Assignment 2: HELOC Data Modeling

Date: November 2, 2020

Sumbit (in the ipynb format) via Moodle before 11:59pm November 16, 2020.

Consider again the same HELOC dataset as in the first assignment, and use your UID as the random seed to split the data into training (80%) and testing (20%) sets. In this assignment you are required to perform the generalized additive models plus different methods of feature engineering. You will need to submit your works in the Python notebook format, with reproducible Python codes and adequate description. Indicate your name and UID in the first cell of the notebook.

(1). (25%) Select these five features (x1, x5, x16, x17, x20) for this assignment. Train the generalized additive model together with the binning technique for each selected feature.

(2). (25%) Re-train the generalized additive model in (1) with the piecewise-linear feature engineering.

(3). (25%) Re-train the generalized additive model in (1) with the penalized B-splines with degrees up to 3.

(4). (25%) Compare the models in (1), (2) and (3) in terms of model explainability. Evaluate their model performance on the testing data and report the prediction accuracy. Draw your conclusions about the final model recommendation.