

# **SBWL Tourismusanalyse und Freizeitmarketing**

**Vertiefungskurs 4: Multivariate Verfahren 2**

**Teil 3: Mischmodelle / Modellgestützte Clusteranalyse**

Achim Zeileis

## Inhalt

---

### Mischmodelle

- Idee
- Latent Class Regression
- EM-Algorithmus
- Beispiel

## Idee

---

**Beispiel:** Es gibt  $k$  Cluster, wobei die Beobachtungen innerhalb eines Clusters eine multivariate Normalverteilung haben. Die Mittelwerte und Kovarianzmatrizen unterscheiden sich dann zwischen den Clustern. Hier spricht man von **modell-basiertem Clustern**.

## Idee

---

### Bisher:

- *Hierarchisches Clustern*: Explorative Einteilung in Gruppen basierend auf Distanzen von Objekten und Clustern,
- *k-Means*: Einteilung in Gruppen basierend auf einer einfachen Zielfunktion (Fehlerquadratsumme).

### Jetzt: Modell-basiertes Clustern

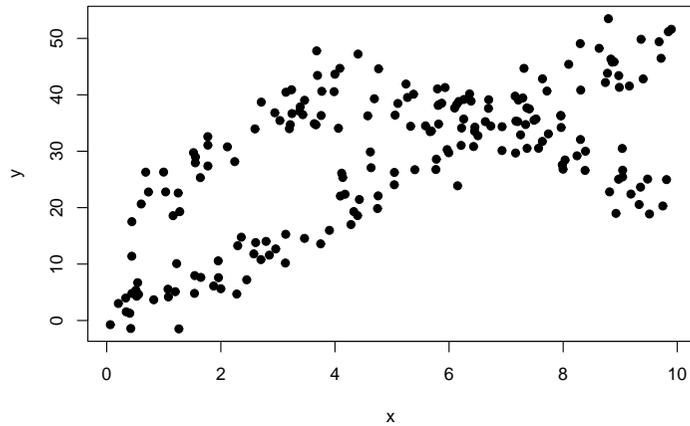
Dabei werden die Daten in Gruppen eingeteilt, wobei in jeder Gruppe ein Modell mit einer Likelihood unterstellt wird, die dann maximiert werden soll.

## Latent Class Regression

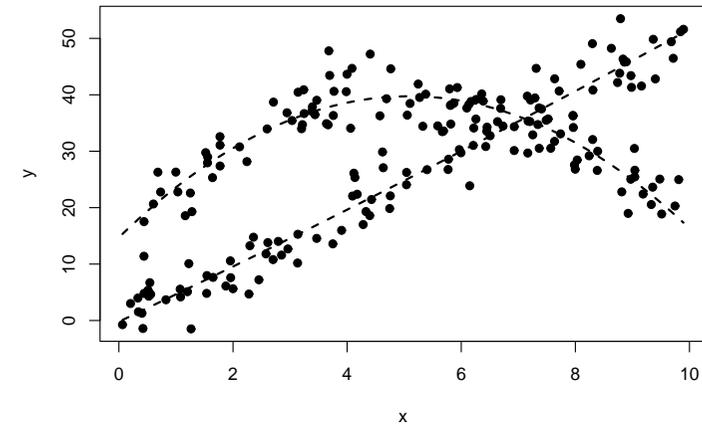
---

**Beispiel:** Es gibt  $k$  Cluster, wobei für die Beobachtungen innerhalb eines Clusters ein GLM angepaßt wird. Die Cluster- oder Klassenzugehörigkeit ist dabei unbekannt (oder unbeobachtet oder **latent**). Dieses Verfahren nennt man daher **latent class regression** oder auch **GLIMMIX** Modell.

