`

CST8390 - Business Intelligence and Data Analytics

Lab 8 - *Regression*

Name: _____ **- Id:** _____

Due Date: Week 11 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? _____.

REMEMBER:

Show your **answers** to the lab professor when you are done.

You should be ready with your results in the result pane and housing file opened in **Notepad++**.

FOR YOUR ANALYSIS:

** **Option 1**: Explain what a **Regression** is and where to use it.*

** **Option 2**: Explain how to determine the **factors** and their **impact** (positive or negative) to the analysis.*

Ottawa, Mar 2020.