GEARS Winter 2022

Perceptron Learning Algorithm

The pdf file, CreditApproval.pdf, has a description of the features of the data set. It is noted the values in the data set was changed by the authors to protect the confidentiality of the data. Use the first 500 data for the training set and the rest for the testing set. That is, the training data set is used to determine the best weights. Once the best weights are determined, you apply them to the testing data set, which you can think as future credit card applications, except that you know the labels so that you can determine how well the best weights that you computed from the training data set classify them by calculating the accuracy. Write a computer code to implement the Perceptron Learning Algorithm for this data set. Set the number of epochs to be 1000. For the output, report the accuracy for the training data as well as the testing data. The accuracy is defined to be the ratio of the number of data that was classified correctly by the algorithm divided by the total number of data. Do this for following cases:

- Data is used unchanged
- Data is normalized using the mean

  $$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_j(i)$$

  and the variance

  $$\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_j(i) - \mu_j)^2$$

  as follows:

  $$x_j(i) = \frac{x_j(i) - \mu_j}{\sigma_j^2},$$

  where $x_j(i)$ is the jth feature of the ith data.
- Plot the accuracy as a function of the epochs for the unnormalized data and normalized data
- What are the most 5 important features that you observed from the Perceptron model?