Ecological Language Acquisition via Incremental Model-Based Clustering

Giampiero Salvi
giampi@kth.se

Outlook

- Introduction
- Method
- Experimental settings
- Results

Introduction

- **Background: ecological theory of language acquisition (Lacerda et al., 2004)**
  - **the infant is naïve: no innate linguistic knowledge**

Introduction

- Background: ecological theory of language acquisition (Lacerda et al., 2004)
  - the infant is naïve: no innate linguistic knowledge
- **Aim (long term): mathematical modelling of the learning process**
  - **acoustic features classification**
  - **time integration into meaningful sequences**

Introduction

- Background: ecological theory of language acquisition (Lacerda et al., 2004)
  - the infant is naïve: no innate linguistic knowledge
- Aim (long term): mathematical modelling of the learning process
  - acoustic features classification
  - time integration into meaningful sequences
- **Aim (this study): spectral features classification**
  - **unsupervised**
  - **incremental**

Method

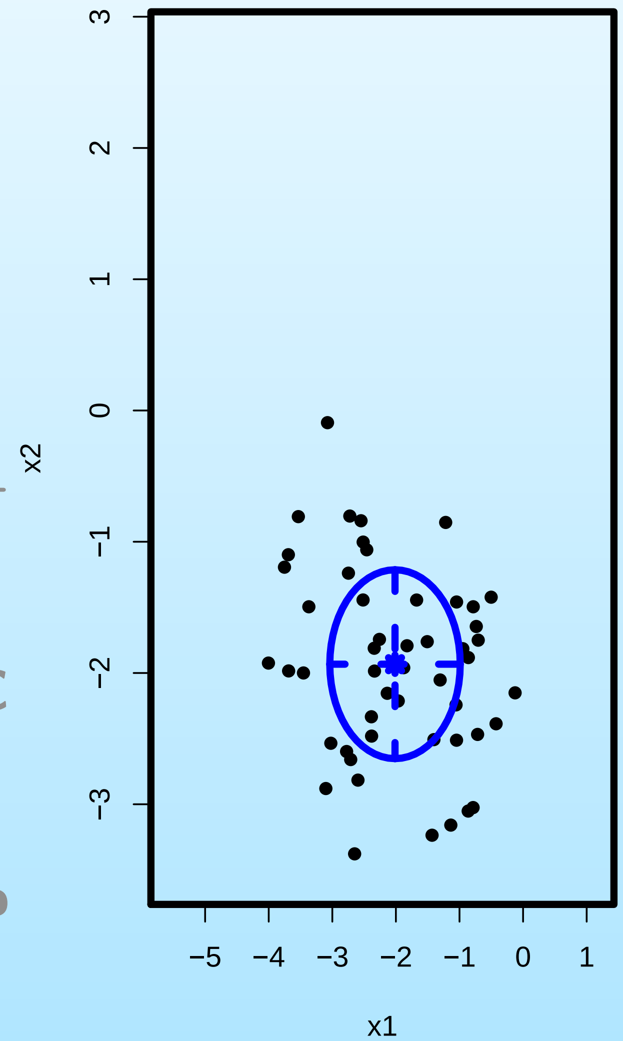
- **Model-Based Clustering (Fraley and Raftery, 1998)**
  - data modelled as mixture of probability distributions
  - each distribution represents a cluster
  - each data point belongs to each cluster with a certain probability
  - model parameters estimated via Expectation Maximisation
  - different models compared via Bayes information criterion (BIC)

Method

- Model-Based Clustering (Fraley and Raftery, 1998)
  - data modelled as mixture of probability distributions
  - each distribution represents a cluster
  - each data point belongs to each cluster with a certain probability
  - model parameters estimated via Expectation Maximisation
  - different models compared via Bayes information criterion (BIC)
- **Incremental Model-Based Clustering (Fraley et al., 2003)**
  - **introduced for large datasets**

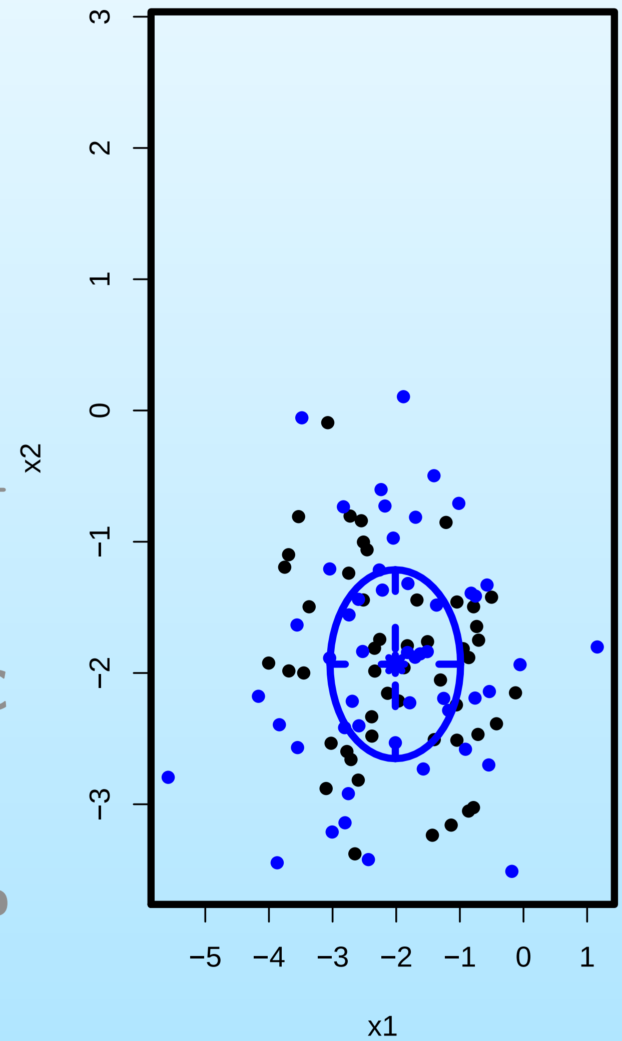
Algorithm

1. **start with a MCLUST model**
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



