All,

The goal for your fifth assignment is to use Cross Validation (CV) to select the best model to fit three datasets. Two of the data sets are for regression models, one is a classification problem. Your task is to fit various models to this data (using a combination of linear, nonlinear, and cross terms of the predictors) and then use CV to determine which of these models is the best fit.

The datasets:

Banknote Authentication: Data were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. Wavelet Transform tool was used to extract features from images.

Attribute Information:

1. variance of Wavelet Transformed image (continuous)

2. skewness of Wavelet Transformed image (continuous)

3. curtosis of Wavelet Transformed image (continuous)

4. entropy of image (continuous)

5. class (integer)

Your goal is to predict the class (0,1) as a function of the first 4 predictors.

Yacht Hydrodynamics: Prediction of residuary resistance of sailing yachts at the initial design stage is of a great value for evaluating the ship's performance and for estimating the required propulsive power. Essential inputs include the basic hull dimensions and the boat velocity. The Delft data set comprises 308 full-scale experiments, which were performed at the Delft Ship Hydromechanics Laboratory for that purpose.

Attribute Information:

1. Longitudinal position of the center of buoyancy, adimensional.
2. Prismatic coefficient, adimensional.
3. Length-displacement ratio, adimensional.
4. Beam-draught ratio, adimensional.
5. Length-beam ratio, adimensional.
6. Froude number, adimensional.

The measured variable is the residuary resistance per unit weight of displacement:

7. Residuary resistance per unit weight of displacement, adimensional.

Your goal is to model the residuary resistance as a function of the other 6 predictors.

Concrete Strength: Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. These ingredients include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate.

The variables:

Cement (component 1) -- quantitative -- kg in a m3 mixture -- Input Variable
Blast Furnace Slag (component 2) -- quantitative -- kg in a m3 mixture -- Input Variable
Fly Ash (component 3) -- quantitative -- kg in a m3 mixture -- Input Variable
Water (component 4) -- quantitative -- kg in a m3 mixture -- Input Variable
Superplasticizer (component 5) -- quantitative -- kg in a m3 mixture -- Input Variable
Coarse Aggregate (component 6) -- quantitative -- kg in a m3 mixture -- Input Variable
Fine Aggregate (component 7) -- quantitative -- kg in a m3 mixture -- Input Variable
Age -- quantitative -- Day (1~365) -- Input Variable
Concrete compressive strength -- quantitative -- MPa -- Output Variable

Your task is to model the concrete compressive strength in terms of the other predictors.

Please submits the results of your analysis by midnight on Sunday, February 21.

Best,
Dr Semper