Bit Decoding SGD Logistic Regression

An interception system observes $K = 48$ two-dimensional features, which are represented by the matrix **Xpred** contained in the file **homeworkclass.mat**. Each feature must be used to reveal an unknown bit, that is, a label belonging to the set $\{0, 1\}$.

The sequence of K bits corresponds to 6 ASCII characters. The word composed of these characters is taken from a famous quote linked to a cult movie from the 1980s. To complete the exercise, it is necessary to trace back to the original sentence (for this final task, no methods studied in the course are required, just some intuition and elements of "general culture"...).

To classify the bits from the features, you are required to implement a binary classification algorithm based on logistic regression. In particular, you are asked to implement K distinct classifiers, one for each bit to be classified.

To design the K classifiers, a training set consisting of feature-label pairs is provided:

$$\{(X_i, Y_i)\}_{i=1}^{n}, \qquad \text{where } X_i \in \mathbb{R}^2, \quad Y_i \in \{0, 1\}$$

The features and labels are stored respectively in the matrix **Xtrain** and in the vector **Ytrain**, which are both found in the file **homeworkclass.mat**.

The training set contains K consecutive data tranches, of unknown length, which correspond to the K classification problems. The first data tranche is distributed according to the model corresponding to the first classification problem, i.e., **Xpred(1,:);** the second according to the model corresponding to the second classification problem, i.e., **Xpred(2,:);** and so on.

Train a logistic classifier on the training set. To this end, apply the stochastic gradient algorithm with a constant step-size.

Given the sequence of vectors $\beta(i)$ produced by the stochastic gradient algorithm, select K values from this sequence that can reasonably correspond to the K classification problems considered. To optimize this selection, evaluate the behavior of the algorithm for different step-size choices.