## Latent Class Regression



## Latent Class Regression



## Latent Class Regression

Wäre die Einteilung in Klassen bekannt und schon in einem Faktor  $f$  kodiert, dann wäre dies ein ganz einfaches lineares Modell (Kovarianzanalyse), das mit `lm` geschätzt werden kann:

$$\text{lm}(y \sim f * (x + I(x^2)))$$

Dieses Modell entspricht dann zwei quadratischen Kurven, die in den durch  $f$  definierten Klassen angepaßt werden.

## EM-Algorithmus

**Problem:** Die Klasseneinteilung ist unbekannt und muß aus den Daten gelernt werden.

**Lösung:** Gegeben ein in  $k$  Klassen angepaßtes Regressionsmodell wird jede Beobachtung der Klasse zugeordnet, zu deren Regressionsgerade es "besser paßt".

## EM-Algorithmus

---

Damit ergibt sich folgender Schätzalgorithmus:

1. Beginne mit einer Einteilung der Beobachtungen in  $k$  Cluster.
2. Passe in jedem Cluster ein Regressionsmodell an.
3. Ordne jede Beobachtung dem Cluster zu, zu dessen Regressionsgerade sie am nächsten liegt.
4. Wiederhole 2. und 3. bis sich die Cluster nicht mehr ändern.

Dieser erzeugt eine **harte** Partition, wobei jede Beobachtung nur zu genau einem Cluster gehören kann ( $f = 1$  oder  $f = 0$ ).

## EM-Algorithmus

---

Dies ist eine Anwendung des **EM-Algorithmus**. Dieser ist weniger ein konkreter Algorithmus als eine generelle Strategie zur Schätzung von Modellen mit unbeobachteten/latenten Variablen.

Er trägt seinen Namen, da sich immer

- **Erwartungsbildung** (hier: über a posteriori Wahrscheinlichkeiten, Schritt 3.)
  - **Maximierung** (hier: der gewichteten Likelihood, Schritt 2.)
- im Modell abwechseln.

## EM-Algorithmus

---

In der Regel schätzt man mit Mischmodellen aber **weiche** Partitionen, die Variable  $f$  gibt dabei dann die *Wahrscheinlichkeit* an, daß eine Beobachtung zu der einer bestimmten Klasse gehört.

1. Initialisiere die Wahrscheinlichkeiten der Clusterzugehörigkeiten.
2. Gegeben die Gewichtung durch die Wahrscheinlichkeiten, passe  $k$  gewichtete Regressionsmodelle an.
3. Berechne für jede Beobachtung die *a posteriori*-Wahrscheinlichkeit der Zugehörigkeit zu Cluster  $j$ .
4. Wiederhole 2. und 3. bis die Veränderung der Gesamtl likelihood des Modells unter eine Gewisse Schranke fällt.

## EM-Algorithmus

---

Allgemein kann in einem Mischmodell die Dichtefunktion geschrieben werden als

$$h(y | x, \pi, \theta) = \sum_{j=1}^k \pi_j f(y | x, \theta_j)$$

also als Mischung von  $j$  Clustern, wobei  $\pi_j$  die *a priori*-Wahrscheinlichkeit für Klasse  $j$  ist und  $f(y | x, \theta_j)$  die Dichte für die Beobachtungen in dieser Klasse.

In einem normalverteilten GLM wäre  $f$  die Dichte der Normalverteilung und  $\theta_j$  der Vektor der Regressionskoeffizienten.

## EM-Algorithmus

---

Die a priori-Wahrscheinlichkeiten müssen sich zu 1 addieren, d.h.

$$\pi_j \geq 0, \quad \sum_{j=1}^k \pi_j = 1$$

Und damit kann dann die a posteriori-Wahrscheinlichkeit, daß eine bestimmte Beobachtung aus Cluster  $i$  stammt, formuliert werden als

$$P(i|x, y, \pi, \theta) = \frac{\pi_i f(y|x, \theta_i)}{\sum_j \pi_j f(y|x, \theta_j)}$$

