Classification

January 14, 2025

[ ]:
```
'''
   Classification -->

   Classification is a supervised machine learning task where the goal is to␣
↪predict
   a label or category for a given input based on learned patterns from␣
↪labeled data.
   The inputs can be text, images, audio, or any other data, and the output is␣
↪typically
   one of a finite set of predefined classes.
'''
```

[ ]:
```
'''
   Key Steps in Classification -->

   Data Collection :

   Gather labeled data relevant to the problem.

   Data Preprocessing :

   Handle missing values.
   Normalize or standardize numerical data.
   Encode categorical data into numerical formats (e.g., one-hot encoding).

   Feature Extraction :

   Select or extract meaningful features from the data
   (e.g., pixel values for images, TF-IDF for text).

   Model Selection :

   Choose an algorithm like Logistic Regression, SVM, Decision Trees, Random␣
↪Forests,
   or Neural Networks depending on the problem.
```

```
    Training : Train the model on the labeled dataset using a loss function to␣
 ↪minimize errors.
    Validation : Evaluate the model on a validation dataset to tune␣
 ↪hyperparameters and avoid overfitting.
    Testing : Test the final model on unseen data to measure its performance.
'''
```

[ ]:
```
'''
    Common Types of Classification -->

    Binary Classification: Predicting one of two possible outcomes (e.g., spam␣
 ↪vs. not spam).
    Multi-Class Classification: Predicting one of three or more possible␣
 ↪outcomes
    (e.g., classifying images into "cat," "dog," or "bird").
    Multi-Label Classification: Assigning multiple labels to each instance
    (e.g., tagging images with "dog," "outdoor," and "sunny").
'''
```

[ ]:
```
'''
    Algorithms for Classification -->

    Linear Models :

    Logistic Regression
    Linear Discriminant Analysis (LDA)

    Tree-Based Models :

    Decision Trees
    Random Forest
    Gradient Boosted Trees (e.g., XGBoost, LightGBM)

    Instance-Based :
    k-Nearest Neighbors (k-NN)

    Neural Networks :

    Multi-Layer Perceptrons (MLPs)
    Convolutional Neural Networks (CNNs) for image classification
    Recurrent Neural Networks (RNNs) for sequential data
    Support Vector Machines (SVMs)
'''
```

[ ]:
```
'''
    Evaluation Metrics -->

```

```
    Accuracy: Overall correctness.
    Precision: Correct positive predictions relative to total predicted␣
 ↪positives.
    Recall (Sensitivity): Correct positive predictions relative to actual␣
 ↪positives.
    F1-Score: Harmonic mean of precision and recall.
    Confusion Matrix: A detailed breakdown of predictions.
'''
```

[ ]:
```
'''
    Applications of Classification -->

    Email Spam Detection
    Image Recognition (e.g., facial recognition)
    Medical Diagnosis (e.g., cancer classification)
    Sentiment Analysis
    Fraud Detection
'''
```