Bayesian Signal Processing

Will Penny

Also featuring:
Steve Roberts, Richard Everson
Dirk Husmeier, Iead Rezek, Tina Good
and Mike Cassidy

Functional Imaging Lab, UCL.

www.fil.ion.ucl.ac.uk/~wpenny/

Overview

- Bayesian principles, methods and model order selection.
- Applied to

Autoregressive models

Gaussian Mixtures

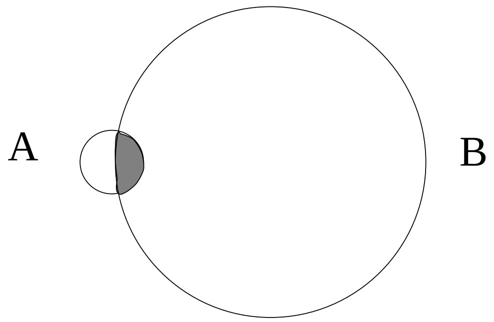
Independent Component Analysis

Neural network classifiers

State-space models

Baye's theorem

- Given $p(A)$, $p(B)$ and $p(A, B)$



- Conditional densities

$$\begin{aligned} p(B|A) &= \frac{p(A,B)}{p(A)} \\ p(A|B) &= \frac{p(A,B)}{p(B)} \end{aligned}$$

- Eliminating $p(A, B)$ gives Baye's rule

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

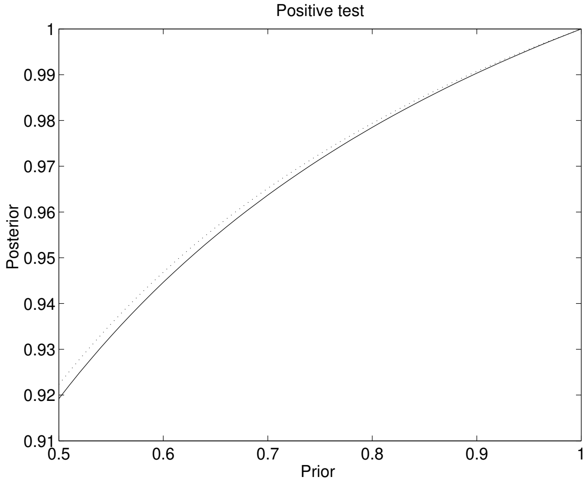
Discrete Bayes

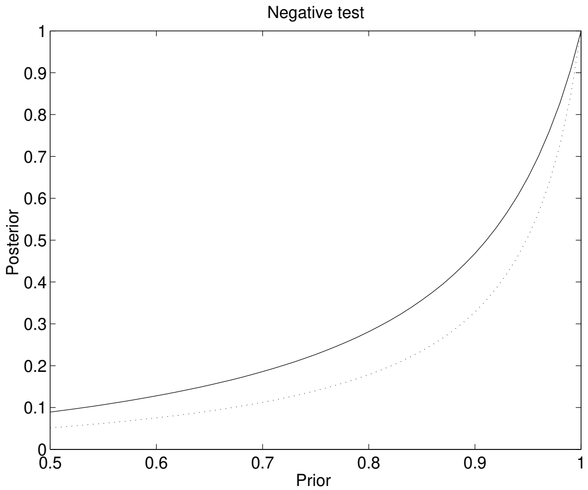
- Disease $D = 0/1$; absent/present
- Test $T = 0/1$; negative/positive
- Sensitivity $p(T = 1|D = 1)$
- Specificity $p(T = 0|D = 0)$
- Bayes rule for a positive test result

$$\begin{aligned} p(D = 1|T = 1) &= \frac{p(T = 1|D = 1)p(D = 1)}{p(T = 1)} \\ &= \frac{Truepos}{Truepos + Falsepos} \end{aligned}$$

Discrete Bayes - Posteriors

- Specificity 92%, sensitivity 91%(–) or 95%(..)

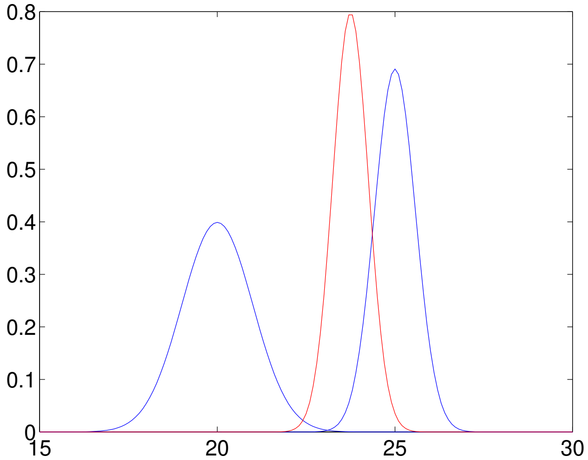




Gaussian Bayes

- Gaussian Prior m_0, β_0

- Gaussian Likelihood m_D, β_D



- Gaussian Posterior

$$\begin{aligned} \beta &= \beta_0 + \beta_D \\ m &= \frac{\beta_0}{\beta} m_0 + \frac{\beta_D}{\beta} m_D \end{aligned}$$