## EM-Algorithmus

---

### E-Schritt:

Berechne a posteriori Wahrscheinlichkeit, daß Beobachtung  $i$  aus Cluster  $j$  ist:

$$\hat{p}_{ij} = P(j | x_i, y_i, \hat{\pi}, \hat{\theta})$$

und mittele dann für die Klassenwahrscheinlichkeiten:

$$\hat{\pi}_j = \frac{1}{n} \sum_{i=1}^n \hat{p}_{ij}$$

## EM-Algorithmus

---

Gegeben  $n$  Beobachtungen  $\{(y_1, x_1), \dots, (y_n, x_n)\}$  ist dann die Likelihood der Parameter  $\pi$  und  $\theta$ :

$$\log L(\pi, \theta) = \sum_{i=1}^n \log h(y_i | x_i, \pi, \theta) = \sum_{i=1}^n \log \left( \sum_{j=1}^k \pi_j f(y_i | x_i, \theta_j) \right)$$

Diese muß nun bzgl. der Parameter maximiert werden. In der Regel ist dies nicht analytisch möglich und auch numerisch schwierig.

Deshalb wird die Optimierung in zwei Teile zerlegt

- E-Schritt: Schätzung von  $\pi$  gegeben  $\theta$ ,
  - M-Schritt: Schätzung von  $\theta$  gegeben  $\pi$ ,
- und dann iteriert.

## EM-Algorithmus

---

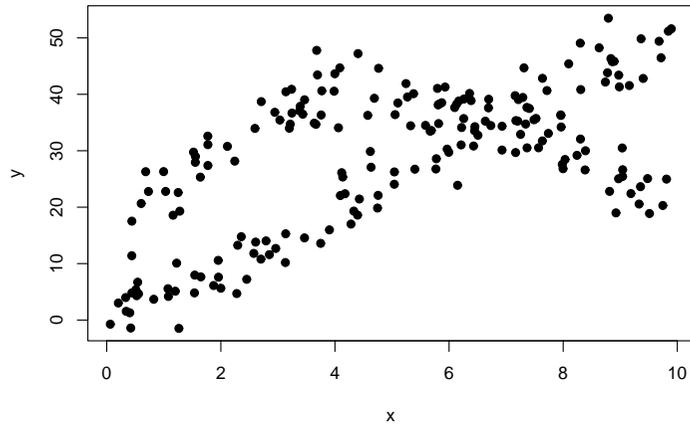
### M-Schritt:

Führe eine gewichtete ML-Schätzung separat für jede Komponente durch. Der Schätzer  $\hat{\theta}_j$  wird dann als Maximalstelle von

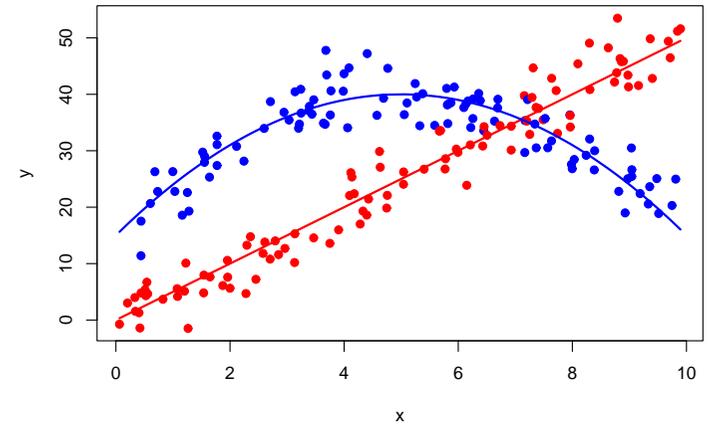
$$\max_{\theta_j} \sum_{i=1}^n \hat{p}_{ij} \log f(y_i | x_i, \theta_j)$$

ermittelt.

