Machine Learning Lab 02

Maximilian Pfundstein (maxpf364)

2018-11-29

Contents

1 Assignment 2: Analysis of Credit Scoring

1.1 Import `creditscoring.xls`

Let's import the data and have a look at it.

Table 1: creditscoring.xls

resident	property	age	other	housing	existcr	job	depends	telephon	foreign	good_bad
4	1	67	3	2	2	3	1	2	1	good
2	1	22	3	2	1	3	1	1	1	bad
3	1	49	3	2	1	2	2	1	1	good
4	2	45	3	3	1	3	2	1	1	good
4	4	53	3	3	2	3	2	1	1	bad
4	4	35	3	3	1	2	2	2	1	good

1.2 Decision Tree Fitting

Task: Fit a decision tree to the training data by using the following measures of impurity:

a. Deviance
b. Gini index

1.2.1 Deviance

The model for the decision tree using deviance.

```
##
## Classification tree:
## tree(formula = good_bad ~ ., data = train, split = "deviance")
## Variables actually used in tree construction:
```

```
##  [1] "duration" "history"  "marital"  "existcr"  "amount"   "purpose"
##  [7] "savings"  "resident" "age"      "other"
## Number of terminal nodes:  22
## Residual mean deviance:  0.7423 = 277.6 / 374
## Misclassification error rate: 0.1869 = 74 / 396
```

The confusion matrix looks as follows:

	bad	good
bad	49	56
good	43	152

Therefore the error rate is:

```
## [1] 0.33
```

1.2.2 Gini

The model for the decision tree using gini

```
##
## Classification tree:
## tree(formula = good_bad ~ ., data = train, split = "gini")
## Variables actually used in tree construction:
##  [1] "foreign"  "coapp"    "depends"  "telephon" "existcr"  "savings"
##  [7] "history"  "property" "amount"   "marital"  "duration" "resident"
## [13] "job"      "installp" "purpose"  "employed" "housing"
## Number of terminal nodes:  53
## Residual mean deviance:  0.9468 = 324.7 / 343
## Misclassification error rate: 0.2247 = 89 / 396
```

The confusion matrix looks as follows:

	bad	good
bad	25	43
good	67	165
Therefo	re the	error rate is:

```
## [1] 0.3666667
```

1.2.3 Conclusions

Question: Report the misclassification rates for the training and test data. Choose the measure providing the better results for the following steps.

Answer: The misqualification rate for the decision tree with deviance is `0.33` compared to the decision tree with gini as the classifier which has a misqualification rate of `0.3666667`. Therefor we will continue with using the decision tree that uses **deviance** as the classifier.

2 Assignment 3: Uncertainty Estimation

```
set.seed(12345)
```

3 Assignemnt 4: Principal Components

```
set.seed(12345)
```