- Sensor fusion

GLM Bayes

- General Linear Model (GLM)

$$Y = Xw + e$$

  eg. Wavelets, Autoregressive models, SPM

- Gauss. Prior $p(w)$ mean 0, precision β_0
- Gauss. Like. $p(e)$ mean 0, precision β_D
- Gaussian Posterior $p(w|D)$

$$\begin{aligned}\Sigma^{-1} &= \beta_D (X^T X)^{-1} + \beta_0 I \\ \hat{w} &= \Sigma X^T \beta_D Y\end{aligned}$$

GLM Empirical Bayes

Given N data points and p coefficients:

1. Estimate posterior mean and covariance
2. Estimate number of well-determined coefficients

$$\gamma = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + \beta_0}$$

3. Estimate prior precision

$$\frac{1}{\beta_0} = \frac{1}{\gamma} \sum_{i=1}^{p} w_i^2$$

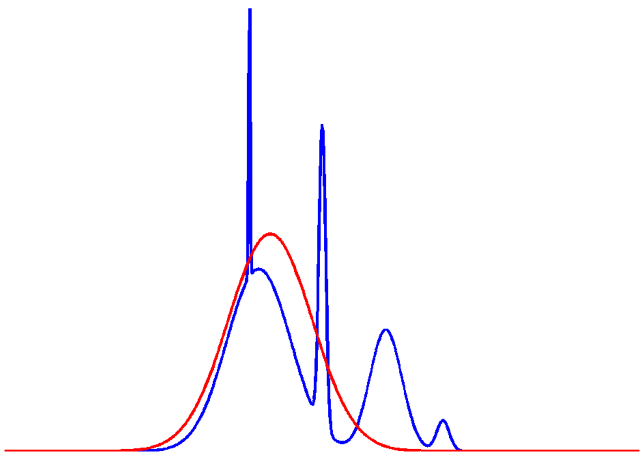
4. Estimate data precision

$$\frac{1}{\beta_D} = \frac{1}{N - \gamma} (Y - Xw)^T (Y - Xw)$$

Variational Bayes

Given data D and model parameters θ_i:

- True posterior $p(\theta|D)$
- Approximate posterior $q(\theta|D)$



- Minimise (plot by Zoubin G.)

$$KL = \int q(\theta|D) \ln \frac{q(\theta|D)}{p(\theta|D)} d\theta$$

GLM - Variational Bayes

- Given the likelihood and priors we can compute the optimal form of $q(\theta|D)$ - via the calculus of variations
- Factorising the posterior

$$q(\theta|D) = \prod_i q(\theta_i|D)$$

and choosing suitable priors helps to make this possible.

- For GLMs, a Gaussian prior on w and Gamma priors on the precisions leads to Gaussian and Gamma posteriors.

Model order selection

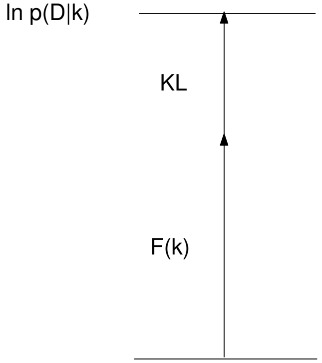
- Given model order k

$$p(k|D) = \frac{p(D|k)p(k)}{p(D)}$$

- For uniform $p(k)$ pick maximum

$$p(D|k) = \int p(D|\theta, k)p(\theta|k)d\theta$$

- Variational Bayes

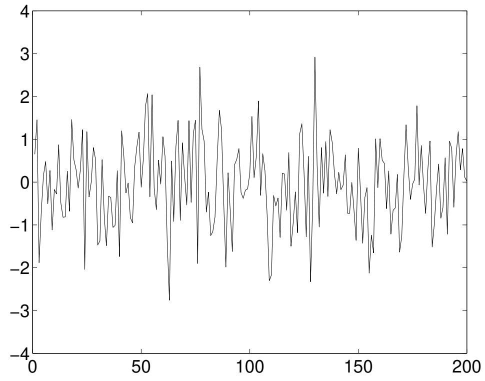


- Empirical Bayes

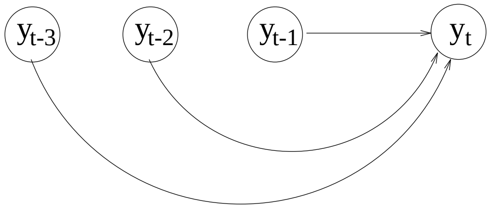
- Minimum Description Length (MDL)

