

Classification of Clinical Time Series with Constrained estimation of Mixtures of HMMs

Ivan G. Costa¹, Alexander Schönhuth²,
Christoph Hafemeister³, Alexander Schliep³

¹Center of Informatics, Federal University of Pernambuco, Brazil

²School of Computing Science, Simon Fraser University, Canada

³Computational Molecular Biology Department,
Max Planck Institute for Molecular Genetics, Germany

Personalized Medicine

- Clinical diagnosis based on patient's genetic information (gene expression)
- Machine learning methods to
 - classify (1) disease type or (2) response to treatments
- Challenges:
 - clinical studies - few patients (< 100)
 - gene expression – noise, missing data and many genes (> 10.000)
 - patient classification - subjectiveness and error prone
 - errors costs lives

Multiple Sclerosis (MS)

- Autoimmune disease
 - leads to neuronal function loss
 - multiple genetic causes
 - Prevalence: 266,000 in the U.S.
- Treatment with IFN β
 - stop disease progression
 - works only for half of the patients
 - side effects

Classification of Treatment Response

- Clinical Time Series (Baranzini *et al.*, 2005)
 - 52 MS Patients classified for IFN β treatment
 - Good or bad responders
 - Expression of 70 genes over 7 time points
- Classification method (IBIS)
 - 75% of test set accuracy
 - uses only first time point
 - feature selection indicates relevant genes

Baranzini, S.E. *et al.* (2005) Transcription-based prediction of response to ifnbeta using supervised computational methods. *PLoS Biol*, **3**, e2.

Caveats

- Temporal information is relevant
 - patients have own response time (Lin *et. al* 2008)
- MS can have multiple causes
 - response groups may have heterogeneous expression patterns
- Expert classification can be wrong

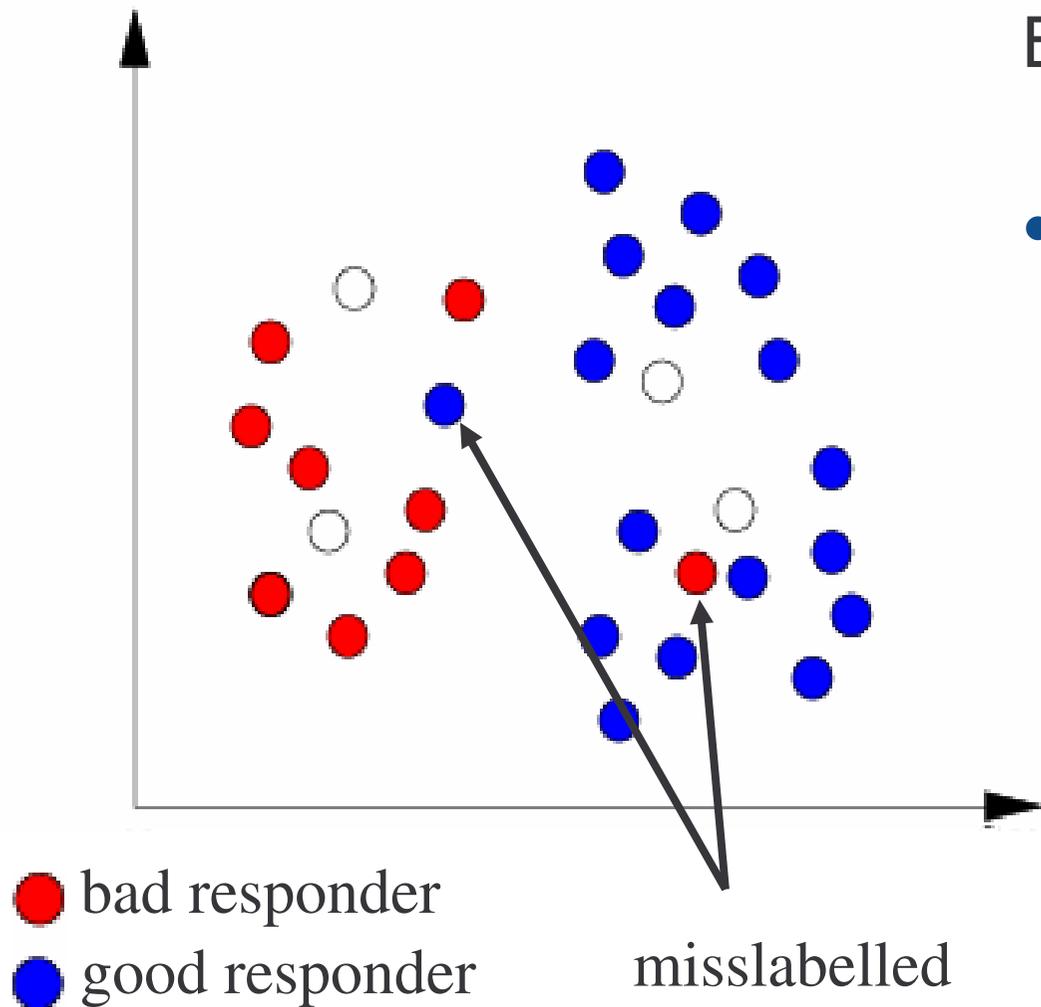
Lin, T. H., Kaminski, N., and Bar-Joseph, Z. (2008). Alignment and classification of time series gene expression in clinical studies. *Bioinformatics*, 24(13), i147–i155.

Our Approach

Mixture Model Classification

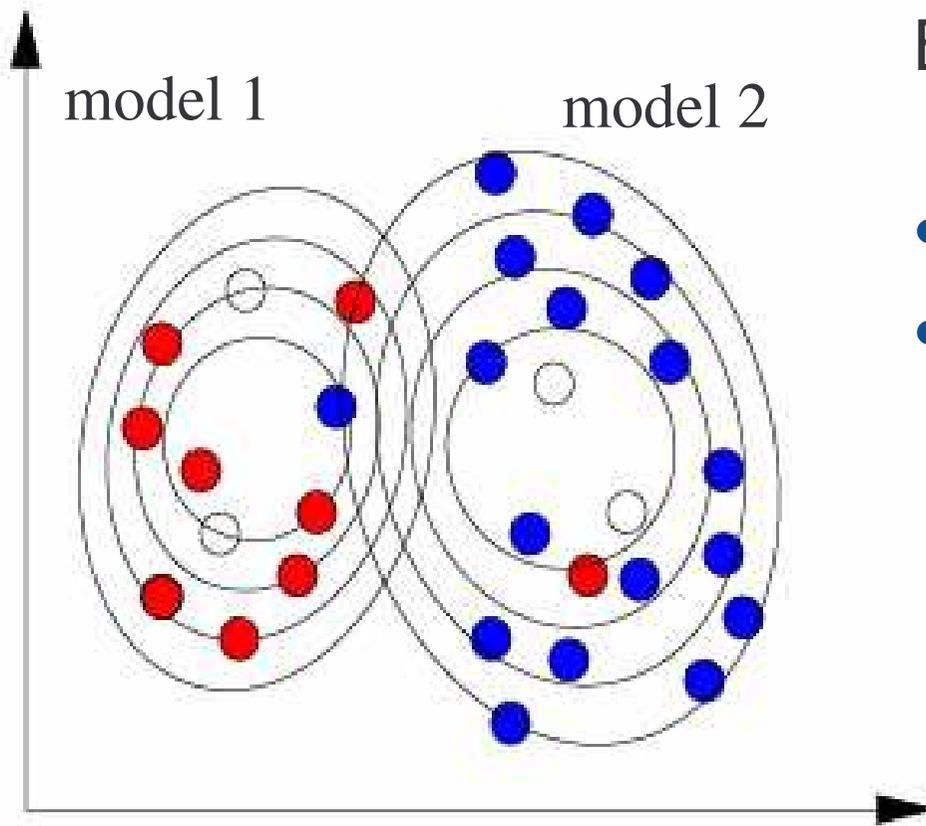
- Mixture Estimation with constraints (semi-supervised)
 - robustness to **wrong labels** 
 - explore **sub-groups** within classes
- Models: linear HMMs
 - **temporal dependencies** in a asynchronous manner

