



भारतीय सूचना प्रौद्योगिकी संस्थान गुवाहाटी
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY GUWAHATI

CS 360: Machine Learning Lab
Practice Assignment 4

Instructions: This is only for practice. Complete it by 12:00 PM today. Your completion will be reviewed by the Teaching Assistants.

1. Download [The Iris species dataset](#) regarding the classification of Iris flower species based different given predictors. Write a program to do the followings:

- (a) Read the dataset.
- (b) Randomly split the dataset into training and testing splits using k-fold cross validation, for $k=5$.

Hint: Consider 10% of the training set as validation set.

- (c) Design the hypothesis (\hat{y}) to classify the Iris species using Multi-class Logistic regression to estimation of the model parameters (θ) or (ω) for the different training-testing splits obtained from 1(b). Apply the followings for this:

- Batch gradient descent method for calculating error-gradient.
- One-vs-all strategy for multi-class scenario.

*** without using in-built Python packages /libraries for both.**

- Consider model parameters are to be initialized (θ) or (ω) to zero.
- Apply hyper-parameter tuning for selecting the best model for the following hyper-parameters, considering the given values:

1. learning rate, α : {0.0001, 0.1}
2. ρ : {0.001, 0.01} where, ρ signifies absolute error-difference in two consecutive epochs.
3. epoch, (T): {50, 100}

- (d) Calculate and visualize the confusion matrix, considering the designed hypothesis (\hat{y}) and the actual targets (y), for all the training: testing splits.

- (e) Calculate and report the overall accuracy, class-wise accuracy, precision and recall values, with respect to average of the k-folds and the best fold, in separate sheets, and also write the result analysis.

2. Download [The Wine quality dataset](#) regarding 13 constituents present in wine, without the information of the wine types (target). Write a program to do the followings:

- (a) Read the dataset.
- (b) Cluster the dataset into different categories of wine using K-means clustering method (**without using in-built Python library/ packages**). Consider the following for this:

- Number of initial centroids, K to be 3, 5, 10, and 15.

Hint: Apply K-means clustering, separately for all the given K -values.

- Stopping criteria: No changes in the cluster assignment.
- (c) Report and visualize the obtained final clusters with respect to different K , and write the result analysis.