AR Models

- Models a sample from a time series as a linear combination of previous samples



- Generative Model



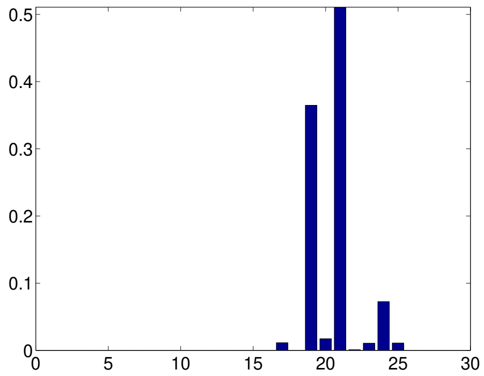
- Probability

$$p(y_t) = \mathsf{N}\left(\sum_{k=1}^{K} a_k y_{t-k}, \sigma^2\right)$$

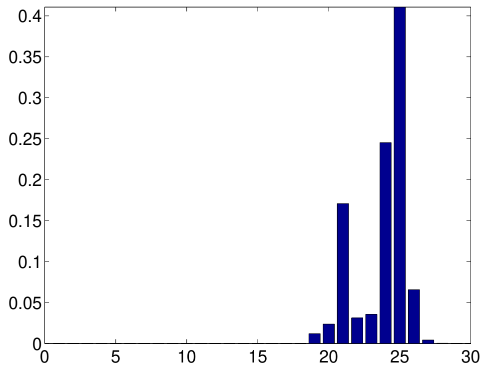
AR model-order

For true $k = 25$, 3 seconds of data

- MDL



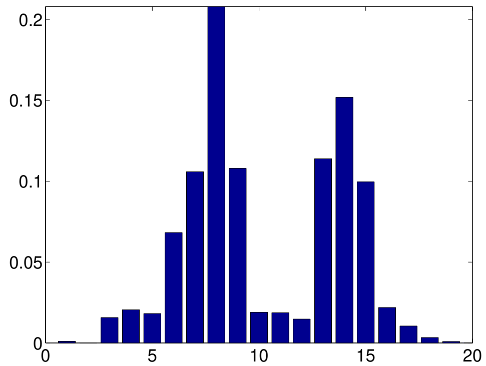
- VB



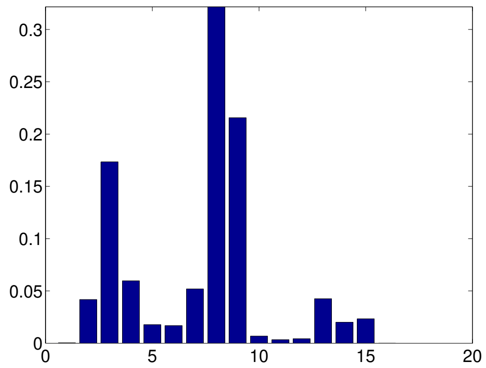
EEG AR model-order

Using VB-AR on 1-second blocks of EEG

- Awake



- Anaesthetised



Multivariate AR

- AR for multiple time-series

$$y_t = \sum_{k=1}^{K} y_{t-k} a_k + e_t$$

- AR coefficients are matrices

$$a_1 = \begin{pmatrix} 0.40 & 0.30 \\ 1.20 & 0.70 \end{pmatrix}$$

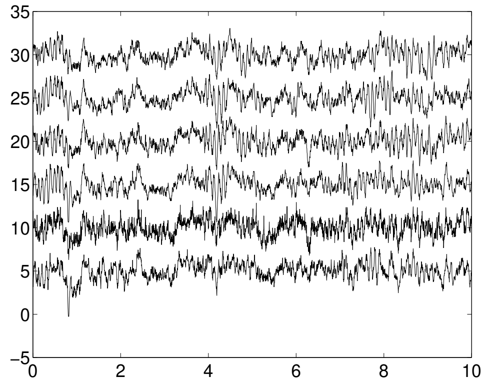
$$a_2 = \begin{pmatrix} 0.35 & -0.40 \\ -0.30 & -0.50 \end{pmatrix}$$

- Coherence and phase spectra available

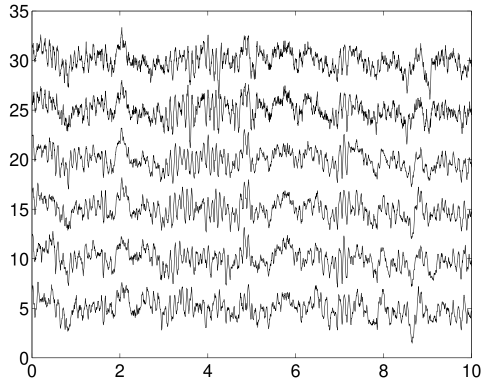
Multivariate EEG data

EEG recorded during

- Baseline



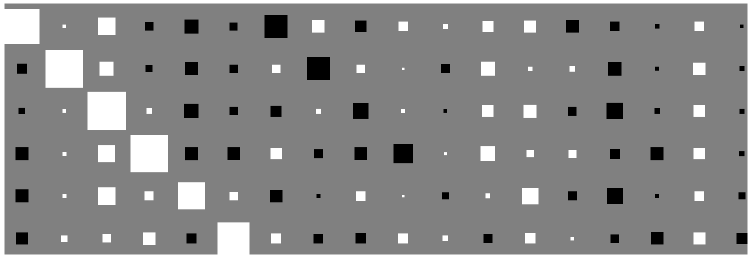
- Maths task



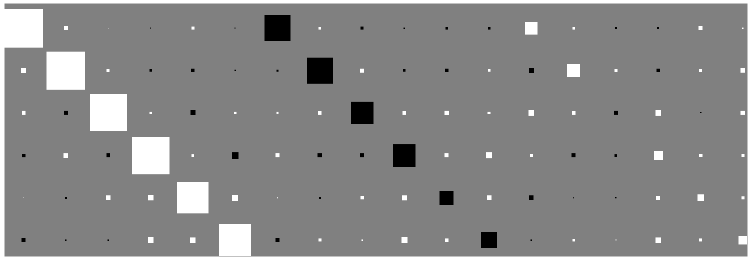
Multivariate AR - Priors

Estimated AR coefficients

- Global Prior



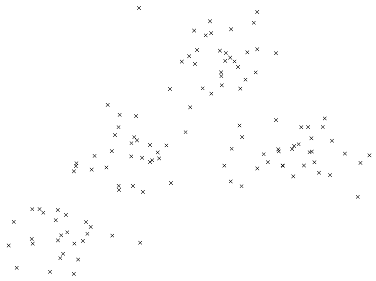
- Interaction Prior



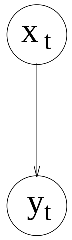
Prior	Evidence
Interaction	8.0 $\pm$ 5.0
Lag	6.2 $\pm$ 4.0
Lag-Interaction	-1.6 $\pm$ 3.4
Global	-12.8 $\pm$ 5.0

