithm more flexible but prone to overfitting.

6. The Naive Bayes assumption of independent features simplifies probability calculations but is often not true in real data. This can negatively impact its classification accuracy.

7. Kernel functions in SVMs transform data into a higher dimensional space to find an optimal separating hyperplane. Common kernels include polynomial and radial basis function.

8. High model complexity can lead to low bias but high variance (overfitting). Simpler models have higher bias but lower variance. Tradeoff must be optimized.

9. TensorFlow provides tools to build neural network models, define loss functions, train using gradient descent, and deploy for inference.

10. Cross-validation evaluates models on held-out subsets of data to estimate generalization performance. It prevents issues with overfitting to the training set.

11. Overfitting can be reduced via regularization, early stopping, reducing model complexity, augmentation, dropout, etc.

12. Regularization adds a penalty term to the loss function that discourages complex models. This improves generalization and prevents overfitting.

13. Hyperparameters are manually set parameters that control model complexity. Tuning them optimizes model performance by balancing under/overfitting.

14. Precision measures positive class accuracy. Recall measures positive class coverage. Accuracy can be misleading if classes are imbalanced.

15. The ROC curve plots true positive rate vs false positive rate. The area under the curve measures how well a classifier distinguishes positive/negative cases.

Submitted by

CHIMMA DIVYA NAGASAI

21HU5A4201

Chebrolu Engineering College