Graduate Program in Software

SEIS 763: ML

Assignment #2 (100 points)

Due Date: June 11st

Write a MatLab (or a programming language of your choice) program to perform or answer the following tasks:

1. Use Matlab command "load `patients`" to load patient self evaluation dataset.
2. You can also load the data from "patients.csv" file.
3. Use variables **`Age, Gender, Height, Weight, Smoker, Location, SelfAssessedHealthStatus`** to build a linear regression model to predict the systolic blood pressure.
4. What are the regression coefficients (thetas)?
5. How do you interpret those numbers?
6. If you need to identify one outlier record, which record is a potential outlier? How do you reach this conclusion?
7. If you need to identify one or few useless features (independent variables or predictors), which one(s) will you choose? Why do you reach this conclusion?

Please follow the instructions below to submit your assignment.

1. If you use Matlab, please name your MatLab program as **"a2.m"**.
2. If you do not use Matlab, please have your non-Matlab program reads in your data from the "C:\tmp" folder.
3. Zip and E-mail the program files and data file (only for non-Matlab program) to the instructor at clai@stthomas.edu before the class on the due date.
4. **TYPE** your answers to a WORD document. Print and bring your program/document to the class on the due date.
5. Please follow the above instructions. Otherwise, your assignment won't be graded and *****NO***** points will be given to your assignment.

Note: Assignment will be collected right before the class. ***Don't forget to include*** your name on the top of your assignment.