Gaussian Mixtures

- Models multivariate data as a set of Gaussian clusters



- Generative Model

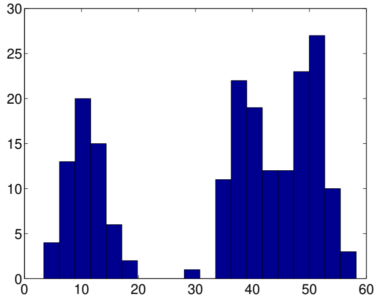


- Probability

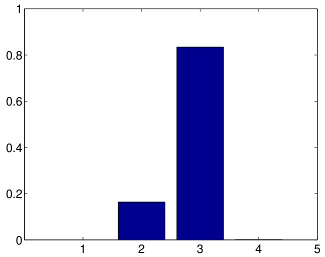
$$
\begin{aligned}
p(x_t = k) &= \pi_k \\
p(y_t) &= \sum_{k=1}^{K} p(x_t = k) \mathrm{N}(\mu_k, \Sigma_k)
\end{aligned}
$$

Gaussian Mixtures - Model order

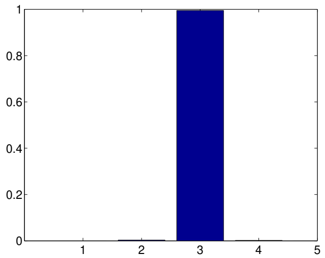
Two-hundred data points drawn from 3-clusters



- MDL model order



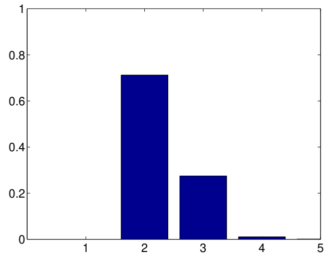
- VB model order



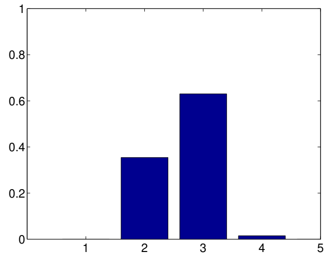
Gaussian Mixtures - Model order

With only 50 points from the same 3-clusters

- MDL model order



- VB model order



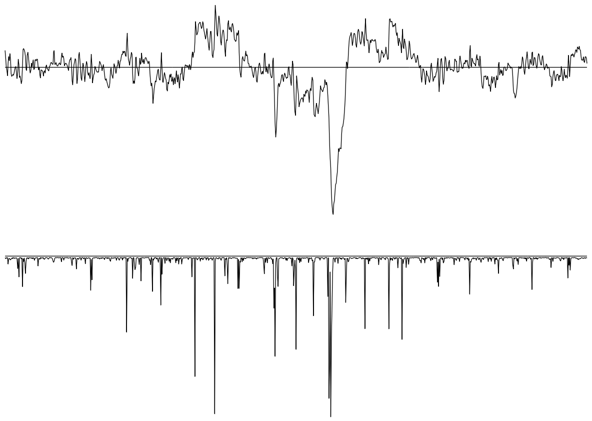
AR-GMM model

- Model noise in AR model with zero mean GMM

- Bayesian weighted least squares

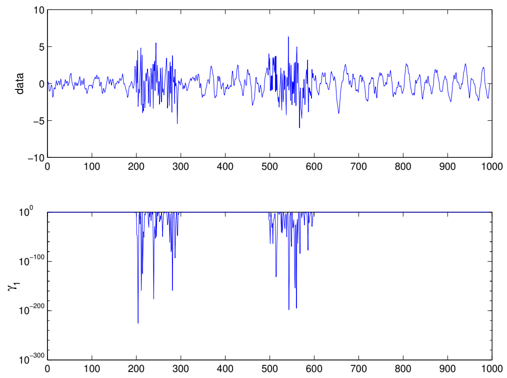
- Full VB implementation

- Robustness to artifacts

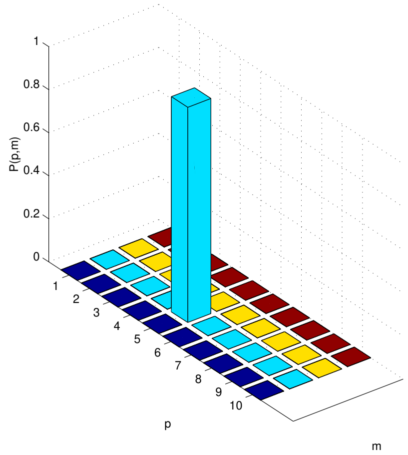


AR-GMM: Music

- On Music data



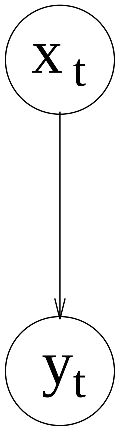
- Model order selection



ICA

- Models multivariate data as a linear combination of independent sources

- Generative Model



- Probability

$$
\begin{aligned}
p(x_t) &= \prod_{k}^{K} p(x_t^k) \\
p(y_t) &= A x_t
\end{aligned}
$$