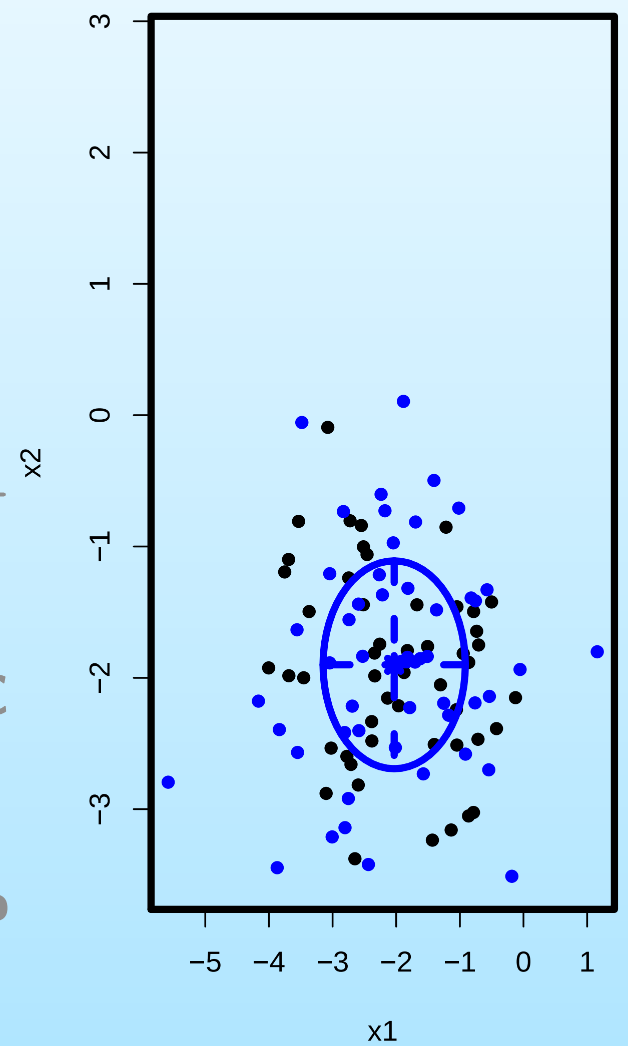
Algorithm

1. start with a MCLUST model
2. **get new data**
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



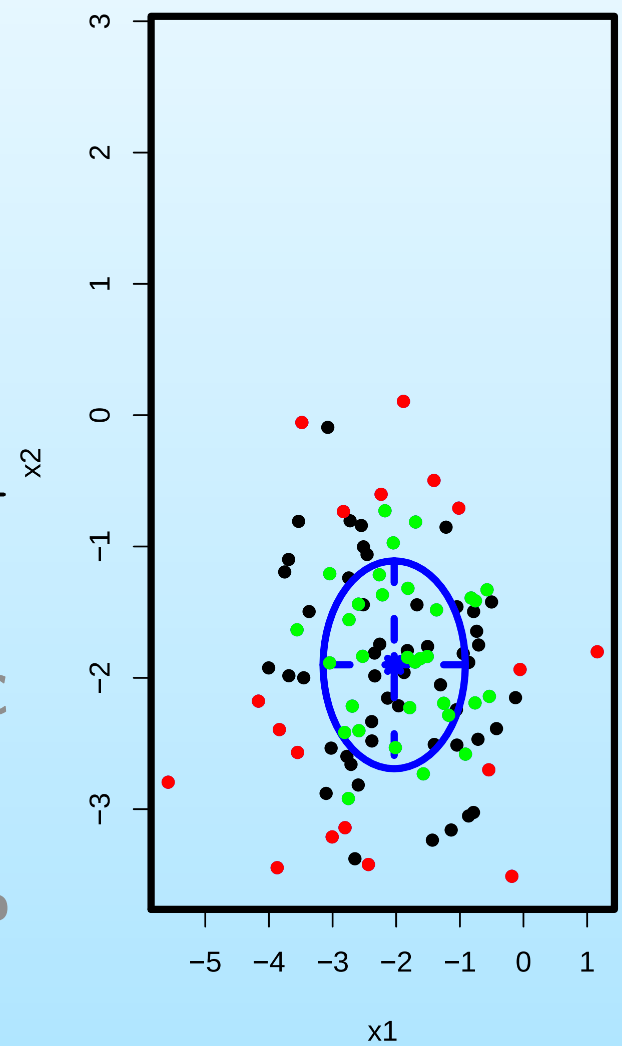
Algorithm

1. start with a MCLUST model
2. get new data
3. **adjust old model to new data**
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



