NEURAL DATA ANALYSIS

ALEXANDER ECKER, PHILIPP BERENS, MATTHIAS BETHGE

COMPUTATIONAL VISION AND NEUROSCIENCE GROUP

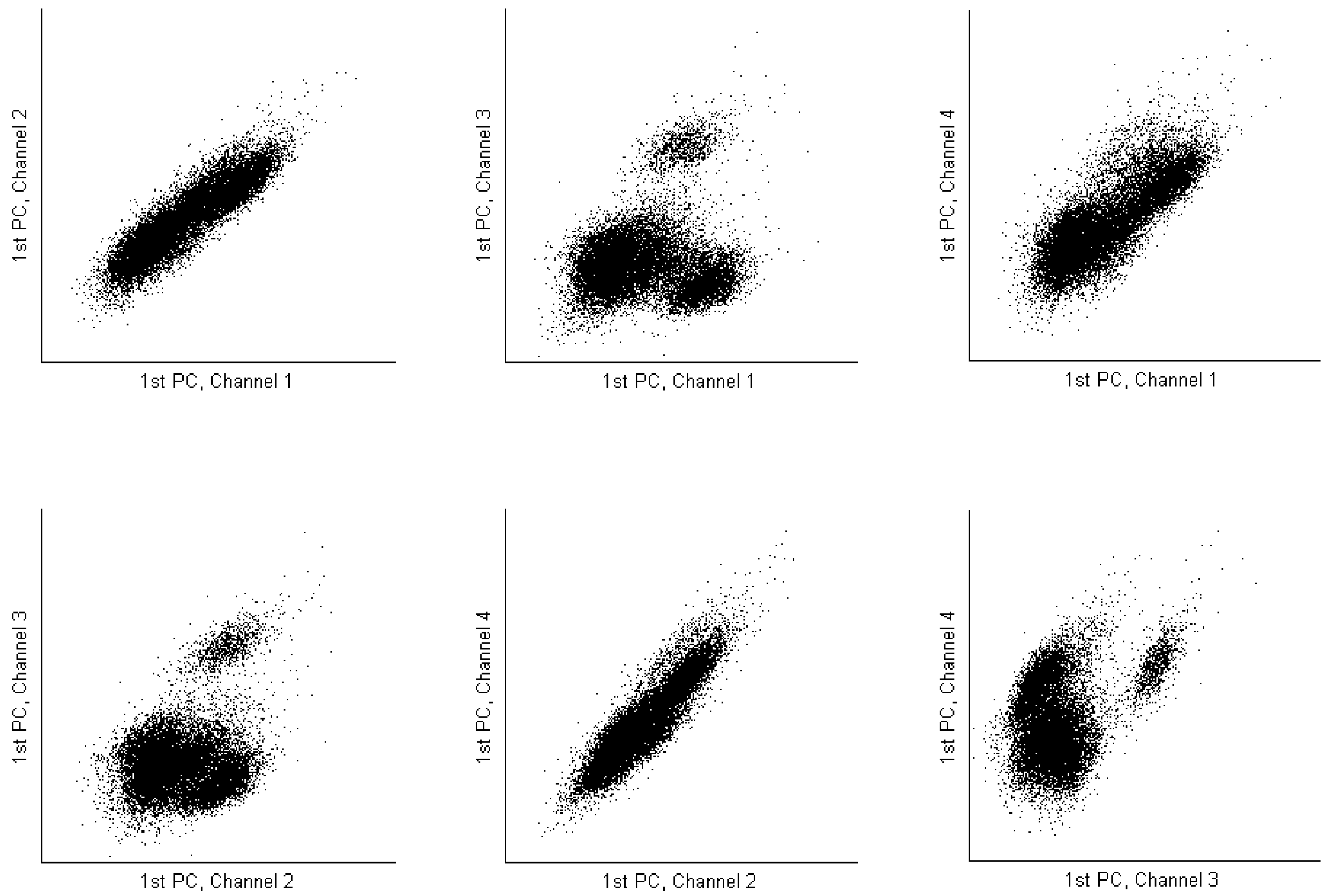
Week	Date	Topic
1	14.04.15	Introduction, Spike detection, Feature extraction
2	21.04.15	Spike sorting with Mixture of Gaussians
3	28.04.15	Identifying single neurons (Cross-correlograms, LDA)
4	05.05.15	Spike inference from calcium data (Deconvolution)
5	**12.05.15**	**Current problems (presentations 1)**
6	19.05.15	Visualizing and analyzing spike trains 1 (Raster plot, PSTH, Latency)
7	02.06.15	Visualizing and analyzing spike trains 2 (Tuning curves, Correlations)
8	09.06.15	Single cell receptive field estimation (STA, STC, LNP-model)
9	16.06.15	Analyzing neural populations 1 (Generalized linear models, Ising models)
10	23.06.15	Analyzing neural populations 2 (Population dynamics, GPFA)
11	**30.06.15**	**Current problems (presentations 2)**
12	07.07.15	Analyzing neural populations 3 (Decoding)
13	14.07.15	Backup (tba)
14	21.07.15	Wrap-up, Discussion