Generative Model Classification (Supervised)



- Estimate models using
train set X and class Y
- $p(x|\theta_{\text{red}})$ e $p(x|\theta_{\text{blue}})$

Generative Model Classification (Supervised)

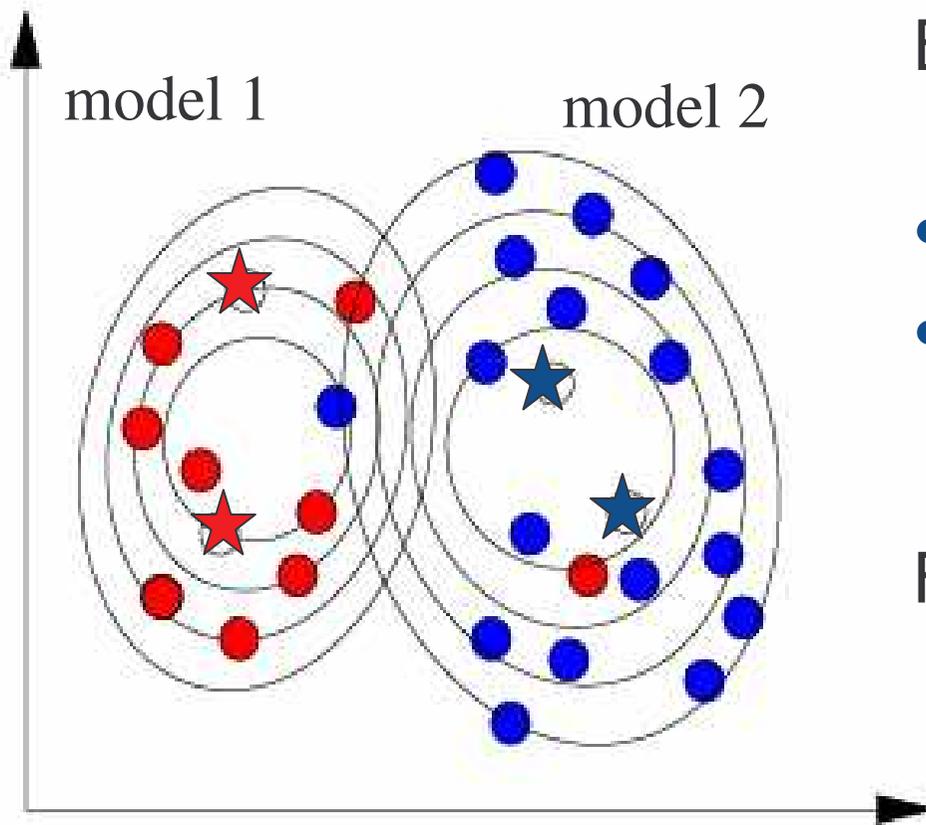


- bad responder
- good responder

Estimate models using
train set X and class Y

- $p(x|\theta_{\text{red}})$ e $p(x|\theta_{\text{blue}})$
- e.g.
 - $p(x|\theta_{\text{red}}) \sim N(x|\mu, \sigma)$

Generative Model Classification (Supervised)



- bad responder
- good responder

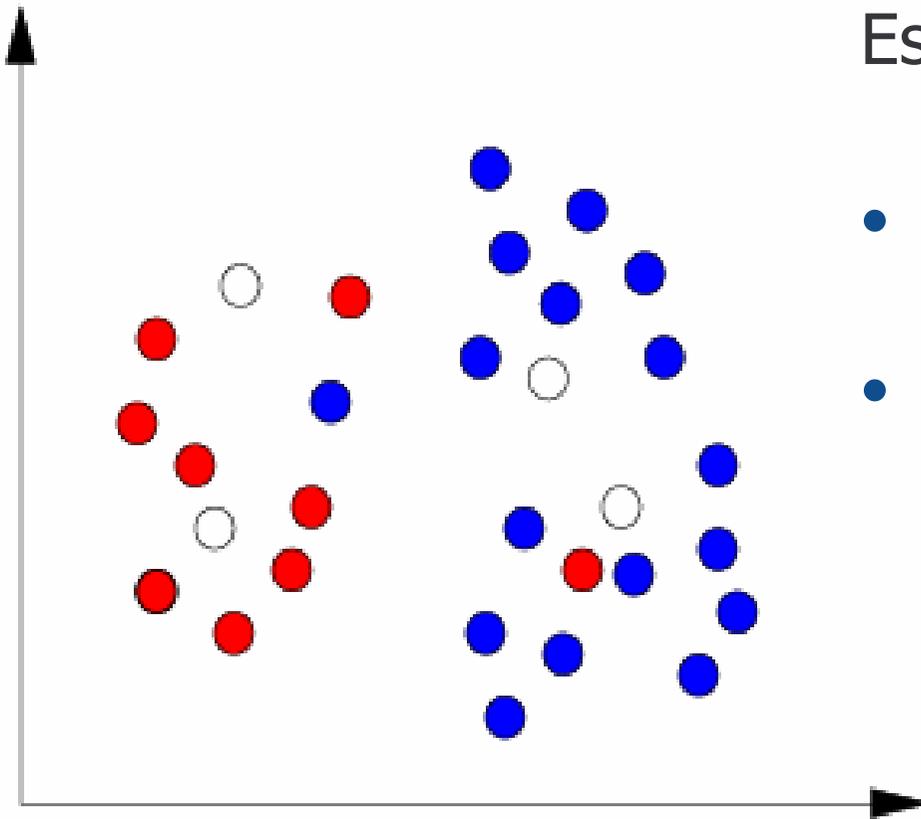
Estimate models using
train set X and class Y

- $p(x|\theta_{\text{red}})$ e $p(x|\theta_{\text{blue}})$
- e.g.
 - $p(x|\theta_{\text{red}}) \sim N(x|\mu, \sigma)$

