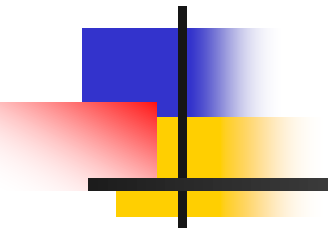

Demo Regression Model Essentials

Dr. Goutam Chakraborty

Outline

- Start with the diagram titled “Diagram1_BAN5743”
  - Cleaned out some additional nodes
- Add Impute node (Modify tab) with following changes and connect to Data partition node
  - Select **Indicator Variables ⇨ Type ⇨ Unique**.
  - Select **Indicator Variables ⇨ Role ⇨ Input**.
- Add Regression node (Model tab), connect to Impute node and run it with default options. Right-click and rename as Regression (No Selection).
  - Run and examine results

Performance of Regression Models

- No Selection:
  - Number of Parameters = 86, ASE (V) = 0.2438, Misclassification (V) = 0.4319

Outline

- Add another Regression node. Rename it as Regression (Stepwise)
  - **Selection Model** ⇨ **Stepwise** on the Regression node Properties panel
  - Run and examine results
  - Compare results with Regression (No selection)

Performance of Regression Models

- No Selection:
  - Number of Parameters = 86, ASE (V) = 0.2438, Misclassification (V) = 0.4319
- Stepwise:
  - Number of Parameters = 13, ASE (V) = **0.2423**, Misclassification (V) = **0.4245**

Outline

- Copy and paste Regression (Stepwise) node. Rename as Regression (Optimal)
  - Select **Use Selection Default ⇨ No** from the Regression node Properties panel
  - Click on **Selection Options** ellipsis button
    - Enter **1.0** as the Entry Significance Level value. Enter **0.5** as the Stay Significance Level value.
    - Change the **Maximum Number of Steps** value to a large value, of your choice (I am using **30** as an example only**)**.
  - *Assuming our interest* in predictions are estimates (or rankings), Select **Selection Criterion ⇨ Validation Error**.
    - Note if our interest in predictions are decisions (instead of estimates), then we need to use **Selection Criterion ⇨ Validation Misclassification**. (Equivalently, you can select **Validation Profit/Loss)**
  - Run and examine results

Performance of Regression Models

- No Selection:
  - Number of Parameters = 86, ASE (V) = 0.2438, Misclassification (V) = 0.4319
- Stepwise:
  - Number of Parameters = 13, ASE (V) = 0.2423, Misclassification (V) = **0.4245**
- Optimal:
  - Number of Parameters = 15, ASE (V) = **0.2417**, Misclassification (V) = 0.4319