Predictive Modeling

Business Analytics

Stefan Feuerriegel

Outline

Outline

k-Nearest Neighbor Classification

Question: Can we predict the credit scoring of consumers based on past applications?

- Age in years, income in €1000, number of credit cards

```
age <- c(35,22,63,59,25,37)
income <- c(35,50,200,170,40,50)
creditcards <- c(3,2,1,1,4,6)
train <- as.data.frame(cbind(age, income, creditcards))
```

- Corresponding credit scoring

```
scoring <- c("Bad", "Good", "Bad", "Bad", "Good", "Good")
```

k-NN Classification in R

- Loading required library `class`

```
library(class)
```

- Predictions via `knn(train, test, labels, k)` for `test` data using historic observations `train` with corresponding `labels`
- Predict scoring for person (age 37, € 30000 income, 2 credit cards)

```
# With 1-nearest neighbor
knn(train, c(37, 30, 2), scoring, 1)

## [1] Bad
## Levels: Bad Good

# With 3-nearest neighbor
knn(train, c(37, 30, 2), scoring, 3)

## [1] Good
## Levels: Bad Good
```

- Output: predicted label in 1st row out of all possible labels (2nd row)