For test sample x' find y'

$$p(y_i'=\text{red}|x_i') \sim p(x_i'|\theta_{\text{red}})$$

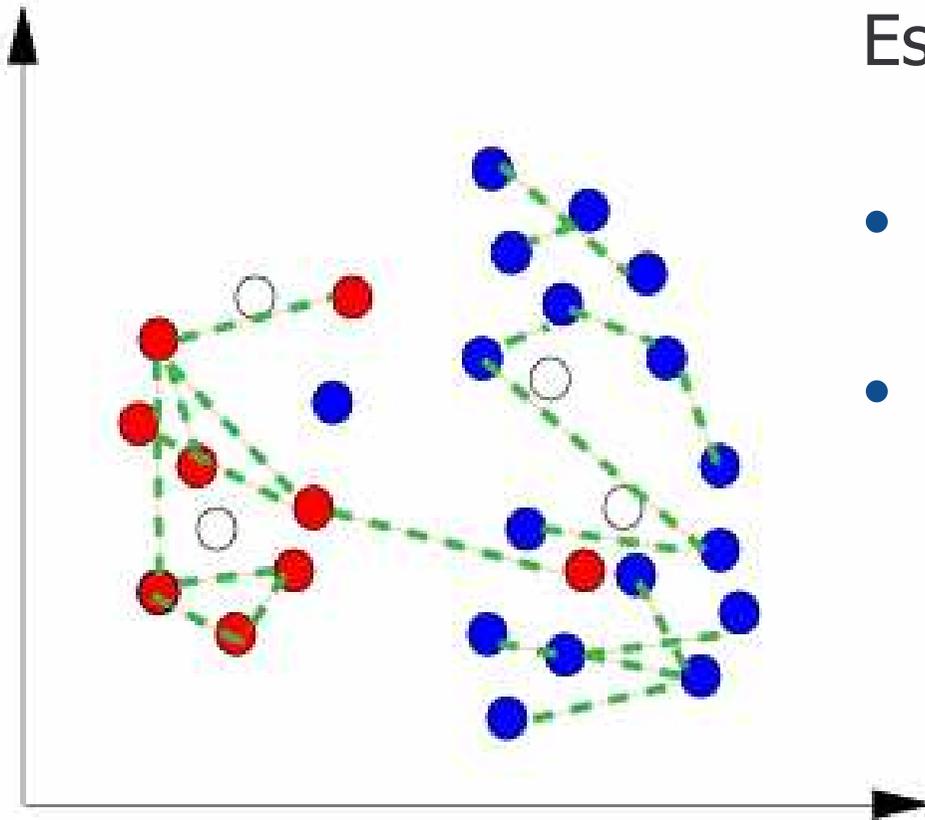
Constrained based Mixture Estimation (Semi-supervised)



Estimate models using data set X and constraints W

- positive – pairs at same class
- negative – pairs at distinct class

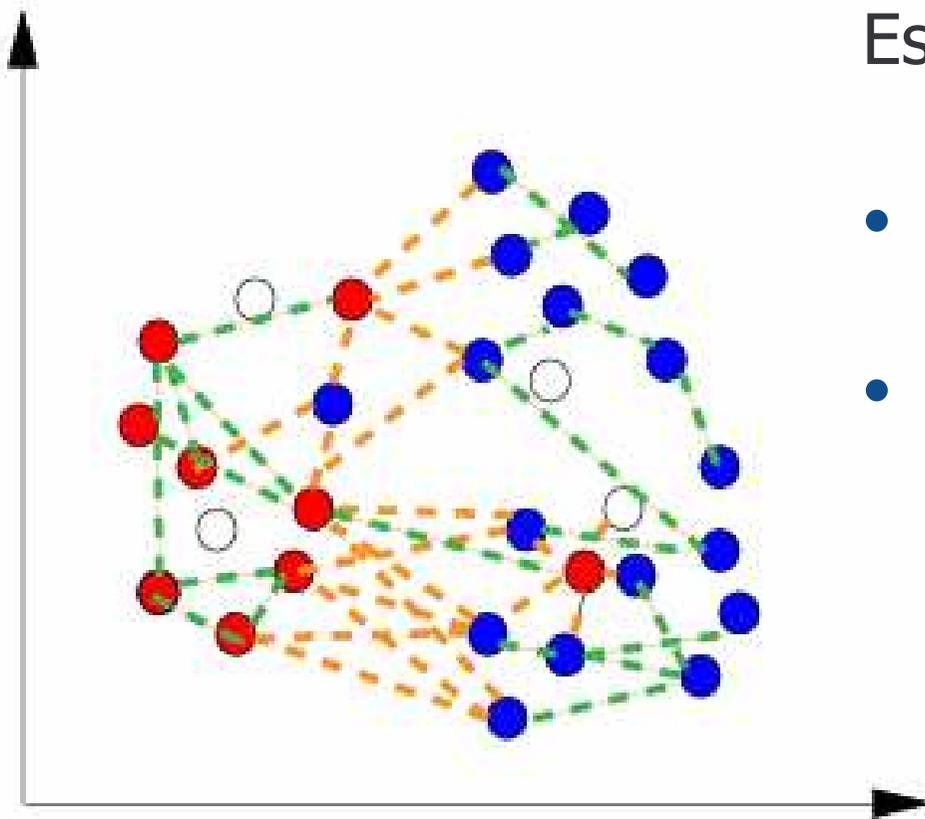
Constrained based Mixture Estimation (Semi-supervised)



Estimate models using data set X and constraints W

- **positive** – pairs at same class ($w+$)
- **negative** – pairs at distinct class ($w-$)

Constrained based Mixture Estimation (Semi-supervised)



Estimate models using data set X and constraints W

- **positive** – pairs at same class ($w+$)
- **negative** – pairs at distinct class ($w-$)

Constrained based Mixture Estimation (Semi-supervised)

- Maximize the complete likelihood (Lu and Leen, 2004):

$$P[X, Y | W, \Theta] = P[X | Y, \Theta] P[Y | \Theta] P[W | \Theta]$$

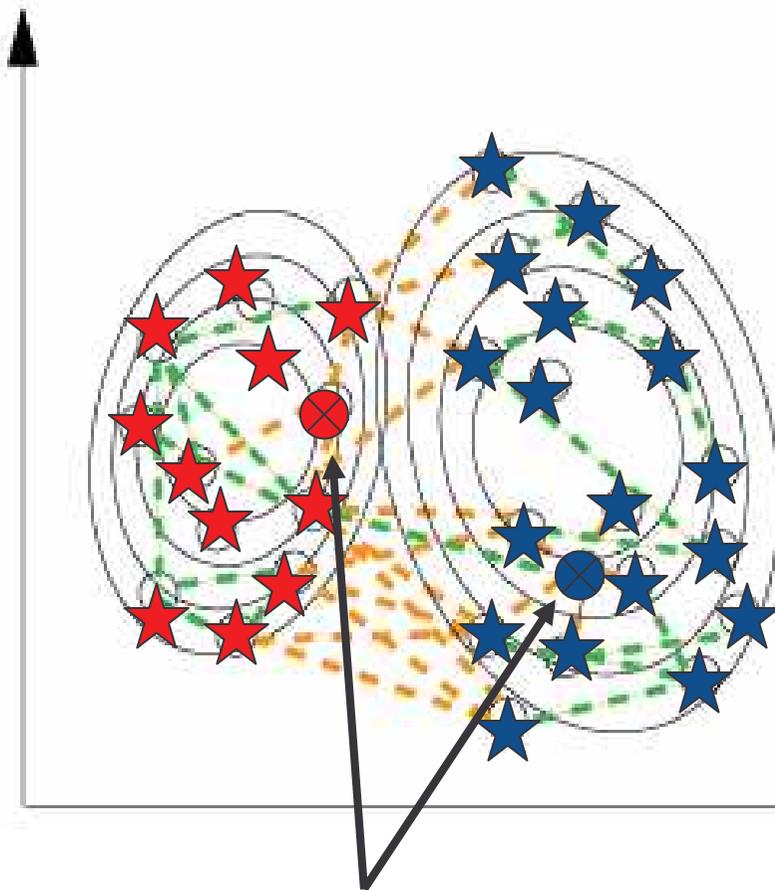
where X is the time series data, Y class assignments, Θ the model parameters, W the constraints

- Constraints penalize solutions

$$P[W | \Theta] \approx \exp \sum_i \sum_{j \neq i} -w_{ij}^+ 1\{y_i \neq y_j\} - w_{ij}^- 1\{y_i = y_j\}$$

Lu, Z. and Leen, T. (2005). Semi-supervised learning with penalized probabilistic clustering. In *Advances in Neural Information Processing Systems 17*, 849–856. MIT Press.