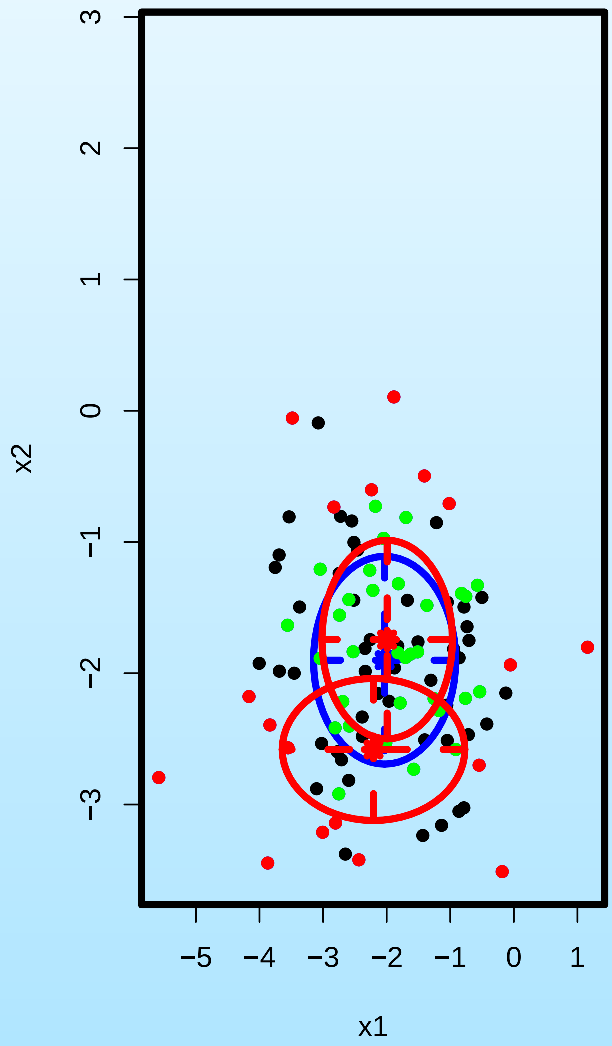
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. **divide new data into well and poorly modelled points**
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



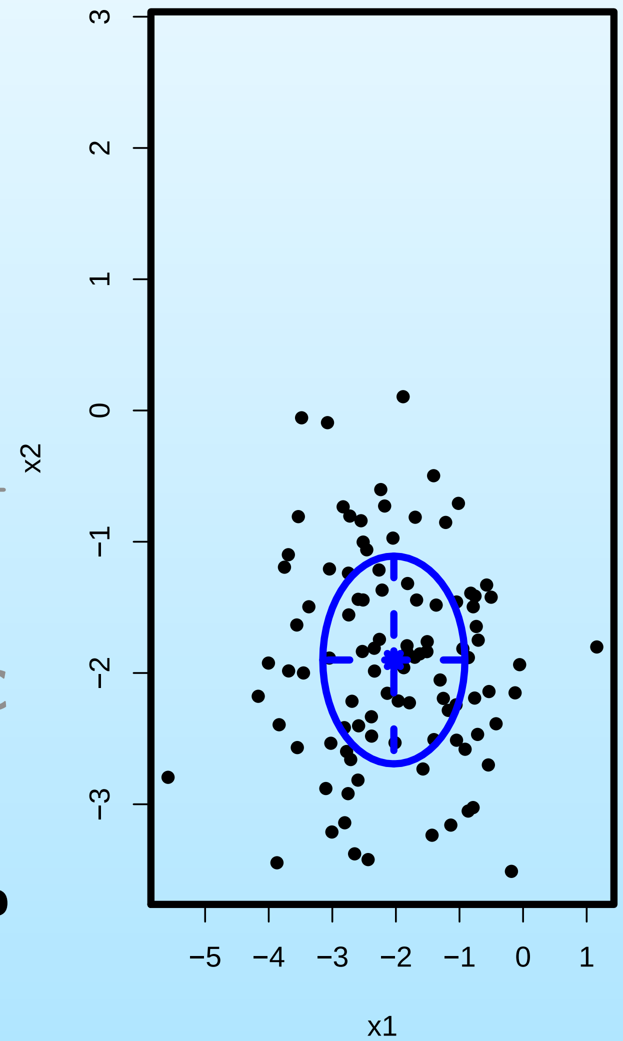
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. **try a more complex model, if better BIC set as best and go back to 4**
6. set the current best model and go back to 2



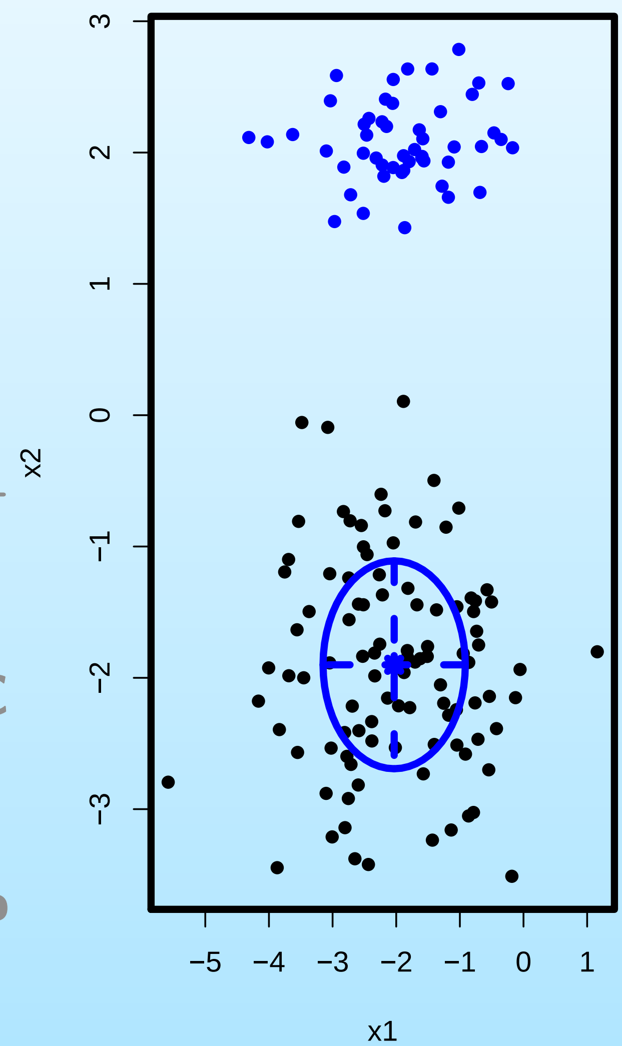
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. **set the current best model and go back to 2**



