CSCI446/946 Big Data Analytics

Week 6
Regression

School of Computing and Information Technology
University of Wollongong Australia

Task One

- Apply the linear regression algorithm to the "income" dataset by using the following two settings of input variables
  - Age, Education, and Gender (called Model A).
  - Age and Education (called Model B).
  - Observe the outputs and interpret them.

Task One

- Let Age = 41 and Education = 12,
  - Predict the expected income
  - Compute the confidence interval on the expected income
  - Compute the prediction interval on the expected income
  - Observe and interpret their difference

Task Two

- Apply the logistic regression algorithm to the "churn" dataset by using the following three settings of input variables
  - Age, Married, Cust_years, and Churned_contacts (called Model A)
  - Age, Married, and Churned_contacts (called Model B)
  - Age and Churned_contacts (called Model C)
- Observe the outputs and interpret them

Task Two

- Use Model C to
  - Plot the ROC curve and show the AUC value
  - Plot FPR and TPR against the probability threshold for classification
  - Observe and interpret the two plots

Q&A

Images Courtesy of Google Image