

Assignments

Read the lecture notes to answer the following.

1. Logistic Regression and Binary Classification:
 - a. What is Binary classification?
 - b. Explain how logistic regression uses features of a dataset to perform binary classification using the Sigmoid Function.
2. Cross-entropy Cost Function:
 - a. What is the expression of cross-entropy cost function? Define each term.
 - b. What is the expression of the loss function associated with the cross-entropy cost function?
 - c. Plot the loss as a function of the weighted sum of the features.
 - d. Consider the following example. There is a dataset of emails which must be classified as spam (class: 0) or non-spam (class: 1). A logistic regression model classifies an email as a spam email. However, it is a non-spam email based on the data. Can you expect a lower loss or a higher loss? Explain your answer using the above plots.
3. A binary classifier predicts the classes as shown in the table. Identify them as true positive (TP), true negative (TN), false positive (FP) or false negative (FN). Assume that the label Car is the positive label.

Actual Label	Predicted Label	TP, TN, FP, or FN?
Car	Van	
Car	Car	
Van	Car	
Van	Van	

4. Assume that we have the following confusion matrix.

	Actual Positive	Actual Negative
Predicted Positive	80	8
Predicted Negative	6	75

- What are the numbers of TP, TN, FP, and FN?
 - What is the accuracy of the classifier?
 - What is the precision of the classifier?
 - What is the recall of the classifier?
 - What is the F1 score of the classifier?
5. Assume that we have a dataset of stars with the following classes.
- White dwarf
 - Brown dwarf
 - Red dwarf
 - Blue Giant
 - Red Giant
- How many binary classifiers are required to develop a logistic regression model for the above dataset using one-vs-rest method?
 - How many binary classifiers are required to develop a logistic regression model for the above dataset using one-vs-one method?
 - Explain how the one-vs-rest method works for the above dataset?
 - Explain how the one-vs-one method works for the above dataset?
6. Develop a logistic regression model for the Iris-dataset available in the Scikit-learn Python package. Try with different options and values for penalty, C and solver of the LogisticRegression() class and find the optimum hyperparameters and values. Evaluate your best model using a confusion matrix and a classification report. Comment on your results.

References:

1. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
2. https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html