Constrained based Mixture Estimation (Semi-supervised)



mislabelled → “misclassified”

Estimate models using data set X and constraints W

- **positive** – pairs at same class ($w+$)
- **negative** – pairs at distinct class ($w-$)

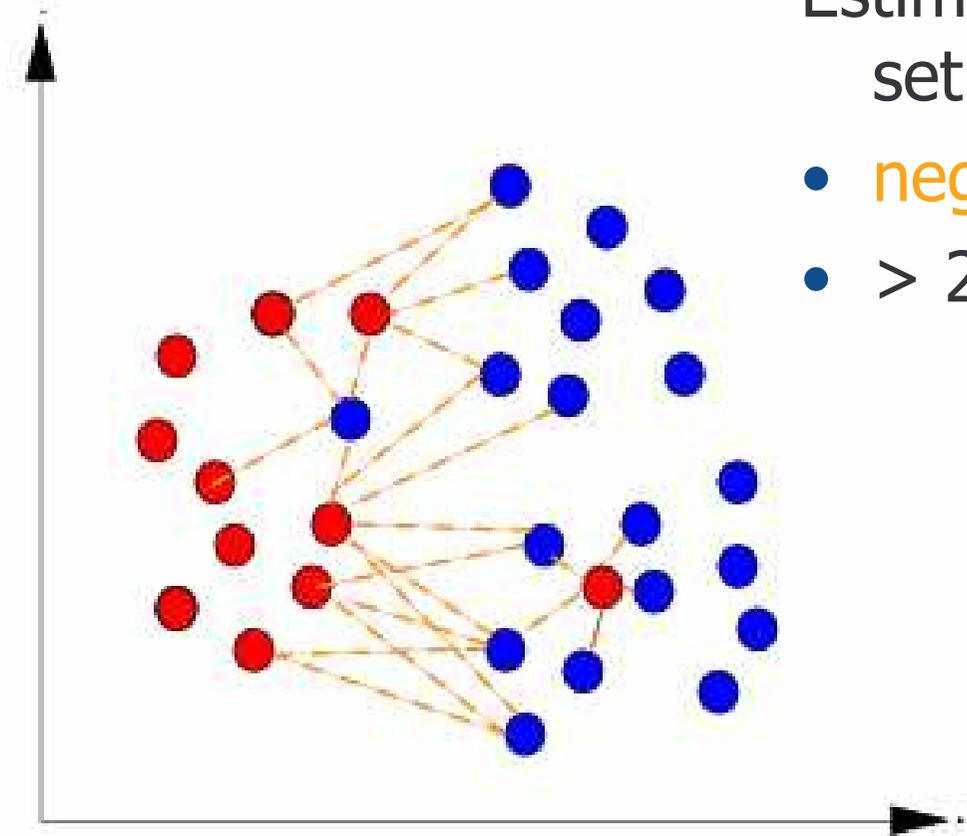
Constrained E.M. algorithm
(Lu and Leen, 2005)

- $p(y_i = \text{red} | x_i) =$
▶ $p(x_i | \theta_{\text{red}}) \cdot p(y_i = \text{red} | W)$

Our Approach

- Mixture Estimation with constraints (semi-supervised)
 - robustness to **wrong labels**
 - explore **sub-groups** within classes ←
- Models: linear HMMs
 - **temporal dependencies** in a asynchronous manner

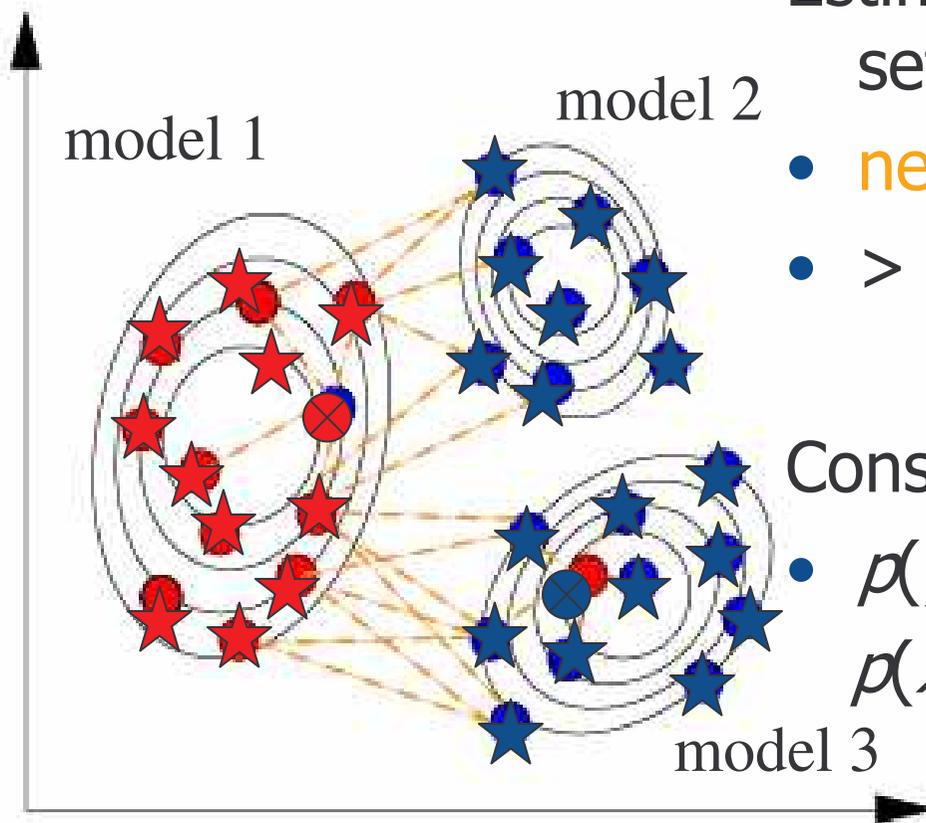
Finding Sub-Groups



Estimate models using data set X

- **negative** constraints only
- > 2 models

Finding Sub-Groups



Estimate models using data set X

- **negative** constraints only
- > 2 models

Constrained E.M. algorithm

- $p(y_i = \text{red} | x_i) = p(x_i | \theta_{\text{red}}) \cdot p(y_i = \text{red} | W)$

