MKT 511 Marketing Analytics

Tutorial 3 – Market Analytics

UNIVERSITY OF MANNHEIM
Business School

AACSB ACCREDITED · EQUIS ACCREDITED · ASSOCIATION OF MBAS ACCREDITED

Market Analytics

- **Part A) Market Basket Analysis**
  - Exercise 1
  - Exercise 2
- **Part B) Segmentation**
  - Exercise 3
  - Exercise 4
  - Exercise 5

Part A) Market Basket Analysis

Association Rules

- A **rule** is the association of one set of items conditional on another, e.g., $\{x\} \rightarrow \{y\}$
- Three measures to assess rules:
  - $\boldsymbol{support}\{\boldsymbol{x}, \boldsymbol{y}\}$ is the proportion of all transactions that contain some set
  - $\boldsymbol{confidence}(x \rightarrow y) = \frac{support\{x,y\}}{support\{x\}}$
  - $\boldsymbol{lift}(x \rightarrow y) = \frac{support\{x,y\}}{support\{x\} \times support\{y\}}$

Part A) Market Basket Analysis
Exercise 1

- Transaction data for a Belgian supermarket

- Exploring the data and transforming it into a useful form (**a**)
- Finding association rules in the given data set (**b**)
- Selecting those rules that fulfill certain criteria (**c**)
- Simulating transaction profitability (**d**)

Part A) Market Basket Analysis
Exercise 2

- Non-transaction data of customer segments

- Converting the data into discrete form (**a**)
- Finding association rules in the given data set (**b**)
- Selecting those rules that fulfill certain criteria (**c & d**)

Part B) Segmentation

Exercise 3

- Music service subscription data

- Exploring the data and removing segment information (**a**)
- Applying various cluster algorithms to find groups in the data (**b & c**)

Part B) Segmentation
Exercise 3

- Plotting the results (**d**)

```
labels
```

integer code, currently one of 0,1,2,3,4 and 5. If

labels= 0,

no labels are placed in the plot;

labels= 1,

points and ellipses can be identified in the plot (see `identify`);

labels= 2,

all points and ellipses are labelled in the plot;

labels= 3,

only the points are labelled in the plot;

labels= 4,

only the ellipses are labelled in the plot.

labels= 5,

the ellipses are labelled in the plot, and points can be identified.

Part B) Segmentation

Exercise 4

- Music service subscription data

- Applying model-based clustering (**a & b**)

Table 11.1 Interpretation of the Bayesian Information Criterion (BIC) when comparing two models. Lower BIC is better, and the difference in BIC indicates the strength of evidence. Adapted from Raftery [159], p. 139

BIC difference	Odds of model superiority	Strength of the evidence
0–2	50–75%	Weak
2–6	75–95%	Positive
6–10	95–99%	Strong
>10	>99%	Very strong

R for Marketing Research and Analytics (Chapman & Feit, 2019)

Part B) Segmentation
Exercise 4

- Recoding variables to binary factors (**c**)
- Fitting categorical latent class analysis (**d**)

`formula`	A formula expression of the form `response ~ predictors`. The details of model specification are given below.

Model specification: Latent class models have more than one manifest variable, so the response variables are `cbind(dv1,dv2,dv3...)` where `dv#` refer to variable names in the data frame. For models with no covariates, the formula is `cbind(dv1,dv2,dv3)~1`. For models with covariates, replace the `~1` with the desired function of predictors `iv1,iv2,iv3...` as, for example, `cbind(dv1,dv2,dv3)~iv1+iv2*iv3`.

```
music_LCA_3 <- poLCA(cbind(age, sex, householdIncome, kidsAtHome, commuteCar,
                           musicEnthuse, subscribeToMusic) ~ 1,
                     data = musicdata_lca, nclass = 3)
music_LCA_4 <- poLCA(cbind(age, sex, householdIncome, kidsAtHome, commuteCar,
                           musicEnthuse, subscribeToMusic) ~ 1,
                     data = musicdata_lca, nclass = 4)
```

Part B) Segmentation
Latent Class Analysis

- Can **use k-means clustering approach** to **test for normality**
  - Run cluster analysis with very high k (e.g., 100)
  - Eliminate those observations that are grouped into very small and unusual clusters (outliers)
  - Multivariable technique to find non-normal observations

Part B) Segmentation
Exercise 5

- Music service subscription data

- Splitting the data into training and test set (**a**)

Description

`sample` takes a sample of the specified size from the elements of `x` using either with or without replacement.

Usage

```
sample(x, size, replace = FALSE, prob = NULL)
```

Arguments

`x`	either a vector of one or more elements from which to choose, or a positive integer. See 'Details.'
`n`	a positive number, the number of items to choose from. See 'Details.'
`size`	a non-negative integer giving the number of items to choose.

```
sample(x = nrow(musicdata_raw), size = nrow(musicdata_raw)*0.65)
# works, since nrow( ) returns an integer
sample(x = musicdata_raw, size = nrow(musicdata_raw)*0.65)
# doesn't work, since musicdata_raw is not a vector of integer
```

Part B) Segmentation
Exercise 5

- Fitting different classification models (**b, c & e**)
- Determining important variables (**d**)

```
(music_rf_sub <- randomForest(subscribeToMusic ~ ., data = music_train,
                              ntree = 5000, importance = TRUE))
```

- ( ) around an assignment to simultaneously run the code and print the results (Tutorial 1)
- ntree should be roughly 5 – 10 times the sample size

Tutorial 3 – Market Analytics

UNIVERSITY OF MANNHEIM
Business School

Questions?

Please feel free to ask all of them in the Q&A Forum on ILIAS!

Leonie Gehrmann

leonie.gehrmann@uni-mannheim.de