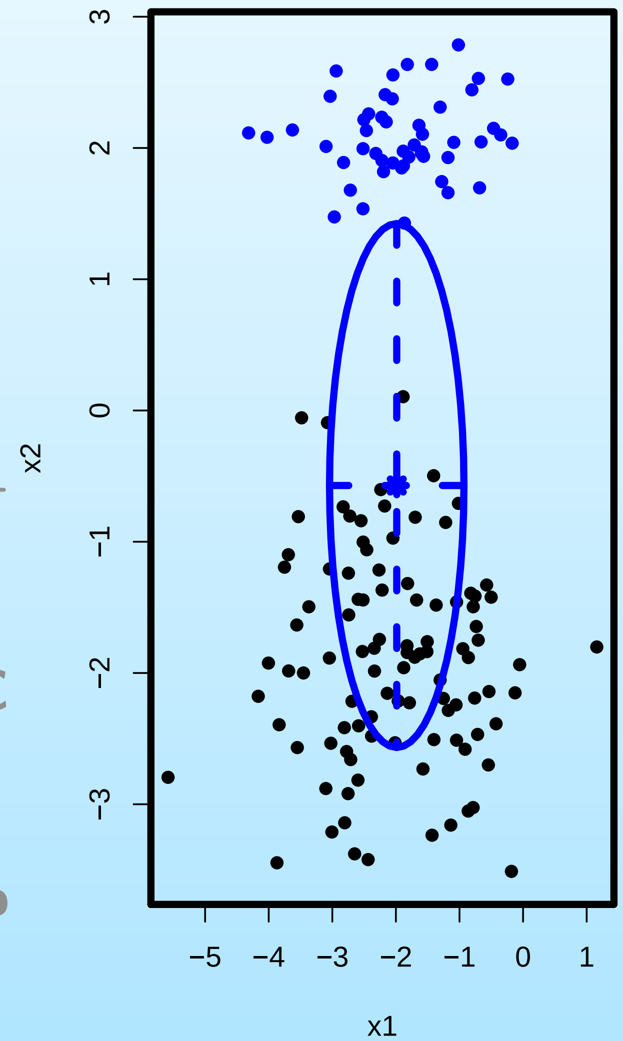
Algorithm

1. start with a MCLUST model
2. **get new data**
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



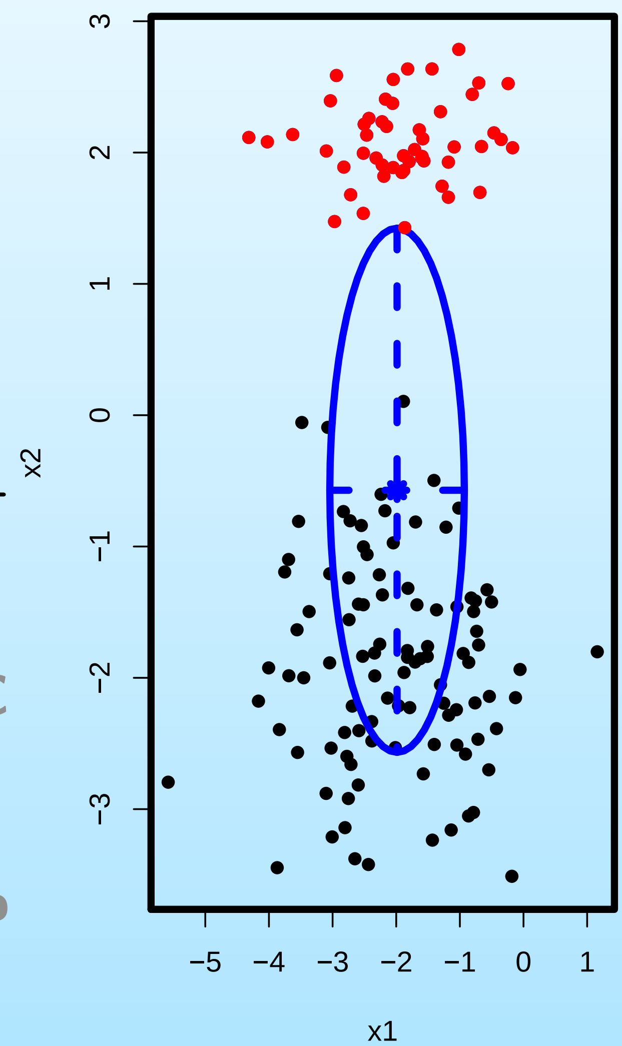
Algorithm

1. start with a MCLUST model
2. get new data
3. **adjust old model to new data**
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



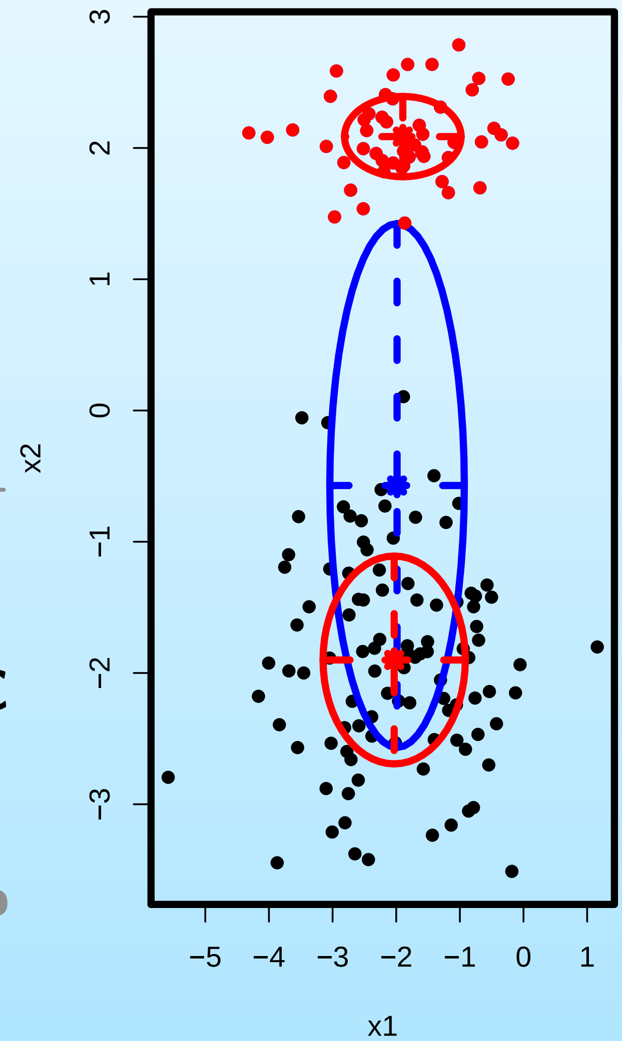
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. **divide new data into well and poorly modelled points**
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



