Week #8 - Assignment

Predictive Modeleling

Due: Sunday

Contents

This week's assignment focuses on binary predictive modeling using the logistic regression model. I used the same data set that was used in the previous few weeks to build a logistic predictive model for predicting the occurrence of diabetes. You are expected to use the same data set you used in the three of your assignment to build a predictive logistic regression model.

Please read the note carefully and reproduce my analysis to make sure you understand the concepts and my analysis logic in the analysis.

The write-up of your assignment should be the same as my case study (section 3 of my class note for this week). To be more specific, you are expected to use my case study as a template to complete this assignment. The following are the major components I expected you to include in your report.

- Introduction - description of what you plan to do in the analysis

- Description of data and variables

    - information on the data collection process.

    - list of variable names and **definitions**.

- Research question(s) - what is the objective of the analysis

- Variable transformation and discretization

    - list the numerical variables you standardize

    - list of the variable you discretize

- Data split - the proportions of data for training and testing sets

- Candidate models - you can the candidate model you used in the previous assignment on the multiple logistic regression model.

- The final model selection

    - Cross-validation method

    - ROC approach (**This is optional**)