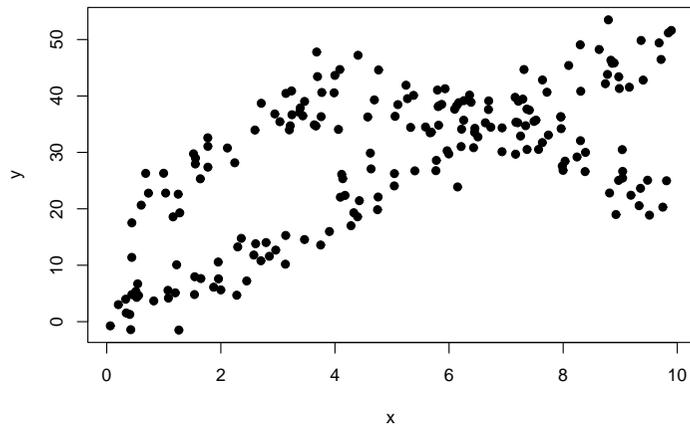
## EM-Algorithmus



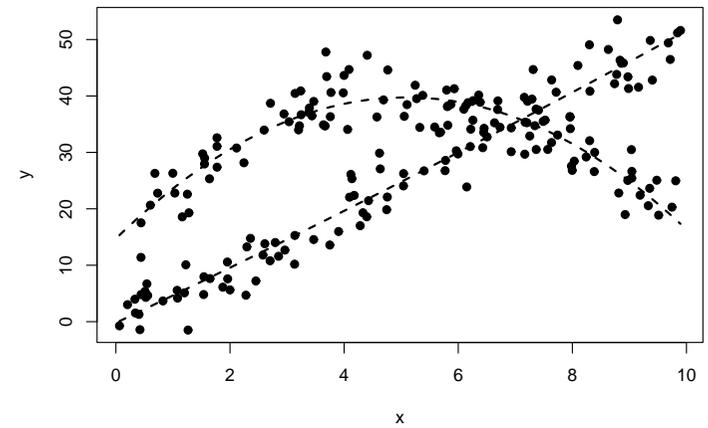
## EM-Algorithmus



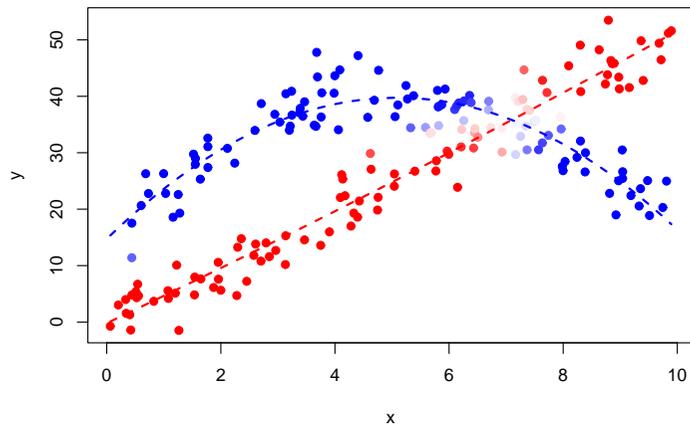
## EM-Algorithmus



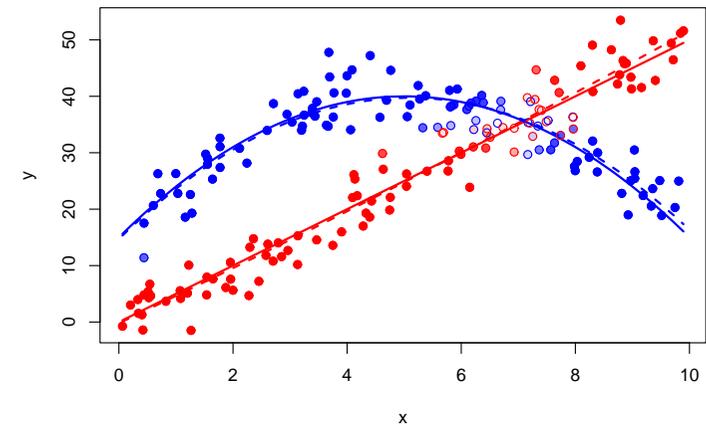
## EM-Algorithmus



## EM-Algorithmus



## EM-Algorithmus



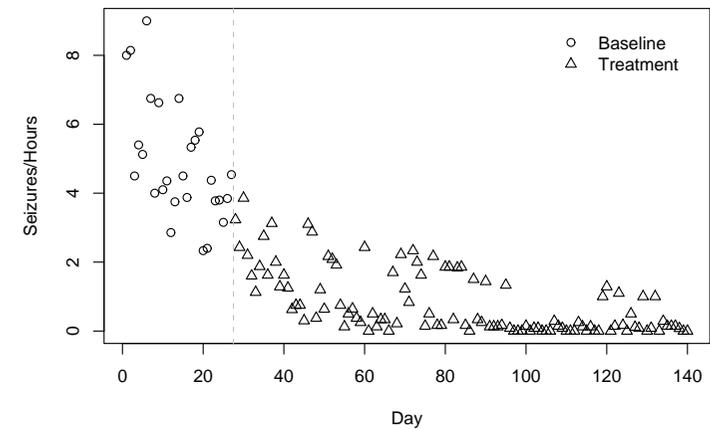
## Beispiel

**Beispiel:** Behandlung eines epileptischen Patienten über 140 Tage hinweg.

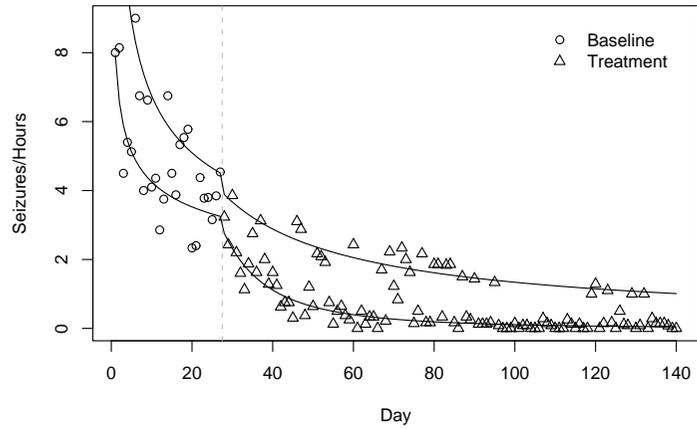