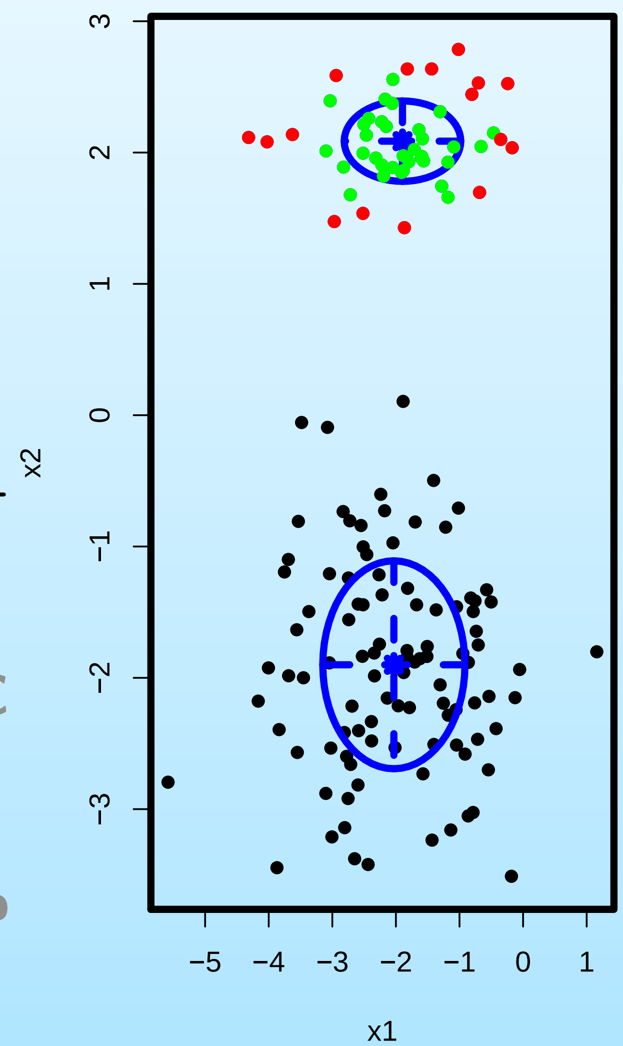
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. **try a more complex model, if better BIC set as best and go back to 4**
6. set the current best model and go back to 2



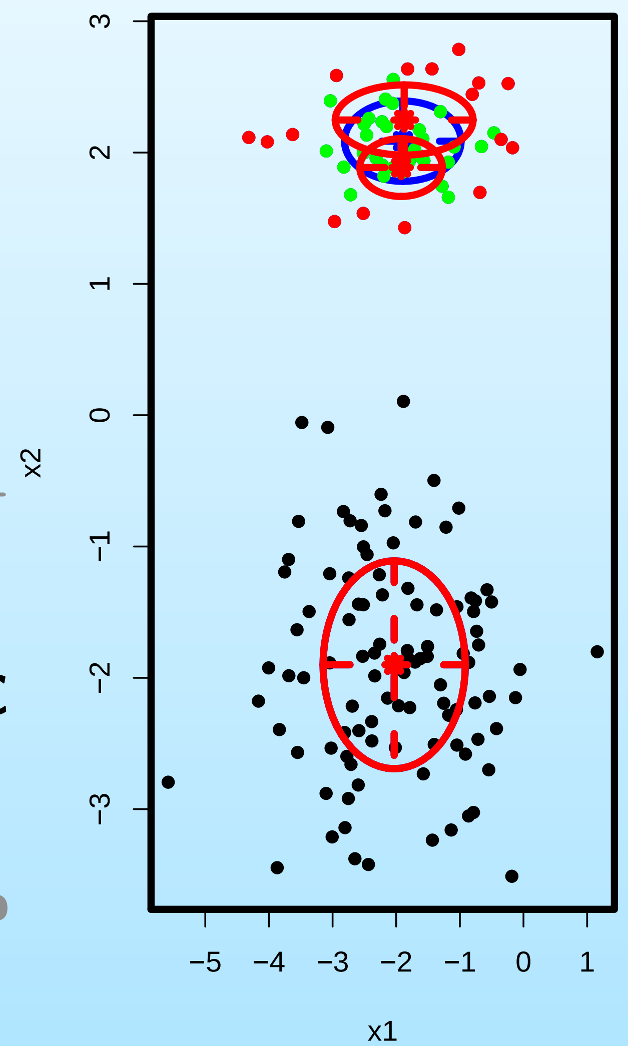
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. **divide new data into well and poorly modelled points**
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2



