MGMT 675
AI-ASSISTED FINANCIAL ANALYSIS

RICE | BUSINESS
Jones Graduate School of Business

CLASSIFICATION

CATEGORICAL TARGET VARIABLES

- Binary (off/on, yes/no, …)
- Multiclass
- Same sets of models: linear, trees, neural nets, …

BINARY EXAMPLES

- Random forest
- Gradient boosting
- Linear (logistic regression)

BINARY DATA

- Upload irrelevant_features.xlsx to Julius
- Ask Julius to read it
- Tell Julius that y2 is the target variable and x1 through x50 are the features
- y2 is a “high-low” version of y1 = x1 + noise.

RANDOM FOREST

- Ask Julius to do a train-test split and train a random forest on the training data.
- Ask Julius to produce a confusion matrix for the training data and a confusion matrix for the test data.
- Ask Julius to produce a ROC curve for the test data and to explain it.

LINEAR MODEL (LOGISTIC REGRESSION)

- For binary variables but can be extended
- Transform binary variable to 0 and 1 dummy variable
- Choose parameters α, β_i to maximize fit of

$$\frac{1}{1 + e^{-\alpha - \beta_1 x_1 - \cdots \beta_n x_n}}$$

  to the dummy variable.
- Can do shrinkage

TO BE CONTINUED