UCLA Computer Science Department

CS 260 Machine Learning Project

You are asked to implement two machine learning algorithms. The first one is kNN (k should be an input, by default should be 1-NN), and the second one is a Neural Network. The goal is to fully understand the underpinnings of these two algorithms. You can assume that the input is 2-D.

To test your algorithms, you need to create a training set (N data points), a validation set (M data points), and a test set (K data points). You will be asked to have each algorithm output the following measures of performance for each of the three sets: accuracy, precision, recall, sensitivity, specificity, F-measure, ROC curves, confusion matrices, and Mathews Correlation Coefficient (in case the data is imbalanced).

Each data point *i* is represented by two dimensions $x_i = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \in R^2$, and x is an an array of points. For example, your testing set will have N points, and so your input will be $\in R^{2xN}$:

$$Input\_Test\_Set = \begin{bmatrix} f_{11} & \cdots & f_{1N} \\ f_{21} & \cdots & f_{2N} \end{bmatrix}$$

Each data point will be associated with one of two labels (one will be a positive class and the other a negative class). For example: Cancer or No Cancer.

$$Labels\_Test\_Set = [Cancer, No\ Cancer]$$

You will need to test your code using your own data sets, or data sets that you find. We will then provide you with our own training, testing, and validation set to test how well your code works.

Honesty Measure: You are to work independently! You are not to use any sources from the internet (or copy code from somewhere). There will be no collaboration. The goal is for you to completely understand how these two machine learning algorithms work.