Our Approach

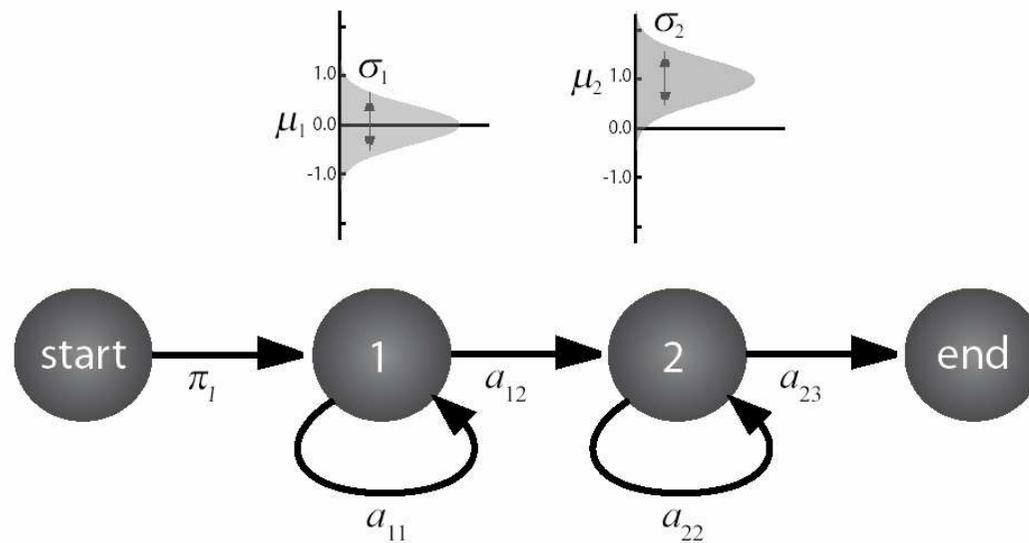
- Mixture Estimation with constraints (semi-supervised)
 - robustness to **wrong labels**
 - explore **sub-groups** within classes
- Models: linear HMMs
 - **temporal dependencies** in a asynchronous manner 

linear HMM

We make the class models to be a linear HMM

– $p(x|\theta_{\text{red}}) \sim P(x|\lambda_{\text{red}})$

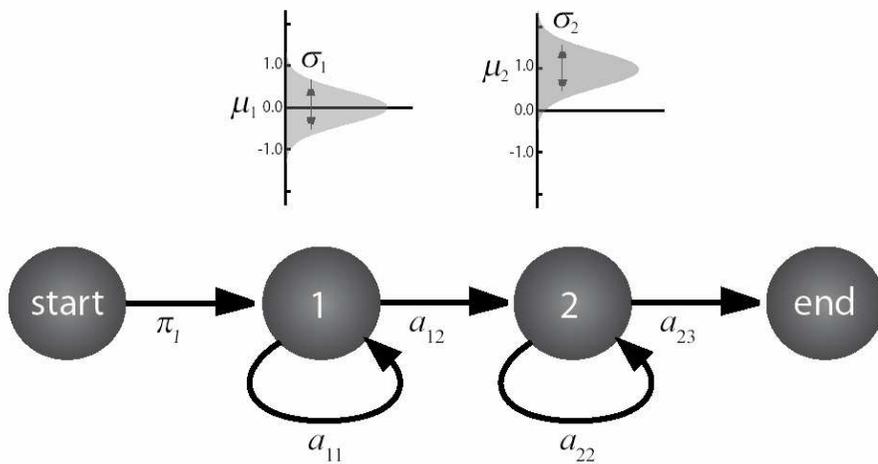
where $\lambda_{\text{red}} = (\pi, A, (\mu_1, \dots, \mu_k), (\sigma_1, \dots, \sigma_k))$



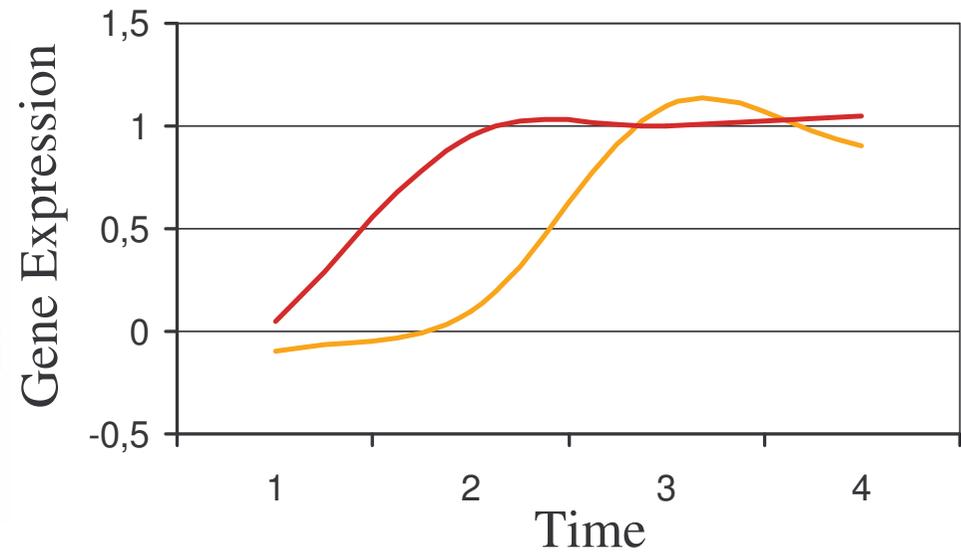
up-regulation model

linear HMM

up-regulation model



up-regulation time-series



State 1 1 2 2
State 1 2 2 1 2

Experiments

- Comparison with
 - IBIS (Baranzini *et al.*, 2005)
 - SVM Kalman (Borgwardt, *et al.*, 2006)
 - HMM Discriminant Learning (Lin *et al.* 2008)
- Experiments
 - 5 times 4-fold cross validation
 - feature selection and number of sub-groups
 - Based on training error
 - HMM with 4 states

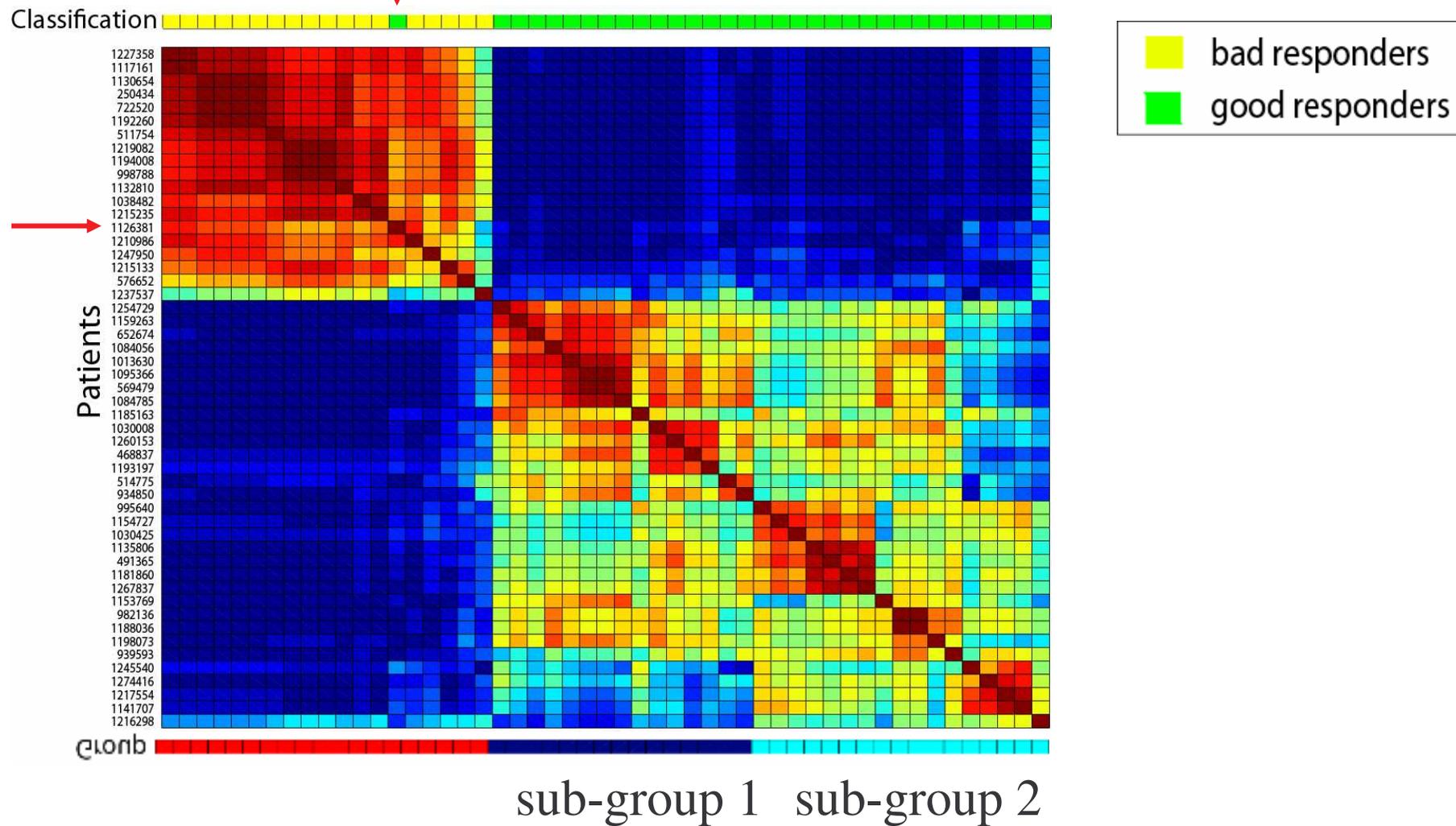
Borgwardt, K. M., *et al.* (2006). Class prediction from time series gene expression profiles using dynamical systems kernel. Pacific Symposium on Biocomputing, 11, 111–122.

