University of Applied Sciences and Arts Northwestern Switzerland
School of Business

DS-04 Classification - Lab 2: Logistic Regression

Dr. Gwendolin Wilke

Task: Fit a Logistic Regression Model

- Fit a logistic regression model to the Data Set `Default` of the `ISLR` library to predict the class variable `default`.

Image source: https://cmci.colorado.edu/classes/INFO-4604/fa17/

Split the data in training and test data using a 20% - 80% split:

- Since standardization of the input variables is not needed for logistic regression, don't reuse the training/tets split you did for kNN, but instead apply the split to the original dataset.
- Call the resulting data sets `test.data` and `training.data`.
- (Remark: We could also use the standardized inputs. For the sake of the exercise, please don't, so that we all get the same slopes.)

Fit a Logistic Regression model with "balance" as only predictor:

```
attach(Default)
glm.fit <- glm(default~balance,family="binomial", data=training.data)
```

- Inspect the model summary and interpret the results. Are the parameters significant? What do they tell you?

```
summary(glm.fit)
```

- Predict the test data probabilities :

```
pred.test.probs = predict(glm.fit, test.data, type = "response")
```

- Calculate the predicted test class values:

```
pred.test.classes = rep("No",nrow(test.data)) # initialize the decision vector
pred.test.classes[pred.test.probs > 0.5] = "Yes"  # populate the decision vector
```

- Calculate the test error rate:

```
mean(pred.test.classes != test.data$default)
```

Fit a Logistic Regression model with ”student” as only predictor.

Fit a Logistic Regression model with all predictors.

Compare the parameters of the last 2 models.

- Can you observe the effect of confounding?

Evaluate all your models:

- Calculate the test error rate of all models and compare them. What is your best model?
- Now compare the test error rates with the test error rate of the ”trivial classifier” that always predicts “No”! (Use `table(test.data$default)` to calculate it.)
- Is your best model better than that?