ICA - EEG

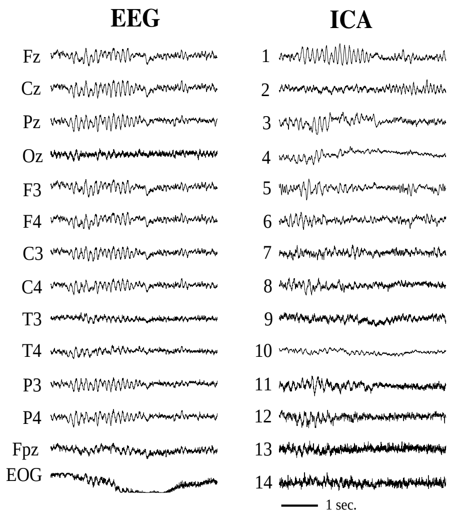
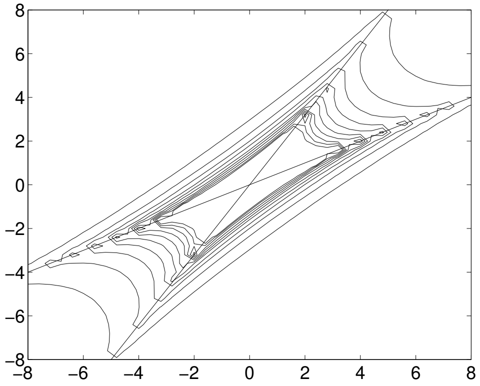


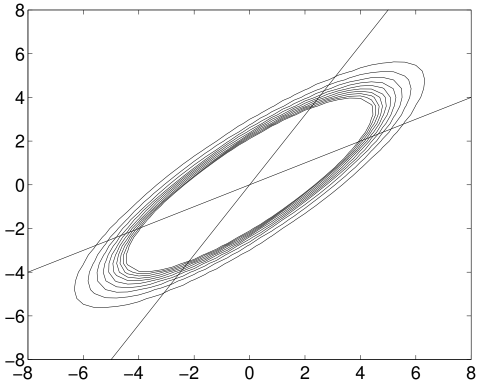
Figure 1: Left: 4.5 seconds of 14-channel EEG data. Right: an ICA transform of the same data, using weights trained on 6.5 minutes of similar data from the same session.

ICA - 2D

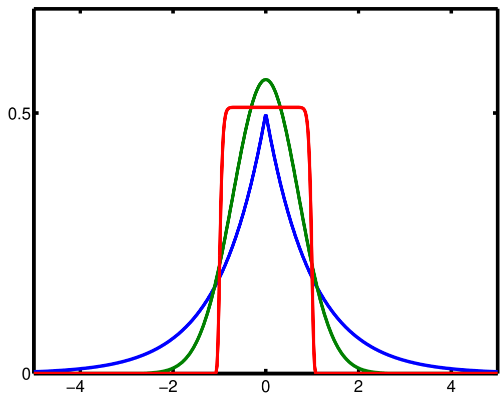
- Heavy-tailed density



- Gaussian



Generalised Exponentials

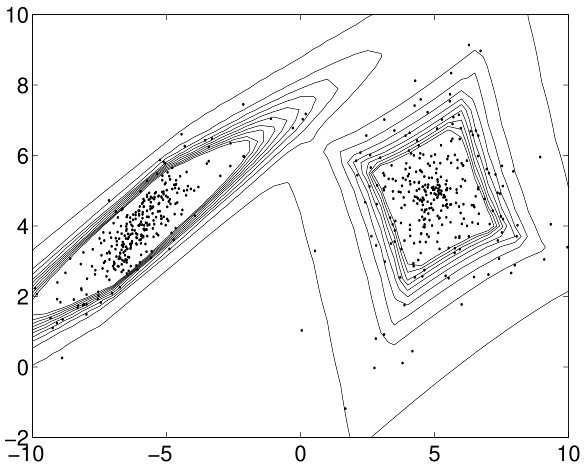


- Width parameter w and Shape parameter R eg. $R = 1$ for a Laplacian , $R = 2$ for a Gaussian and $R \geq 20$ for a Uniform density.

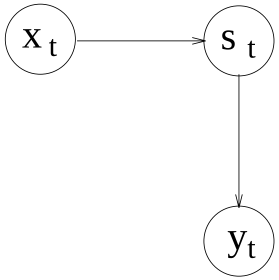
- Width parameter w and Shape parameter R eg. $R = 1$ for a Laplacian , $R = 2$ for a Gaussian and $R \geq 20$ for a Uniform density.

ICA Mixture model

- Non-Gaussian cluster analysis



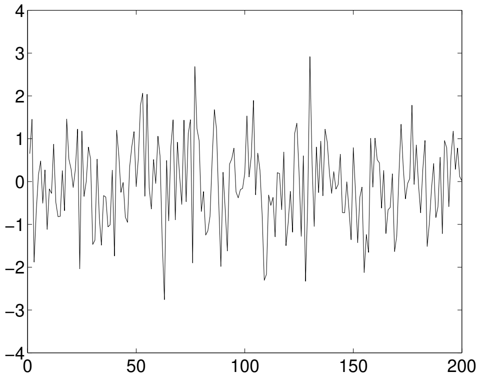
- Generative model



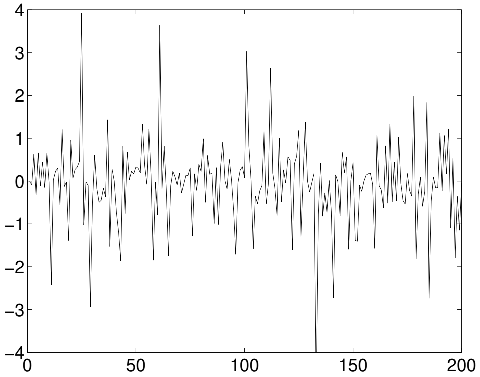
- High-dimensional data

GAR sources

- Generalised AR (GAR) model with R=2



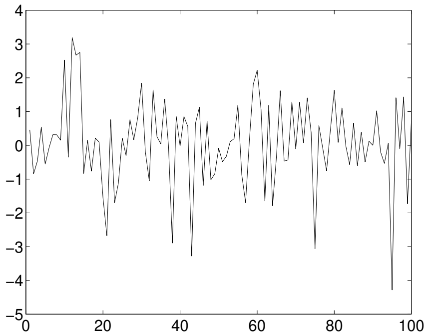
- With R=1



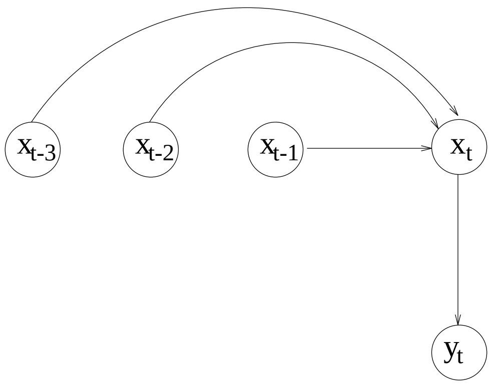
- Better suited to biomedical data

ICA-GAR model

- ICA with temporally correlated sources; EEG, Music



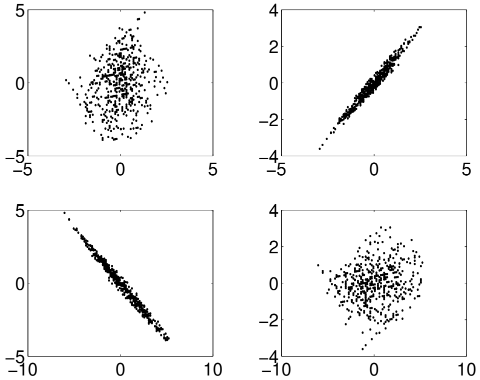
- Generative model



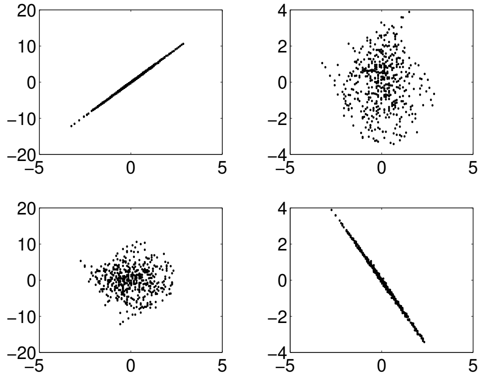
- MDL model order selection criterion

ICA-GAR unmixing

- Inverse-Cosh sources



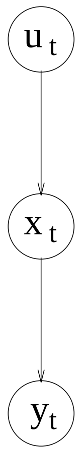
- GAR sources



Multilayer Perceptron Classifier

- Classifies multivariate data points as belonging to one of two classes

- Generative Model

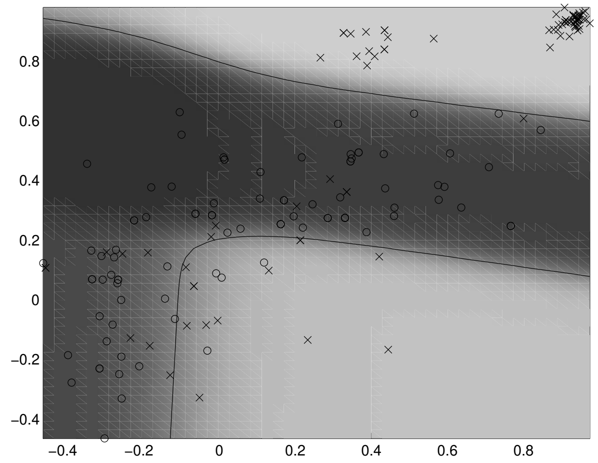


- Probability

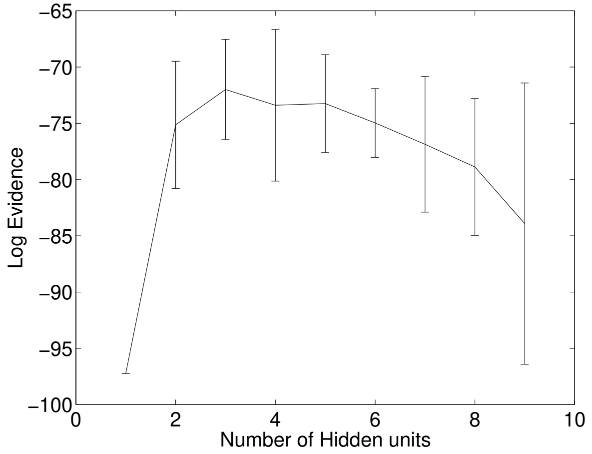
$$
\begin{aligned}
x_t^k &= \sigma(W_1^k u_t) \\
p(y_t = 1) &= \sigma(W_2 x_t)
\end{aligned}
$$

MLP Empirical Bayes

- Classifying EMG Tremor data

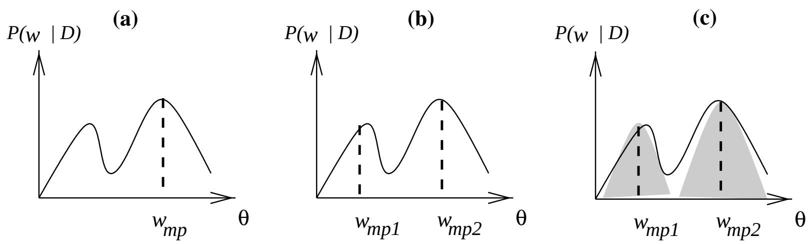


- Model order selection



MLP Empirical Bayes

- GMM approximation to posterior



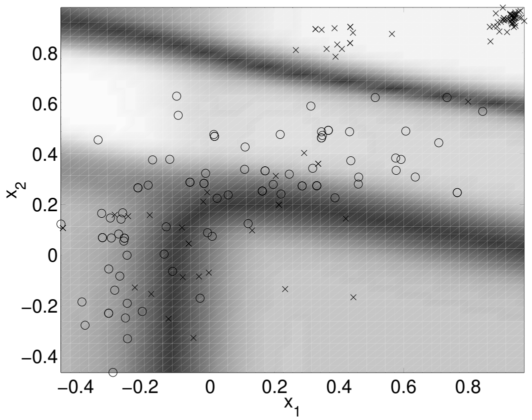
- 'Moderate' classification

$$p(y = 1|D) \;=\; \int p(y = 1|w)p(w|D)dw$$

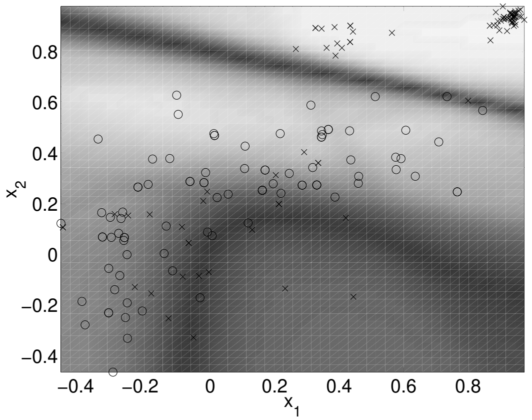
by averaging over the posterior density.

MLP Confidence

- Confidence of unmoderated network



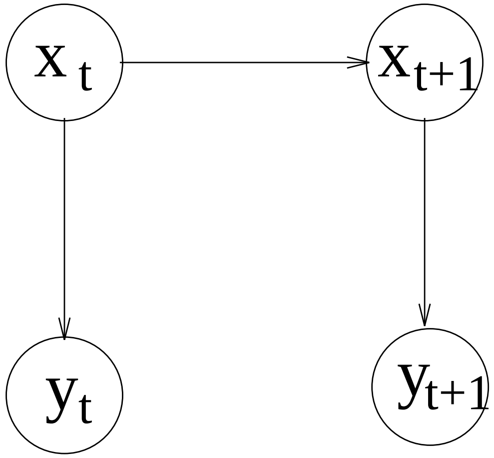
- Confidence of moderated network



Linear Dynamical Systems

- Models multivariate time-varying observations as a linear function of a linearly evolving hidden variable

- Generative Model



- Probability

$$
\begin{aligned}
p(x_{t+1}) &= \mathrm{N}(Ax_t, Q) \\
p(y_t) &= \mathrm{N}(Cx_t, R)
\end{aligned}
$$

Kalman filters etc.

- For a linear dynamical system

$$
\begin{aligned}
p(x_{t+1}) &= \mathrm{N}(Ax_t, Q) \\
p(y_t) &= \mathrm{N}(Cx_t, R)
\end{aligned}
$$

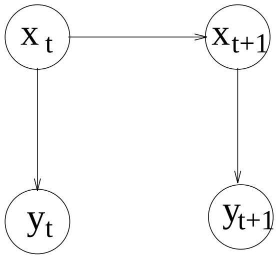
- If A, Q, C and R are known estimation of next state given previous time-series by Bayes rule produces the Kalman filter . Using future values also gives the Kalman smoother

- Fix A,C and R and infer Q online

Hidden Markov Models

- Models multivariate time-varying observations as a discrete switching between stationary processes

