

MACHINE LEARNING

MGMT 675

AI-Assisted Financial Analysis

Kerry Back



RICE | BUSINESS

Jones Graduate School of Business

MACHINE LEARNING IN FINANCE

- Fraud detection
- Credit risk analysis
- Return prediction
- Valuation
- Text analysis
- Time series forecasting

MODELS

- Linear
- Trees
- Neural networks
- Others

REGRESSION VS CLASSIFICATION

- Regression means to predict a continuous variable (not necessarily linear regression).
- Classification is to predict a categorical variable. Binary or multiclass.

TRAIN AND TEST

- Training means fitting a model (like linear regression).
- Objective is to make accurate predictions on new data.
- To assess performance, we have to check the model on “new data” (data not used in training).
- Split data into random train and test subsets. Train on training data. Test on test data.

TEST CRITERIA

- How do we decide if performance is good or bad?
- For continuous variables,
 - usually want to achieve a low sum of squared errors
 - equivalently, achieve a high R^2 .

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

- Categorical also based on prediction errors

EXAMPLE DATA

- Download ml1.xlsx from the [course website](#)
- Upload it to Julius and ask Julius to read it and describe it.
- The data was created by generating 51 sets of 100 standard normals.
 - The first 50 sets are labeled x_1, \dots, x_{50} .
 - The last set was used as the noise to generate y_1 as $x_1 + \text{noise}$.
 - So, x_2, \dots, x_{50} are irrelevant for y_1 .

LINEAR REGRESSION EXAMPLE

- Ask Julius to do a train-test split of the data with 20% of the data in the test set.
- Ask Julius to train a linear regression on the training data with x_1, \dots, x_{50} as the features and y_1 as the target.
- Ask Julius to compute the R-squared on the test data.
- Ask Julius to report the parameter estimates.

TBD