CST8390 - Lab 7
Regression

Due Date: Week 10 in own lab sessions

Introduction

The goal of this lab is to perform linear regression on housing file.

Steps for Linear Regression:

1. Open the housing.arff file (uploaded in Brightspace) in a text editor to read about the data. Fill in the following questions:

    a. Number of instances:

    b. Number of attributes:

    c. Attribute Information:

2. Start Weka and open the file housing.arff. Find the following information from the preprocess tab. The median is the middle value of a sorted list, so click on the edit tab, and sort the columns and find the middle element:
    a) Median House Value (class) x $1000 :

    b) Median number of rooms per dwelling :

    c) Median per capita crime rate :

3. Click on the Classify tab and choose "LinearRegression" from Functions. Modify the algorithm parameters so that outputAdditionalStats is "true". Ensure that "class" is set for what value is being computed. Run the algorithm to output the weights of the regression. (***Answer should be typed in. Snippet or screenshot not permitted.***)

    a. What is the linear regression model for this set?

    b. Which are the two highest factors which have a positive influence on the housing price?

    c. Which are the two highest factors that have a negative influence on housing price?

In order to get grades,

1. **Upload answer document to Brightspace.**
2. **You should be ready with your results in the result pane and housing file opened in notepad++.**