

#### Answer 1

When the depth is 2 and 3, the recall is the highest. This is because as the tree depth increases, the model can more finely partition the data, leading to better performance. When the depth is 1, the precision is the lowest. This is because, with a very small maximum depth, the decision tree model cannot effectively distinguish and classify the data. Only when the depth increases does the precision improve.

The F1 score reaches its highest value when the depth is 2 and 3.

Micro-average: This aggregates the precision, recall, and F1 score across all class predictions, calculating a global score. It is suitable for imbalanced datasets because each sample is weighted equally.

Macro-average: This calculates the average precision, recall, and F1 score for each class individually. It is suitable for balanced datasets because each class is weighted equally.

Weighted-average: This calculates the average of precision, recall, and F1 score for each class, weighted by the proportion of samples in each class. It is suitable for imbalanced datasets because the weights account for the differences in sample sizes.

#### Answer 2

According to my code, the entropy value is 0.8930108100446013, the Gini index value is 0.4276929180718112, and the classification error value is 0.3098591549295775. Information gain typically refers to the difference between the entropy of the parent node and the weighted sum of the entropy of the child nodes. The first split selects the feature perimeter3, and the decision boundary value is 115.35.

#### Answer 3

The numerical values required by the question are all reflected in the code and are not repeated here.

According to the data calculated in the code, using continuous data is indeed beneficial for the model. This is because, after performing PCA dimensionality reduction, using only the first two principal components for model training results in performance metrics (e.g., F1 score, precision, and recall) that are nearly equivalent to those of the original data. This indicates that continuous data retains a significant amount of feature information, even after dimensionality reduction. Additionally, using continuous data results in higher F1 scores, precision, and recall, indicating that the model performs well in distinguishing between classes.