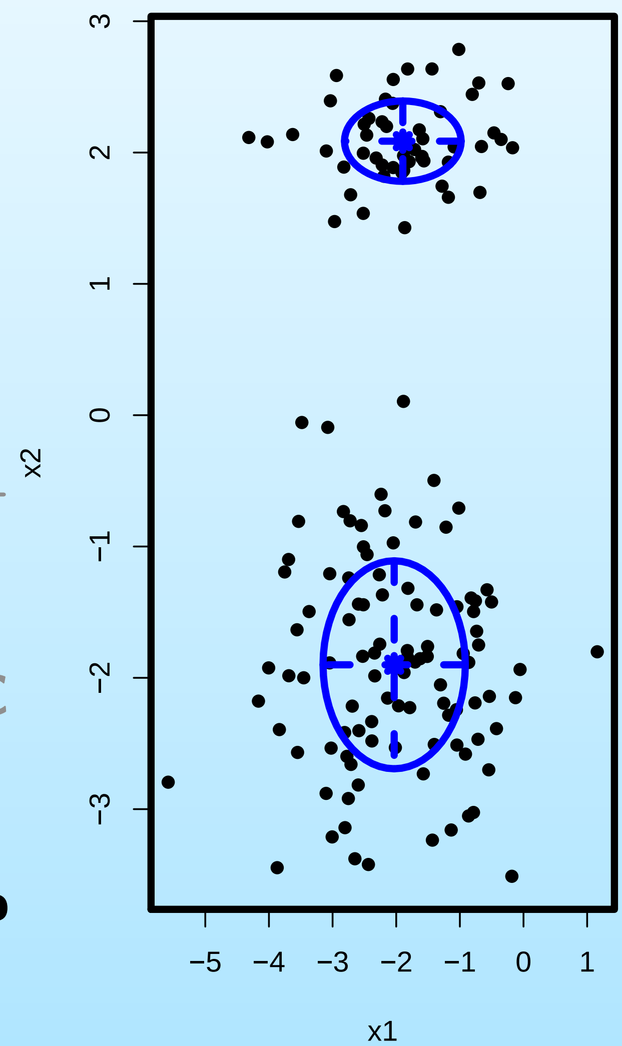
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. **try a more complex model, if better BIC set as best and go back to 4**
6. set the current best model and go back to 2



Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. **set the current best model and go back to 2**



Experimental settings

- **data (ex1, ex2, ex3, ex4, ex5)**
  - **12 minutes from the MILLE corpus**
  - **child directed speech (1 mother talking to her child)**
  - **Mel frequency cepstral coeffs computed every 10ms + differences of first and second order**

Experimental settings

- data (ex1, ex2, ex3, ex4, ex5)
  - 12 minutes from the MILLE corpus
  - child directed speech (1 mother talking to her child)
  - Mel frequency cepstral coeffs computed every 10ms + differences of first and second order
- experimental factors
  - dimensionality of the data: from 3 to 39 dimensions
  - frame length: from 200msec to 3sec

Evaluation

- **problem: there is no reference (at the moment)**

Evaluation

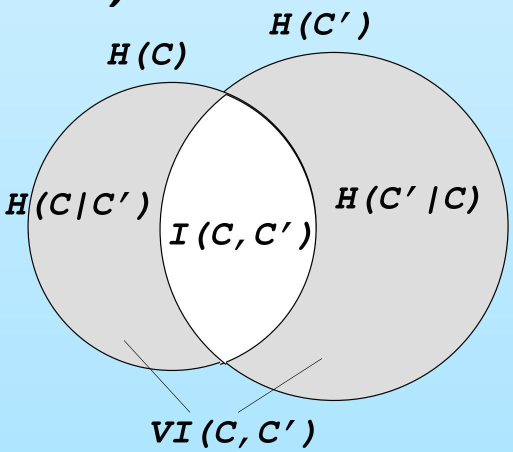
- problem: there is no reference (at the moment)
- **relative evaluation:**

Evaluation

- problem: there is no reference (at the moment)
- relative evaluation:
- time evolution of number of clusters
  - dependency with number of feature coefficients
  - dependency with frame length