Mit zunehmender Dauer der Betreuung nimmt die Anzahl der epileptischen Anfälle ab. Dieser Effekt kann durch Injektionen von Gamma-Globulin ab dem 28. Tag noch verstärkt werden.

Ein Mischmodell ist deshalb nötig, weil es salopp gesprochen "gute" und "schlechte" Tage des Patienten gibt.

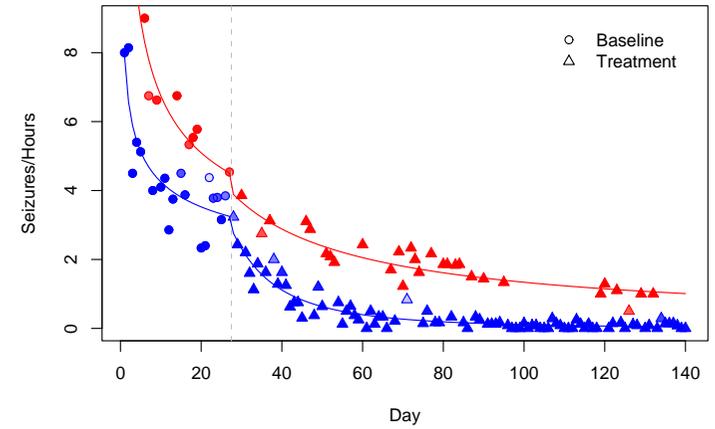
## Beispiel



## Beispiel



## Beispiel



## Tutorium

Mischmodelle in R (*FlexMix.pdf*)