Software Engineering and Data Science

SEIS 763-02: Machine Learning

Assignment #2 (100 points)

Due Date: February 17th

Write a program (Python or Matlab) to find results / answers to the following tasks:

1. Load the patient data from "**ML_HW_Data_Patients.csv**" file.
2. Use variables **Age, Gender, Height, Weight, Smoker, Location, SelfAssessedHealthStatus** to build a linear regression model to predict the systolic blood pressure. You do NOT need to split data into training and testing sets.
3. What are the regression coefficients (thetas)?
4. How do you interpret those numbers in thetas?
5. 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?

Submission Guideline:

1. Please include the WORD document to include your answers to the above questions. Please include **your name** on the top of your WORD document.
2. Please print your program (matlab or python) as **PDF / html** and include the **PDF / html** in your submission. Please name your program as "a2.m/.mlx/.py/.inpyb", depending on the programming language / environment you used.
3. Please also include your program in the formats like .m/.mlx/.py/.inpyb in your submission.
4. Prepare EVERYTHING mentioned in the guideline and submit them on **Canvas** no later than the due date. Please do **NOT** zip your files.
5. Please carefully follow the submission guideline. Otherwise, the instructor may not be able to grade your assignment.