

Lab - 10

CSL2010: Introduction To Machine Learning AY 2021-22

General Instructions

1. You need to upload a zip **<Your_Roll_No>.zip**, which contains two files for the task in **<Your_Roll_No>.py** format and the report for the entire assignment in **<Your_Roll_No>.pdf** format.
2. Provide your collab file link in the report. **Make sure that your file is accessible.**
3. Submit a single report, mention your observations for all the tasks.[Include plots]
4. Report any resources you have used while attempting the assignment.

Any submission received in another format or after the deadline will not be evaluated.

(Due: 5:30 PM, 27 Oct 2021)

Problem 1: Implement a multi-layer perceptron model with two hidden layers to perform multiclass classification on the iris database using the inbuilt function (with an appropriate activation function if necessary). Use a random split of 70:30 for training and testing respectively.

- (a) Vary the learning rate for a fixed size of hidden layers and show the best learning rate value when you run it for 50 epochs. (5 marks)
- (b) Vary the number of epochs from 10-100 in a step of 10 and show the loss value curve, using the best set of hyperparameters. (5 marks)

Link for the IRIS dataset: <https://archive.ics.uci.edu/ml/datasets/iris>

Problem 2: (Regression Task)

Using the dataset of the given [link](#) or [drive link](#) (House Price prediction dataset)

The dataset consists of 80 columns of unique independent variables, 1460 rows in the train data, and 1459 rows in the test data. The goal is to predict the sales prices of houses based on these variables using MLP (Multilayer perceptron).

1. Preprocess the data and discuss the preprocessing steps in the report. (5 marks)
2. Train an MLP model (with at least one hidden layer, and an appropriate non-linear activation function if necessary) using training data by varying different hyperparameters, and predict the sales price of the houses. (5 marks)
3. For test data, show the prediction values for any 4 samples. (2 marks)
4. Compute average mean-square error for complete test data. (2 marks)
5. Try L1 (manhattan distance) and L2 (euclidean distance) loss and compare the results. (3 marks)
6. Plot the loss curve, and a histogram of prediction errors with an appropriate bin size.

(3 marks)

Reference:

https://github.com/willh99/MLP_Regression/blob/master/MLP/SKLearn_LinReg.ipynb