

Normalizing Features:

To normalize the features, we are going to rescale each feature so that it is within the range [0,1]. Let \mathbf{X} be the feature that we are working with (remember that this is one of the columns of the matrix that you have loaded from the data file), and x_i be the i^{th} entry in the column. You can use the following formula to rescale/normalize the feature values:

$$x_i^{Norm} = \frac{x - \min(X)}{\max(X) - \min(X)}$$

where, $\min(X)$ and $\max(X)$ are the minimum and maximum values of the feature (column) across all entries, respectively. Only numeric (continuous) features should be normalized, therefore you will perform this data pre-processing step for the Iris, Spambase and Housing datasets. Normalize the entire dataset once; before you perform cross-validation.

NOTE: Do not normalize the output/label variable(s) (last column of the dataset).