Results

Method	Genes	Test Acc.
IBIS	3	75.0%
HMM Disc	7	85.0%
SVM Kal.	70	87.8%
HMM Const 2.	17	89.62%*
HMM Const 3.	17	90.39%*

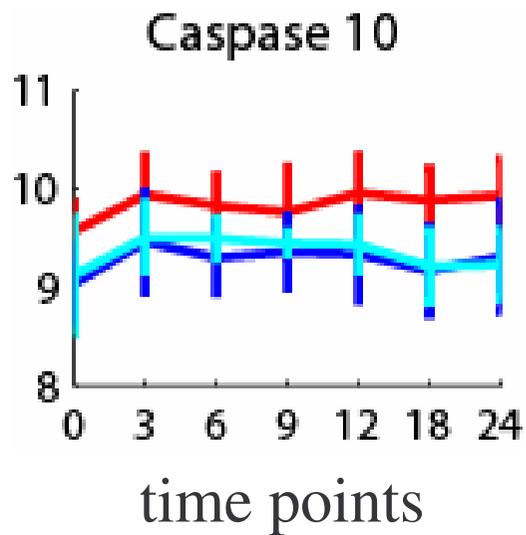
Results - Consensus Matrix

- All 5 x 4-fold classifications – HMM Cons. 3

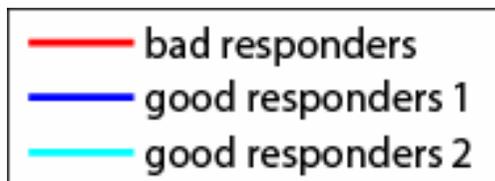
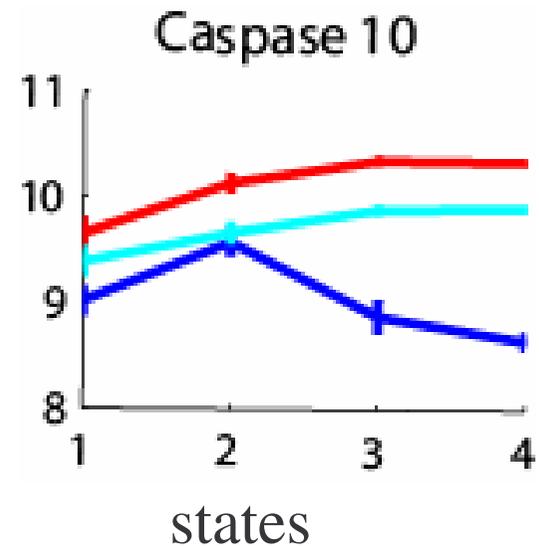


