

DA2_Homework_Classification_2

Employee turnover

We want to determine based whether or not an employee will leave (*i.e.*, *turn*). We'll build 2 models: a glm binomial regression, and we'll also build an logistic equation and compare results to glm. The dimensions we want to consider in the model are:

- Left (*employee left - turnover*)
- Satisfaction (*latest emp survey*)
- Last_Eval (*lastest evaluation*)
- Number_Projects (*average number of project per month*)
- Avg_Mo_Hrs (*average hours per month*)
- Tenure (*years with company*)
- Promotion (*promotion recieved in last year*)

First check data types on Left between 0 (*did not leave*) and 1 (*left*)

Now run glm to estimate your coefficients (*you want to use the smote data to train the model, but retain the original data for pulling testsets - keep the datatypes in sync!*).

Now, that you have coefficients, create a test file with 100 records (*just use sample_n, 100 on the original data*). Using the glm coefficients, build a logistic regression equation, and calcuate probabilities (*write these to the test dataframe*). Just for confidence, also run the test data through the glm fitted model and compare to your equation results to make sure all agree.

Now set all the records with a probabiliy over 50% to 1 (*Left*), and use a confusion Matrix to score.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0   1
##           0 76 10
##           1 10   4
##
##           Accuracy : 0.8
##           95% CI : (0.7082, 0.8733)
##           No Information Rate : 0.86
##           P-Value [Acc > NIR] : 0.9644
##
##           Kappa : 0.1694
##
## Mcnemar's Test P-Value : 1.0000
##
##           Sensitivity : 0.8837
##           Specificity : 0.2857
##           Pos Pred Value : 0.8837
##           Neg Pred Value : 0.2857
##           Prevalence : 0.8600
##           Detection Rate : 0.7600
##           Detection Prevalence : 0.8600
```

```
##      Balanced Accuracy : 0.5847
##
##      'Positive' Class : 0
##
```

Comments

Comment on your test results.

Show the relationship between Employee Satisfaction and whether they left or not. Should be similar to the following:

