**Homework Part 4**

**The Performance of Regression and Logistic Regression Models**

INSTRUCTIONS

Please capture all your work in an Orange file (with extension .ows), a Word file and a handful of Excel files. Please copy the images you generate in Orange into the Word file to illustrate your answers as needed.

IMPORTANT NOTE: If you use certain graphs and/or tables to answer a question or draw a conclusion please include the respective graphs / tables. Do so even if including that graph / table is not specifically asked for in the question.

ASSIGNMENT

For this homework assignment, you’ll use the heart\_disease.tab dataset that comes with the Orange installation. To get it, open a File widget and click the “Browse documentation datasets” button on the bottom left of the window’s menu. For more information on the dataset, go to <http://archive.ics.uci.edu/ml/datasets/heart+Disease>

The goal of this assignment is to start with an unfamiliar dataset, frame the problem(s) you are going to use the dataset to solve, build a set of models and then compare the performance of the models. You will be using both regression and logistic regression models.

Specifically, carry out the following steps:

1. The data attribute types have already been set by Orange. Stick with these preset types but feel free to change the target feature to answer the following questions:
   1. Formulate 2 regression problems that the data can be used to investigate. For example, “I will use attributes a, b, c to predict the target attribute y.”
   2. Formulate 2 logistic regression problems that the data can be used to investigate.
2. Pick one of the regression problems you stated above. For this problem, build 3 different regression models (use the default values of each model). Measure each model’s performance on a test dataset. Which model performed the best? Explain your answer. (Be sure to include any pre-processing of the data before building your models.)
3. Pick one of the logistic regression problems you formulated in 1b above. For this problem, build 4 different logistic regression models (use the default values of each model). Measure each model’s performance on a test dataset. Which model performed the best? Explain your answer. (Be sure to include any pre-processing of the data before building your models.)