Results –Time Series

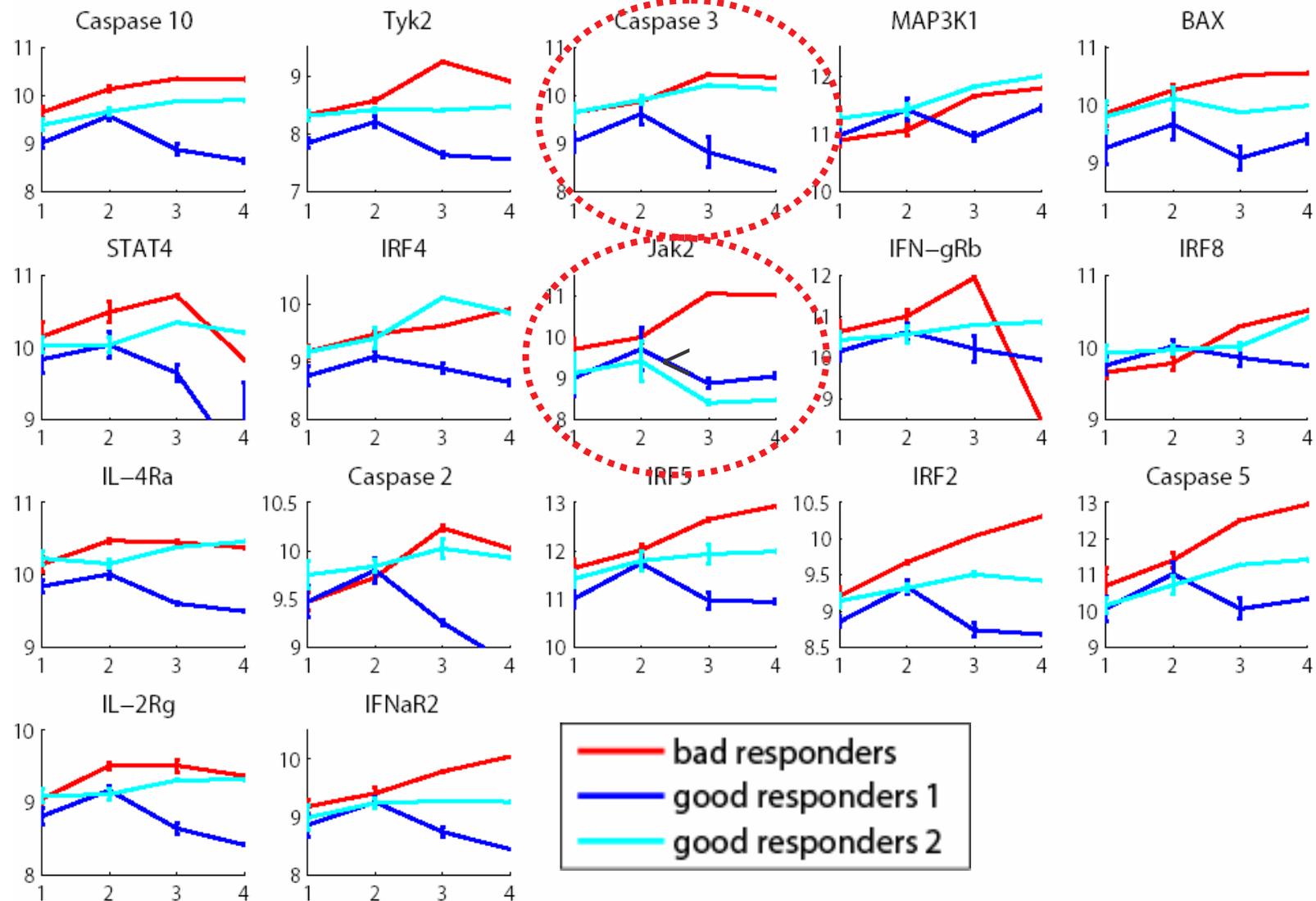
original time series



state aligned series



Results – Selected Features



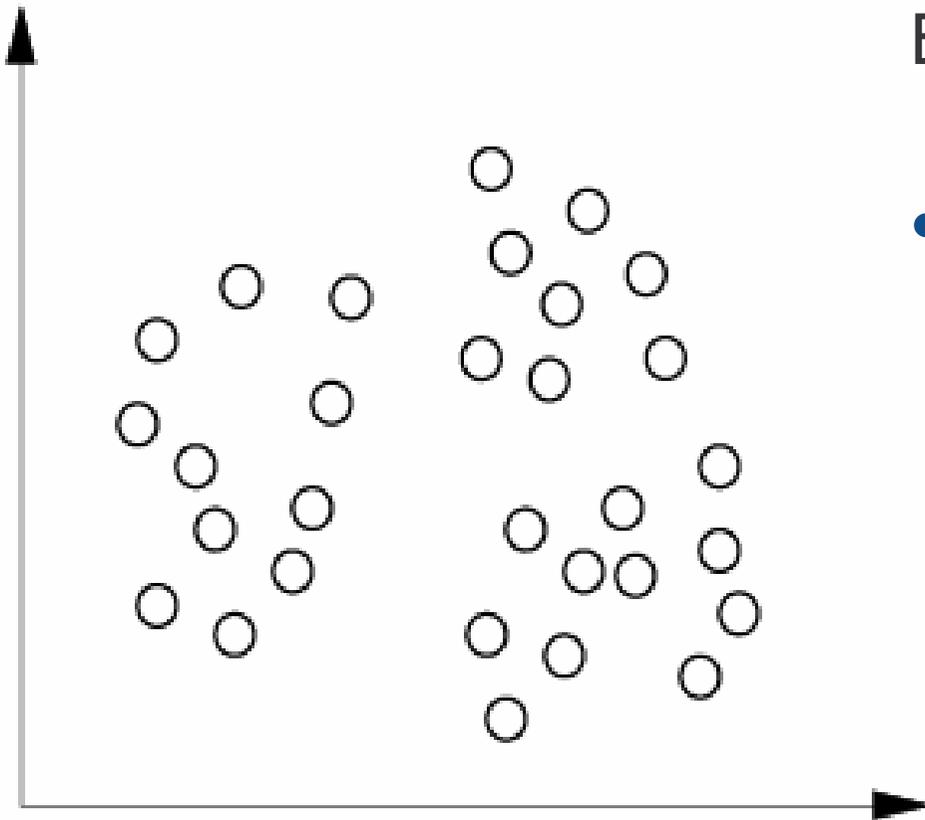
Conclusion

- Increase in classification accuracy
 - robustness to mislabeled examples
 - detection of sub-groups
- MS Treatment Classification
 - mislabeled sample was confirmed
 - sub-groups of good responders are of clinical interest

Acknowledgements

- **Benjamin Georgi**
 - MPI Molecular Genetics
- **Katrin Höfl, Peter van den Elzen**
 - Department of Pathology and Laboratory Medicine,
University of British
- **Sergio Baranzini**
 - Department of Neurology University of California San
Francisco

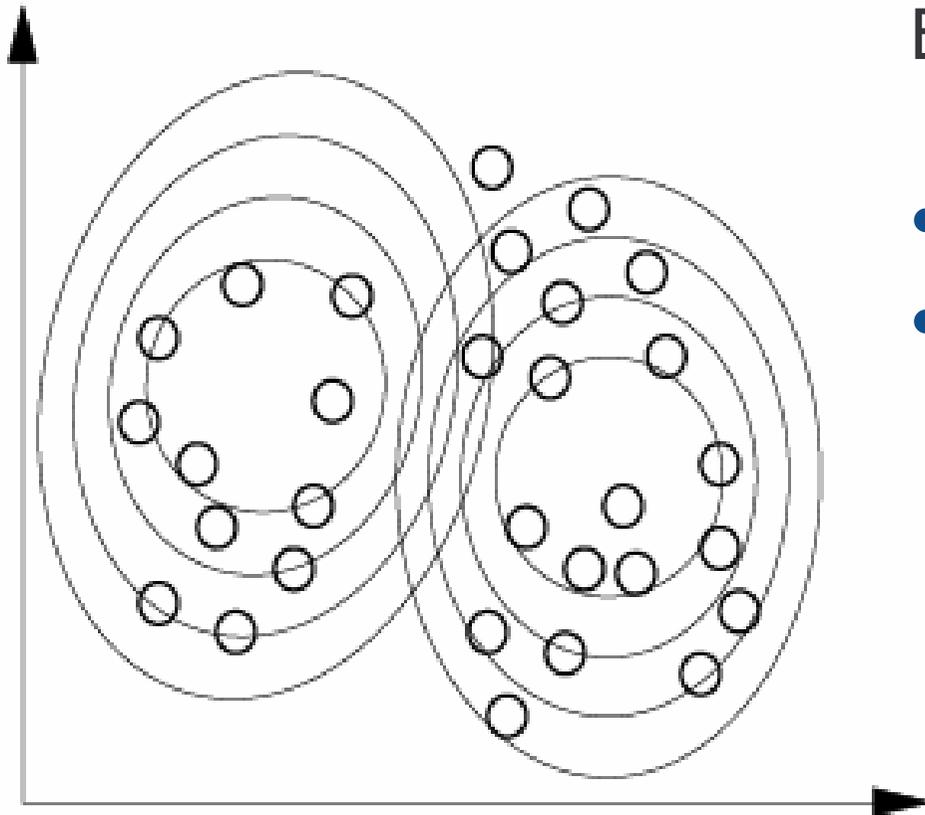
Mixture Estimation (Unsupervised)



