CS 6301 - Machine Learning Lab - Week 2

Date: 11.08.23

ON THE SPOT

(5) Use the cpu dataset available at

https://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html

(a) Read the data file by executing Table Reader and Missing Value Nodes.

(b) Partition the data into training set (75%) and test set (25%) drawing randomly.

(c) Train a linear regression model on the training set to predict diabetes.

(d) Try selecting and omitting certain attributes for the prediction.

(e) Apply the model to the test set.

(f) Re-run the model with different set of attributes.

(g) Tabulate the performance of the linear regression model with Numeric Scorer Node for every run.