HyperionDev

Supervised Learning – Logistic Regression

Model Answer Approach

Visit our website

Auto-graded task

The objective of this task was to classify iris flowers into two categories: `Iris-setosa` and `Not-Iris-setosa`. The process begins by importing the necessary libraries and loading the Iris dataset from a CSV file into a Pandas DataFrame. To gain an understanding of the dataset, the `.info()` and `.describe()` methods are employed.

A copy of the DataFrame is created to preserve the integrity of the original data. The target variable, `Species`, is then encoded from categorical to numerical values, with 0 representing `Iris-setosa` and 1 representing `Not-Iris-setosa`. This approach combines `Iris-versicolor` and `Iris-virginica` into the `Not-Iris-setosa` category.

Following the encoding step, features (denoted as `x`) and the target variable (denoted as `y`) are selected. The features are scaled to enhance model performance. Subsequently, the dataset is split into training and testing sets, with 30% allocated for testing. A random state is set to ensure reproducibility of the results. The stratification parameter is also applied to maintain the same class distribution in the test set as found in the original dataset.

Next, a logistic regression model is instantiated and fitted to the training data. Predictions are then made on the testing set. A confusion matrix is then generated to evaluate the model's performance. The results indicate that the model successfully distinguishes between `Iris-setosa` and `Not-Iris-setosa`. Specifically, it accurately predicted 15 instances of `Iris-setosa` and 30 instances of `Not-Iris-setosa`, with no false positives or false negatives.

Lastly, code is implemented to calculate accuracy, precision, and recall using the relevant formulas. These values are further validated using the scikit-learn library's built-in methods: `accuracy_score`, `precision_score`, and `recall_score`, ensuring the reliability of the performance metrics obtained.

Optional task

For the optional task, the same approach is applied as described above, with the primary difference being the inclusion of all three categories: `Iris-setosa`, `Iris-versicolor`, and `Iris-virginica`, corresponding to the numeric values 0, 1, and 2, respectively.