Estimate models using
data set X **alone**

- $p(x|\theta_{\text{red}})$ e $p(x|\theta_{\text{blue}})$

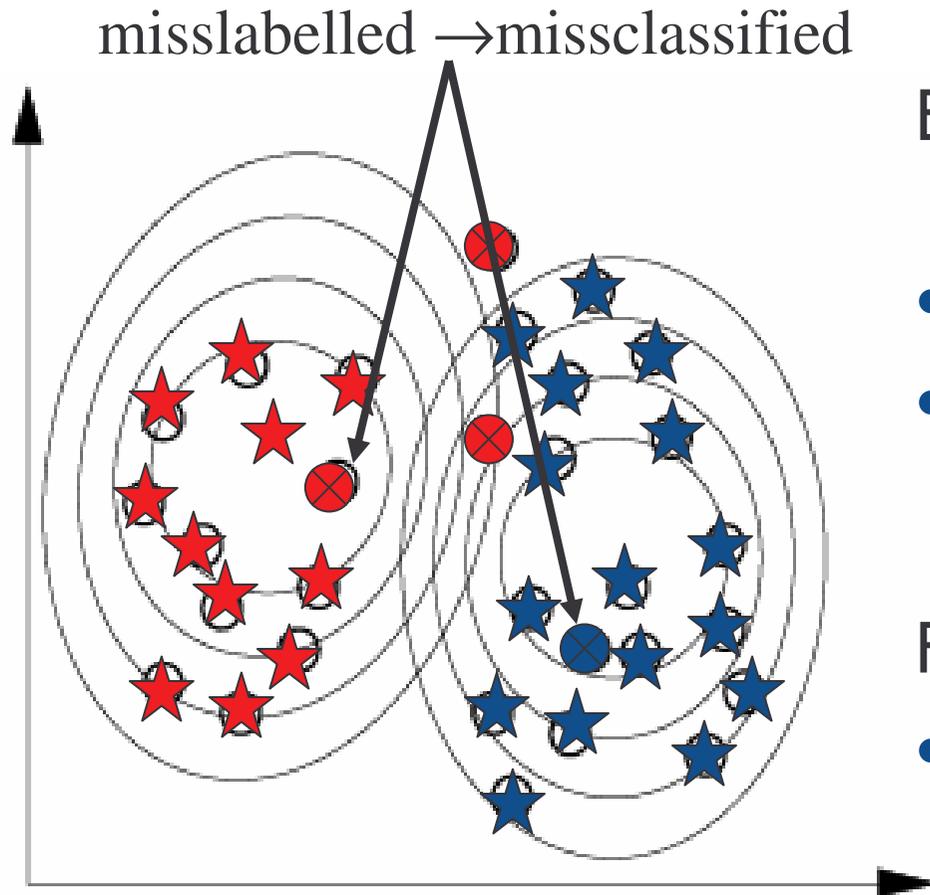
Mixture Estimation (Unsupervised)



Estimate models using
data set X **alone**

- $p(x|\theta_{\text{red}})$ e $p(x|\theta_{\text{blue}})$
- e.g.
 - E.M. algorithm

Mixture Estimation (Unsupervised)



- bad responder
- ☆ rightly classified
- good responder
- ⊗ missclassified
- ...

Estimate models using data set X **alone**

- $p(x|\theta_{\text{red}})$ e $p(x|\theta_{\text{blue}})$
- e.g.
 - E.M. algorithm

For sample x find y

- $p(y_i=\text{red}|x_i) = p(y_i=\text{red}) \cdot p(x_i|\theta_{\text{red}})$

Method – More Details

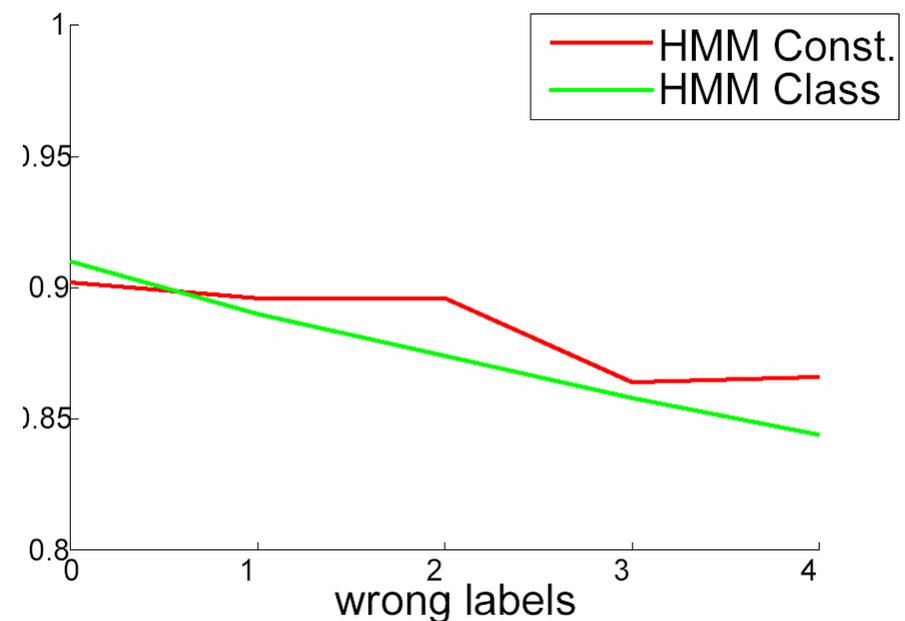
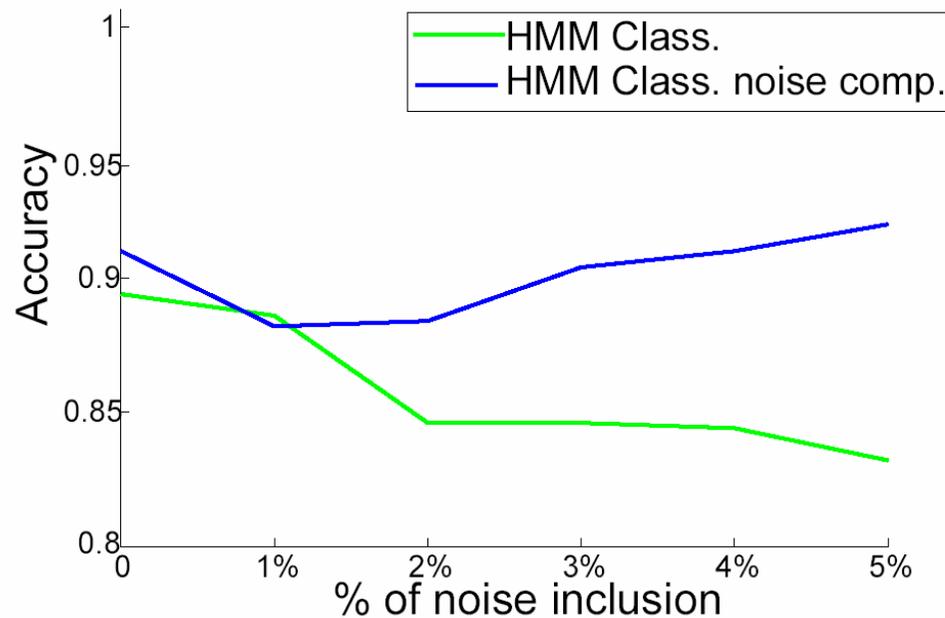
- Feature Selection
 - relative entropy of model class distribution
 - detects sub-group specific features
- HMM Emissions are mixtures
 1. multivariate Gaussians with diagonal covariance matrix
 2. missing data component
 3. noise component
- Consensus Analysis (Monti *et al.* 2004)
 - building co-clustering matrix

Borgwardt, K. M., Vishwanathan, S. V. N., and Kriegel, H.-P. (2006). Class prediction

from time series gene expression profiles using dynamical systems kernel. Pacific

Simulated Data

- Simulated Clinical Time Courses (Lin, et al 2008)



Lin, T. H., Kaminski, N., and Bar-Joseph, Z. (2008). Alignment and classification of time series gene expression in clinical studies. *Bioinformatics*, 24(13), i147–i155.

Classification of INFb Response

Previous Studies

- Use of linear HMM and discriminant learning
 - Discriminant explored in (Lin *et. al*/2008) and (Lin *et. al*/2008)

Lin, T.H. et al. (2008) Alignment and classification of time series gene expression in clinical studies. *Bioinformatics*, 24, i147–i155.

Borgwardt, K.M. et al. (2006) Class prediction from time series gene expression profiles using dynamical systems kernel. *Pac. Symp. Biocomput.*, 11, 547–558.