CYBR 486 - Lab #5: Perceptron

Overview

This lab focuses on building, training, and evaluating a Perceptron binary classifier using the breast cancer dataset from scikit-learn. The objective is to classify tissue samples as either malignant or benign using various features. You will work through loading and preprocessing the dataset, splitting it into training and testing sets, training the Perceptron model, and evaluating its performance using key evaluation metrics.

Objectives

1. Load and preprocess the breast cancer dataset.
2. Split the dataset into training (80%) and testing (20%) subsets.
3. Build and train a Perceptron binary classifier using the training set.
4. Evaluate the model's performance with:
   * Confusion Matrix
   * Accuracy Score
   * Precision Score
   * Recall Score
5. Visualize the results using a confusion matrix heatmap.

Prerequisites

1. Python 3.x installed on your machine.
2. Required Python libraries:
   * scikit-learn
   * pandas
   * numpy
   * seaborn
   * matplotlib

To install the required libraries, run the following command:

bash

Copy code

pip install scikit-learn pandas numpy seaborn matplotlib

Dataset

* Dataset: Breast cancer dataset
* Source: Built-in dataset from scikit-learn
* Description: The dataset contains features of breast cancer tissue samples and a binary classification target variable (0 for malignant, 1 for benign).

Ensure the dataset is automatically loaded through the scikit-learn function load\_breast\_cancer().

Usage Instructions

1. Clone the Repository Clone this repository or download the code file and dataset to your local machine.
2. Run the Script Execute the Jupyter notebook or Python script (e.g., perceptron\_lab.ipynb) to load the dataset, split it into training and testing sets, train the model, and evaluate it.
3. Outputs
   * Information about the dataset, including data types and null values.
   * Model evaluation metrics: confusion matrix, accuracy, precision, and recall scores.
   * Visualization: confusion matrix heatmap.
4. Save Results (Optional) You can save the results of the model's evaluation to a CSV file or image if needed.

Code Structure

1. Step 1: Import Necessary Libraries
   * Import libraries for model training, dataset loading, and evaluation metrics.
2. Step 2: Load and Inspect the Dataset
   * Load the breast cancer dataset and inspect its structure.
   * Check for null values and data types.
3. Step 3: Split the Data into Training and Testing Subsets
   * Split the data into 80% training and 20% testing sets using train\_test\_split().
4. Step 4: Build and Train the Perceptron Model
   * Train the Perceptron model using the training set.
5. Step 5: Evaluate the Model
   * Make predictions on the test set.
   * Evaluate the model using confusion matrix, accuracy, precision, and recall scores.
6. Step 6: Visualize the Results
   * Display the confusion matrix using a heatmap.

Evaluation Metrics

1. Confusion Matrix: A matrix showing the actual vs predicted classifications.
2. Accuracy Score: The percentage of correctly predicted instances.
3. Precision Score: The proportion of true positives among predicted positives.
4. Recall Score: The proportion of true positives among actual positives.

Example Outputs

* Dataset Info: Information on the number of features, data types, and null values.
* Confusion Matrix: A 2x2 matrix displaying the true positives, false positives, true negatives, and false negatives.
* Evaluation Metrics:
  + Accuracy: A score between 0 and 1 representing the percentage of correct predictions.
  + Precision: A score representing how many of the predicted positives were actually positive.
  + Recall: A score showing how many of the actual positives were correctly identified.

Results and Observations

* Evaluate how well the Perceptron model performs on the breast cancer dataset.
* Compare accuracy, precision, and recall scores to understand the balance between false positives and false negatives.
* Visualize the confusion matrix to see how the model classifies malignant and benign samples.