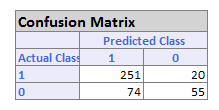
Case Study German Credit by Joshua Troup

**Review the predictor variables and guess from their definition at what their role might be in a credit decision.**

There are 30 predictor variables including the output variable (RESPONSE). Variables NEW\_CAR, USED\_CAR, FURNITURE, RADIO/TV, EDUCATION, and RETRAINING are all CREDIT PURPOSE. Variables DURATION, AMOUNT, INSTALL RATE, CO\_APPLICANT, GUARANTOR are all CREDIT. Variables CHK\_ACCT, SAV\_ACCT, HISTORY, OTHER INSTALL, NIM\_CREDITS are all APPLICANT FINANCIAL STATUS. Variables EMPLOYMENT, MALE\_DIV, MALE\_SINGLE, MALE\_MAR\_WID, PRESENT\_RESIDENT, REAL\_ESTATE, PROP\_UNKN\_NONE, AGE, RENT, OWN\_RES, TELEPHONE, FOREIGN, NUM\_DEPENDENTS are all APPLICANT PERSONAL STATUS.

**Choose one model from each technique and report the confusion matrix and the cost/gain matrix for the validation data. • For the logistic regression model use a cutoff “predicted probability of success” ("success"=1) of 0.5.**

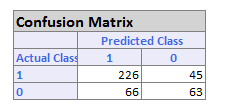


LOGISTIC REGRESSION

74 \* -5 = -370

251 \* 1 = 251

Net gain/cost = -370 + 251 = -119

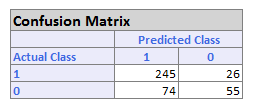


CLASSIFICATION SINGLE TREE

66 \* -5 = -330

226\*1 = 226

Net gain/cost = -330 + 226 =-104



NEURAL NETWORK

74 \* -5 = 370

245 \* 1 = 245

Net gain/cost = -370 + 245 = -125

**Which technique gives the most net profit on the validation data?**

LOGISTIC REGRESSION: -11,900 (DM)

CLASSIFICATION TREE: -10,400 (DM)

NEURAL NETWORK: -12,500 (DM)

All three classifers are showing negative net costs. Classification tree would be the most net profit even though it is still negative.

**d. How far into the validation data do you go to get maximum net profit? (Often this is specified as a percentile or rounded to deciles.)**

Decile 1 = 2,800 DM

Decile 2 = 6,200 DM

Decile 3 = 7,800 DM

Decile 4 = 4,600 DM

**e. If this logistic regression model is scored to future applicants, what "probability of success" cutoff should be used in extending credit?**

The “probability of success” cutoff should be approximately 0.92