Department of Electrical and Computer Engineering
University of Massachusetts Lowell

EECE CDM Problem Set #4

1. Consider the problem of multiclass classification for the species identification using the iris.data set. The data can be obtained online from https://archive.ics.edu/ml/machine-learning-databases/iris/iris.data

a. Using the pandas library function pandas.read_csv() read and store species and characteristics. Assign species (Iris-setosa, Iris-versicolor, Iris-virginica) to enumerated classes (0,1,2) respectively. Assign features (sepal-length, sepal-width, petal-length and petal-width) to enumerated class (0,1,2,3) respectively.

b. Draw the histogram of the features of each class. Note the intersectionality of each feature across each species. Using panda scatterplot note that the features petal-length and petal-width exhibit the tight correlation across class. Construct and train the model using 1-K encoding using the two aformentioned features. Solve the resulting linear equation. Compute the fitting error.

c. Compare the fittng error for the model using all of the features.