1 Fall 25 Machine Learning Final Project - Requirements and Due Dates

(1) Make a group

Form a group of 3-4 students, then email me who is in your group. The email must have all members CC'd. Email me if you need a group.
If you want to work with less than 3 students, you need my approval. (email me)
(Due: Wed Nov 5th 11pm)

(2) Pick data

You need to pick dataset for your project, decide on the model set-up (Response Variable, Regression / Binary Classification / Multi-level Classification) and get it approved.
You can use websites listed under "Databank" on class webpage, or you can find something on the web. If you happen to have your own data (e.g. from your internship) you can use them as well. If the data was used in one of the lecture videos or assignments, you can't use them.
The data must have enough number of predictors, and enough number of observations. Once you picked your data, email me for approval. The email needs to contain data (or link to the data), brief description of the data, your choice of the response variable and model setting.
(Due: Wed Nov 12th 11pm)

(3) What you need to do

You will be applying various machine learning models you learned in this class to the data you selected. Your primary objective should be to come up with a model that shows most predictive power.
In the end here are the requirements:

- On-line presentation of the project. It should take about 10 min + 5 min Q+A. You need to schedule a presentation time with me.

- After the presentation, you will be submitting the PowerPoint slides used for the presentation, and the R code that you used for the analysis. Your code file must have clear section header.

(Due: Fri Dec 11th 9pm)

(4) PowerPoint structure

Refer to the Project Report Outline below.

(5) You will be graded for:

- Clarity of explanation of how analysis was conducted.
- Whether you used appropriate statistical method for given situation.
- Correctness of your interpretations of the outputs from statistical tools.
- You will not be graded on whether your final model is actually the best model for the dataset or not.

(6) How to submit

After the presentation, submit your PowerPoint, R code, and data on Brightspace under Assessment → Assignment → Final Project

(7) I am available

- Throughout the process, I am available for consulting via email and online/in-person meeting.
- You can also submit draft/partial report to me to get feedback.
- If you are having a coding issue and want me to look at, make sure to send entire code file.

- Any questions/concerns, email nmimoto@uakron.edu.

2 Project Report Outline

Overall Structure

For your 10 min presentation, slides should be less than 13 slides including the title slide. The slides should follow the format listed below.

A. Introduction (1-2 slides)
B. Base Model Fit (1-2 slides)
C. ML models (1-3 slides)
D. The best model (1-2 slides)
E. Further analysis (0-1 slide)

- You must use PowerPoint or equivalent for the report.
- Use figures and tables effectively to show more information with less space.
- Key output, and key plots should be included, but they must be carefully picked. Don't just include all the plots, or just omit all the outputs.
- Make each slide dense. Don't use long sentences.
- Use graphics to explain as much as possible.

A. Introduction (1-2 slides)

- Clearly define the project's objective.
- Brief introduction of dataset. Regression or Classification, Inference vs Prediction, etc.
- Brief list of models you tried.
- Specify where you obtained your data (R package, UCI, Kaggle, etc.)
- Table explaining the variables in the dataset. List out each variable in the data, and its class (numeric, categorical, etc.) If it's categorical, list number of levels. If number of variable is large, the table may have to be summarized.

- Show selected scatter plots of (X_i vs Y).
- Give summary statistic for response variable. (Table or Histogram).
- Selected results of Chi-square association test.

B. Base Model Fit (1-2 slides)

- Base model is multiple regression for a regression problem, and Logistic regression for binary classification problem.
- If you are doing multiclass classification, then no base model is required.
- Apply base model to the data, and show summary of final fit / prediction performance.
- List Model Parameters and fit metrics (RMSE, R-square, or AUC)
- Did you look for better fit by omitting some variables? (if applicable)

C. ML Models (1-2 slides)

- Apply ML models that you learned in this course. State what method was used to select hyper-parameter. (Grid search? Built-in function?)
- Show at least 1 model from each of the methods you tried.
- List of actual hyper-parameter values that was used.
- Plot hyper-parameter (X-axis) vs valid.RMSE and train.RMSE (Y-axis) for regression.
- Summary of CV fit, and final Test fit summary.
- Was there any improvement over the base model?
- How many parameter does this model have?
- Any sign of over-fitting?
- Any convergence problems? How long does it take to fit?