Software Engineering and Data Science

SEIS 763: Machine Learning

Assignment #4 (100 points)

Due Date: March 3rd

Write a program with excellent comments using a programming language of your choice to perform **AND** provide answers to the following tasks:

1. Load the patient data from "**ML_HW_Data_Patients.csv**" file (same data file in the last assignment).
2. Use the following **7** variables **Age, Gender, Height, Weight, Smoker, Location, SelfAssessedHealthStatus** to build a linear regression model to predict the systolic blood pressure.
3. Use ****lasso regression**** with ****10-fold cross-validation**** to identify useful predictors.
4. Which top ****TWO**** remaining predictors (with non-zero theta values) are you going to select after the lasso analysis?
5. What is the lambda (λ) value you choose in order to select the top two predictors you identified in the last question?
6. What are the $\boldsymbol{\theta}$ values for the two selected predictors at the lambda (λ) value you identified in the last question?
7. **This question can be more difficult in Python; hence optional if you use Python.** Plot a lasso plot with **OR** without cross-validation. Please have readable tick labels on the X and Y axes in your plot for easy visualization and verification.

Submission Guideline:

1. Please include the WORD document to include your answers (and clearly readable figures/screenshots) 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 "a4.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.