- Generative Model



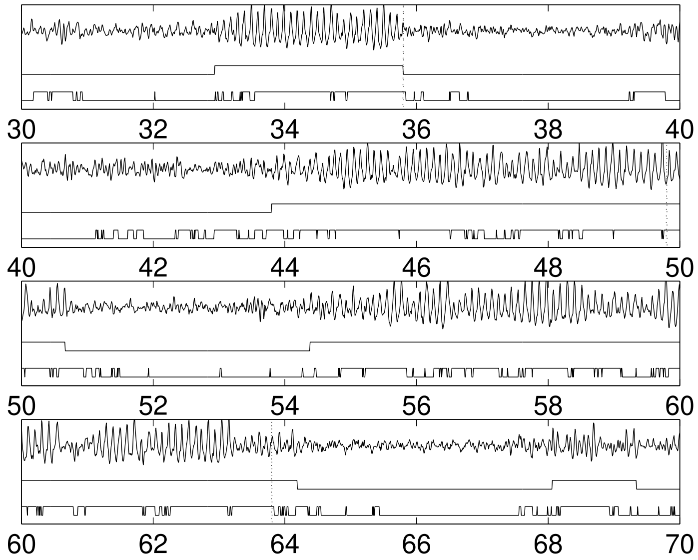
- Composed of

State Transition Model

Observation Model

HMM-AR: Mu-Rhythm

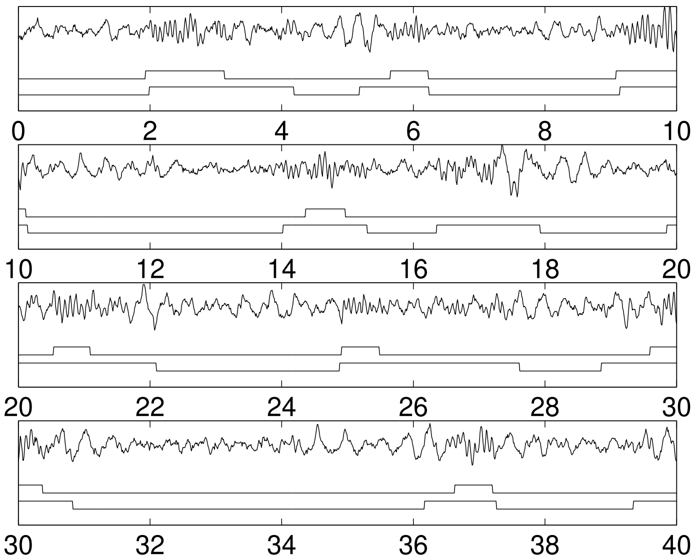
- Cued hand-movement EEG data



- Comparison with mixture-model - HMMs classify sequences

HMM-AR: Sleep-spindles

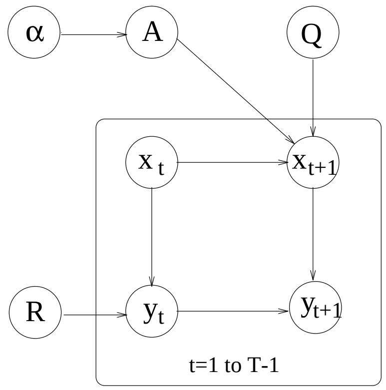
- Detecting sleep spindles



- Comparison with human sleep expert

Nonstationary MAR

- Multivariate autoregressive model with time-varying coefficients



- Probability

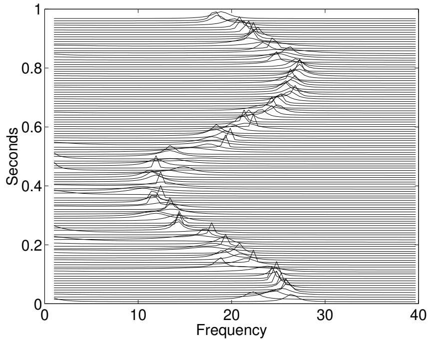
$$
\begin{aligned}
p(x_{t+1}) &= \mathrm{N}(Ax_t, Q) \\
p(y_t) &= \mathrm{N}(M_t x_t, R)
\end{aligned}
$$

- Full VB implementation - Mike Cassidy

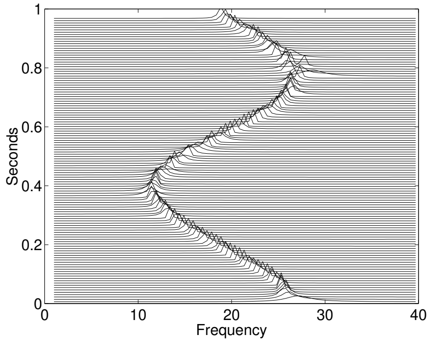
Nonstationary MAR

- NMAR(4) applied to clean frequency modulated signal.

- Inference using Maximum Likelihood:



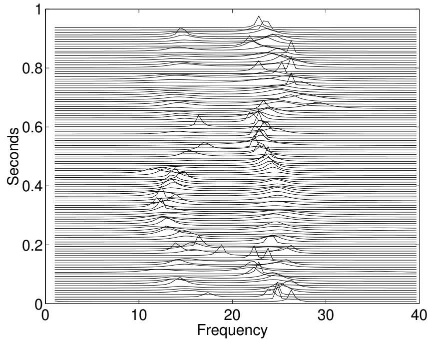
- Inference using Variational Bayes:



Nonstationary MAR

- NMAR(8) applied to noisy frequency modulated signal.

- Inference using Maximum Likelihood:



- Inference using Variational Bayes: