Homework 2

March 2022

1 Multi-class classification

Solve a simple classification problem by four different methods. The Glass Identification dataset (glass.txt) has 11 columns where the first column is the ID of the sample (needs to be ignored), the 2nd to 10th columns are features. The last column is the type of glass. The Glass Identification dataset includes six types of glasses.

In general, there are two types of classification algorithms: binary classification algorithms and multi-class classification algorithms. Binary classification is the case where we have to classify objects into two groups. Generally, these two groups consist of "positive" and "negative" samples, respectively. On the other hand, in multi-class classification, we have to classify objects into more than two classes. In this case, given a set of attributes of glasses, a multi-class classification task would be to determine the type of glasses.

Specifically, the dataset has 214 samples. You can use 90% samples as the training set, and 10% samples as the test set.

Here are four methods you need to implement:

(1) Using logistic regression models to solve the multi-class classification problem. A multi-class classification problem can be split into many different binary classification problems. To understand how this works, let us consider an example. A classification problem is to classify various glasses into six types: float processed building windows, not float processed build-

ing windows, vehicle windows, containers, tableware, and headlamps. Now, this is a multi-class classification problem. You can use logistic regression models to solve the multi-class classification problem as follows.

The above multi-class classification problem can be decomposed into six binary classification problems:
Problem 1 : Float processed building windows vs [Not float processed building windows, Vehicle windows, Containers, Tableware, Headlamps]
Problem 2 : Not float processed building windows vs [Float processed building windows, Vehicle windows, Containers, Tableware, Headlamps]
...
Problem 6 : Headlamps vs [Float processed building windows, Not float processed building windows, Vehicle windows, Containers, Tableware]

Making decisions means applying all classifiers to an unseen sample $\mathbf{X_i}$ and predicting the label for which the corresponding classifier reports the highest confidence score. The logistic regression model ($y_i \in \{0, 1\}$) for binary classification is on page 4 of lecture 1.

(2) Find a linear model with a weight matrix W to solve the multi-class classification problem. For an input sample $\mathbf{X_i} \in \mathbf{R}^9$, we have $y_i = \mathbf{W}^T X_i \in \mathbf{R}^6$. Then the predicted class $c_j = \arg \max_{j \in 1,\ldots,6} p_j$, where $p = softmax(y_i) \in \mathbf{R}^6$. To this end, you need to compute the cross entropy loss and use gradient descent to find a weight matrix W.

(3) Using Linear Discriminant Analysis (LDA) models to solve the multi-class classification problem. Similarly, the multi-class classification problem can be decomposed into many different binary classification problems. In this way, you can use an LDA model to solve a binary classification, which is on page 26 of lecture 1.

(4) Using logistic regression models trained by logistic loss to solve the multi-class classification problem. Similarly, the multi-class classification problem can be decomposed into many different binary classification problems. The logistic loss ($y_i \in \{+1, -1\}$) for binary classification is on page 12 of lecture 1.

Please provide the following results for four methods:

(1) Use Principal Component Analysis (PCA) to project the weight matrix $W \in \mathbf{R}^{9\times 6}$ onto lower-dimensional space while preserving as much of the variance in the weight matrix as possible. In other words, $W' = PW$, where $P \in \mathbf{R}^{2\times 9}$ or $P \in \mathbf{R}^{3\times 9}$, and $W' \in \mathbf{R}^{2\times 6}$ or $W' \in \mathbf{R}^{3\times 6}$ (Specifically, 2-dimensional or 3-dimensional can be selected by yourself). In this way, you can draw six columns of the matrix W'. Besides, you also need to project training samples into the lower-dimensional space based on the matrix P, e.g., $\mathbf{X_i}' = PX_i$, where $X_i \in \mathbf{R}^9$, and $\mathbf{X_i}' \in \mathbf{R}^2$ or $\mathbf{X_i}' \in \mathbf{R}^3$. Specifically, projected training samples and six columns of the matrix W' should be visualized together. You can use existing python or Matlab PCA functions.

e.g., the visualization code in python (plot_pca.py)

(2) Please report the classification accuracy of four methods on the test set.

(3) Discuss the difference and similarity of these four methods from problem formulation and loss function. The discussion is an open problem. You are not limited to the two above perspectives.

More information about the dataset can be referred as follows:
https://archive.ics.uci.edu/ml/datasets/Glass+Identification