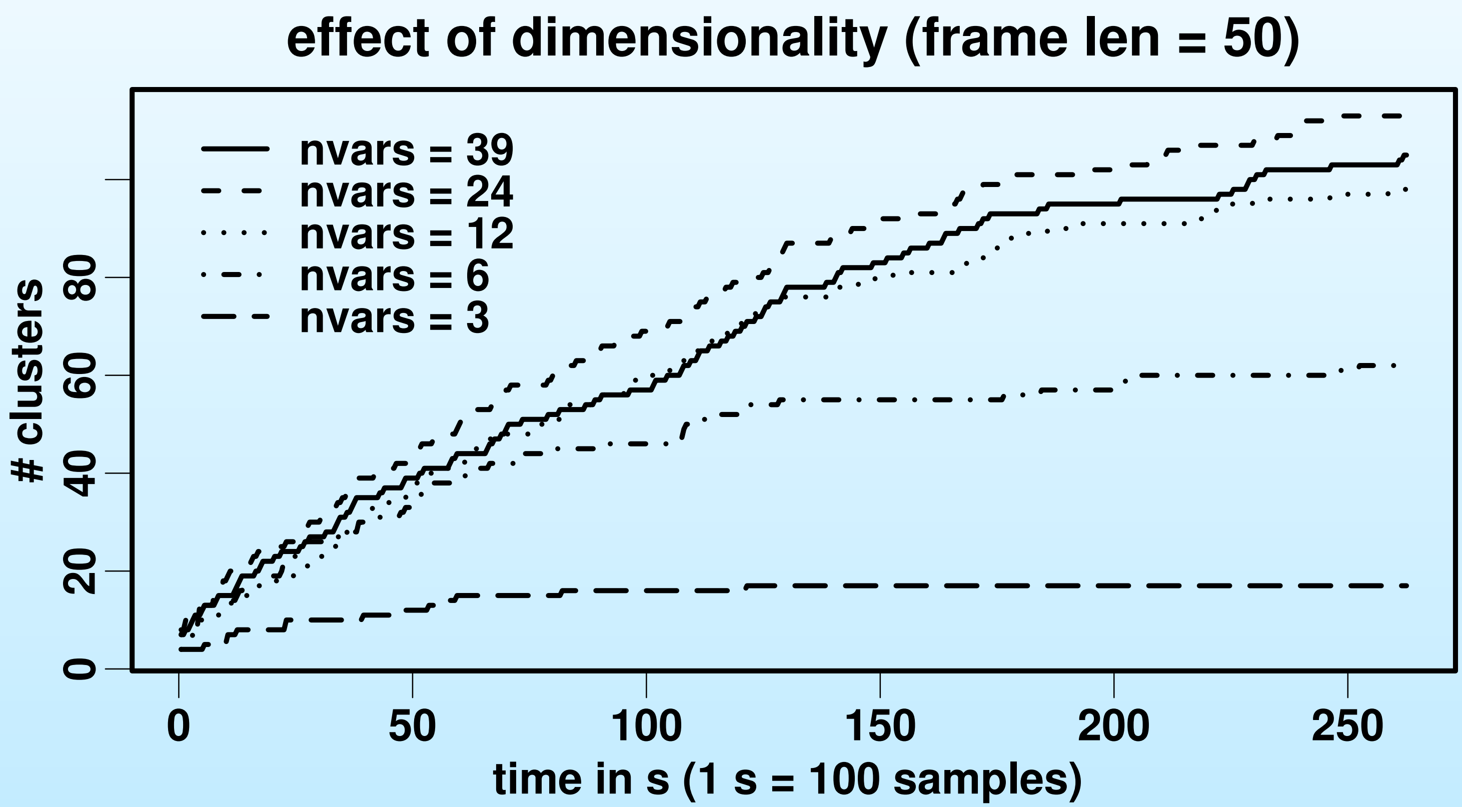
Evaluation

- problem: there is no reference (at the moment)
- relative evaluation:
- time evolution of number of clusters
  - dependency with number of feature coefficients
  - dependency with frame length
- **agreement of classification in different conditions**
  - **variation of information (Meilă, 2002)**



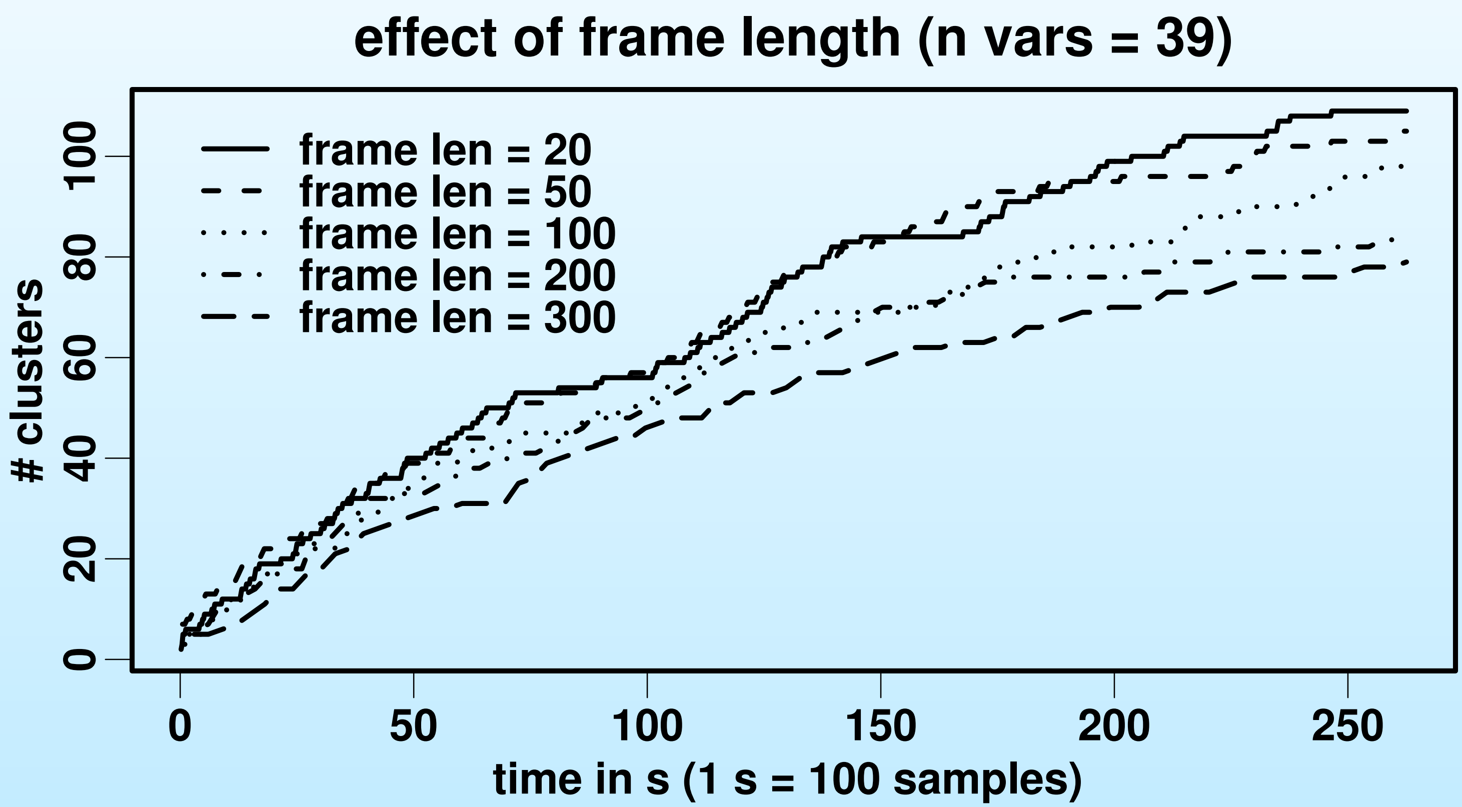
$$\mathbf{V}I(C, C') = H(C|C') + H(C'|C)$$

Results



variation of information	3	6	12	24	39
3	0	0.358	0.435	0.471	0.488
6		0	0.376	0.428	0.460
12			0	0.366	0.407
24				0	0.320
39					0

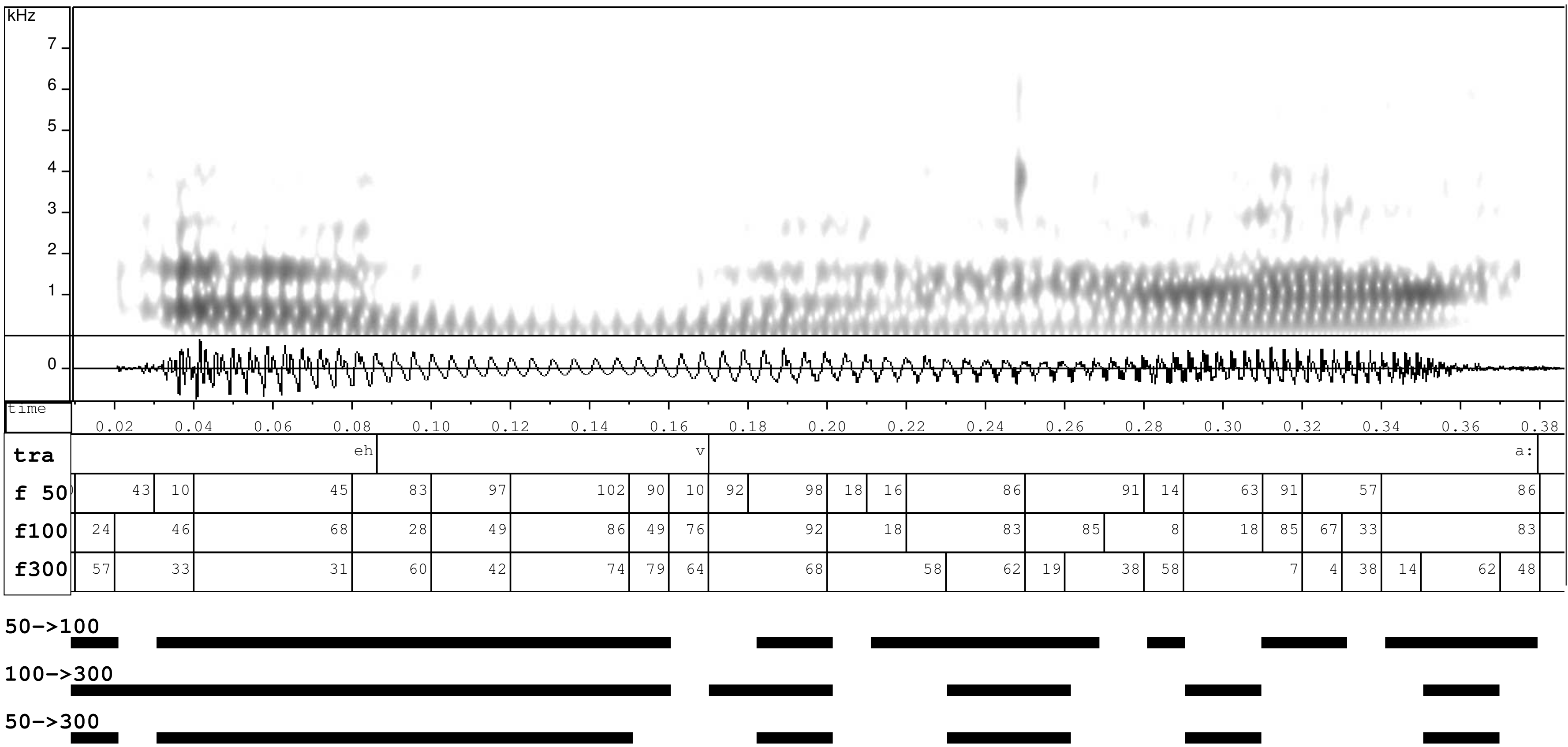
Results



variation of information	20	50	100	200	300
20	0	0.215	0.228	0.253	0.252
50		0	0.195	0.241	0.238
100			0	0.236	0.219
200				0	0.222
300					0

Results

- **example**



Conclusions

- Incremental model-based clustering is a good candidate to model incremental learning
  - gives stable results in different conditions (frame length, dimensionality)
  - the number of clusters increases with new data
  - the rate of increase is larger for high dimensional acoustic features
  - an asymptote is reached at low dimensionality
  - the variation of information can be used to compare classifications
  - probabilistic framework: easier for time integration

Conclusions

- Incremental model-based clustering is a good candidate to model incremental learning
  - gives stable results in different conditions (frame length, dimensionality)
  - the number of clusters increases with new data
  - the rate of increase is larger for high dimensional acoustic features
  - an asymptote is reached at low dimensionality
  - the variation of information can be used to compare classifications
  - probabilistic framework: easier for time integration
- **use IMClust to interpret production and perception data from children studies**

Bibliography

http://www.speech.kth.se/~giampi

Fraley, C., Raftery, A., and Wehrensy, R. (2003). Incremental model-based clustering for large datasets with small clusters. Technical Report 439, Department of Statistics, University of Washington.

Fraley, C. and Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *Computer Journal*, 41(8).

Lacerda, F., Klintfors, E., Gustavsson, L., Lagerkvist, L., Marklund, E., and Sundberg, U. (2004). Ecological theory of language acquisition. In *EPIROB*, pages 147–148.

Meilă, M. (2002). Comparing clusterings. Technical Report 418, Department of Statistics, University of Washington.