CSCI 5622: Machine Learning Due: by noon on Tuesday Jan 29, 2008

Introduction to Supervised Learning: Homework 1

Professor: Greg Grudic; Substitute Teacher: Sam Reid

For additional instructions and required files for finishing this assignment, please visit the website at: `http://www.colorado.edu/physics/pion/csci5622-spring08/`

Problem 1

What is the difference between a classifier and a classification algorithm?

Problem 2

Why can't the training set be used to perform model assessment?

Problem 3

Implement an algorithm in Matlab® that computes the average $\bar{y}$ of all training data point y-values (regression only), and produces a model that predicts $\bar{y}$ for all its predictions.

Problem 4

Implement the K-Nearest Neighbor algorithm for regression in Matlab®.

Problem 5

Implement the hold-out technique for model assessment in Matlab®, using the mean squared error loss function. The function should have the following signature:

function score=evaluate(algorithm,hyperparameters,trainSet,testSet)

Download main.m and auto-csv.csv from the website. Run main() and report its output as a 5x3 matrix.

Auto-csv.csv was taken from `http://archive.ics.uci.edu/ml/datasets/Auto+MPG`. MPG column was put on the right, and data with missing values were deleted.

Problem 6

What attribute is in the leftmost column in auto-csv.csv?