TASK 1A:
SPIKE DETECTION

ISSUES

- **Dealing with four channels**
- **Alignment**
- **Double triggering**

TASK 1B:
FEATURE EXTRACTION


1st PC, Channel 2
1st PC, Channel 1
1st PC, Channel 3
1st PC, Channel 1
1st PC, Channel 4
1st PC, Channel 1
1st PC, Channel 3
1st PC, Channel 2
1st PC, Channel 4
1st PC, Channel 2
1st PC, Channel 4
1st PC, Channel 3

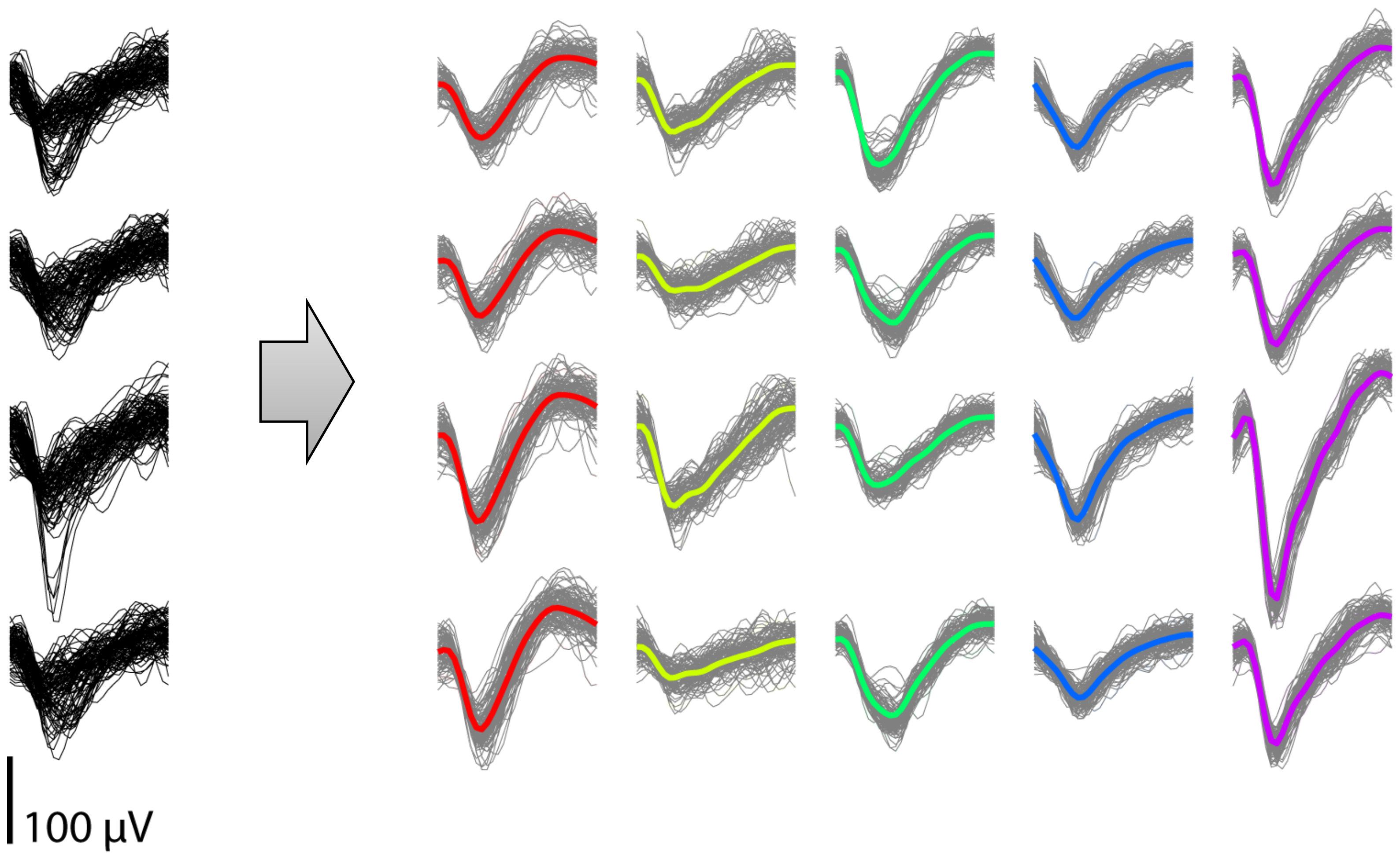
ISSUES

- **Why not perform PCA on all 4 channels at same time?**

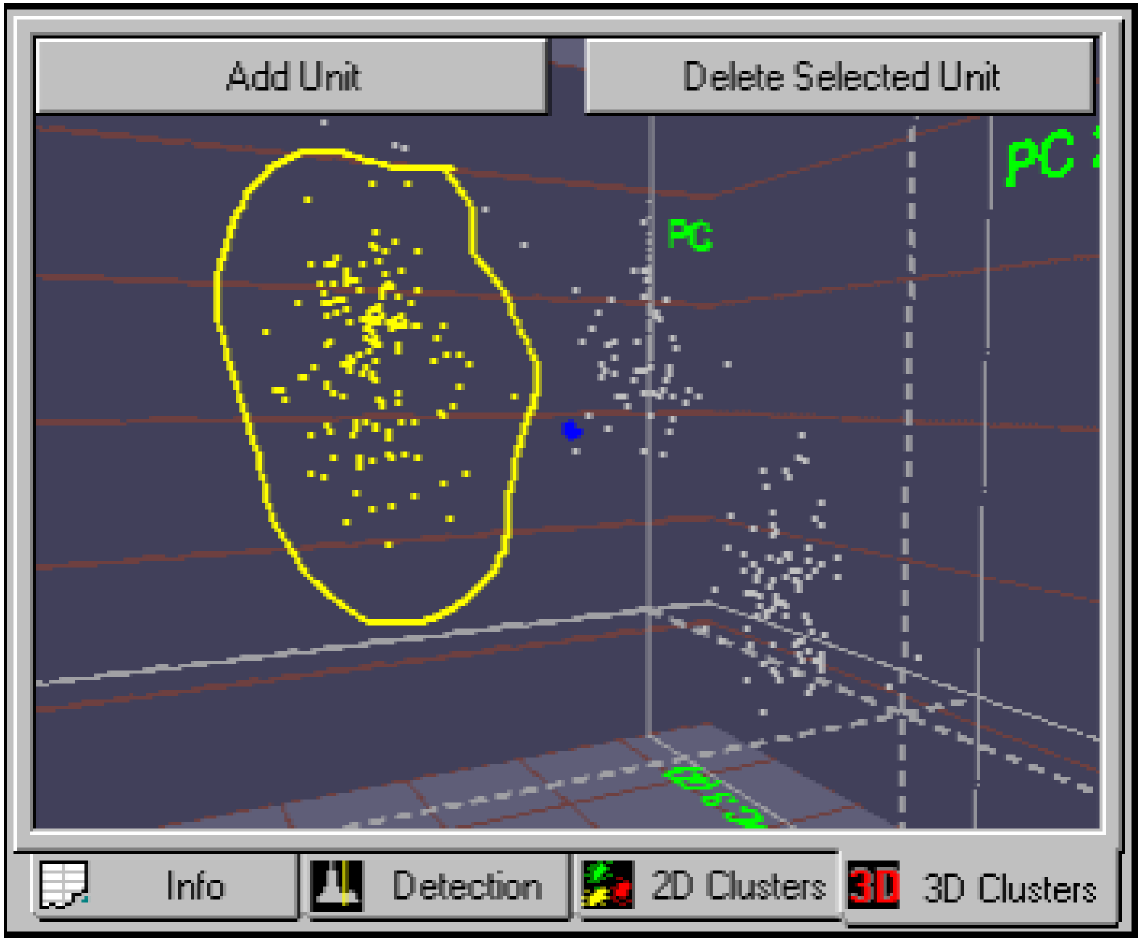
  Relative amplitudes are important
  In simultaneous PCA, they get lost since they are not the largest source of variance

TASK 2:
SPIKE SORTING

SPIKE SORTING



MANUAL CLUSTER CUTTING



PROBLEMS

Subjective

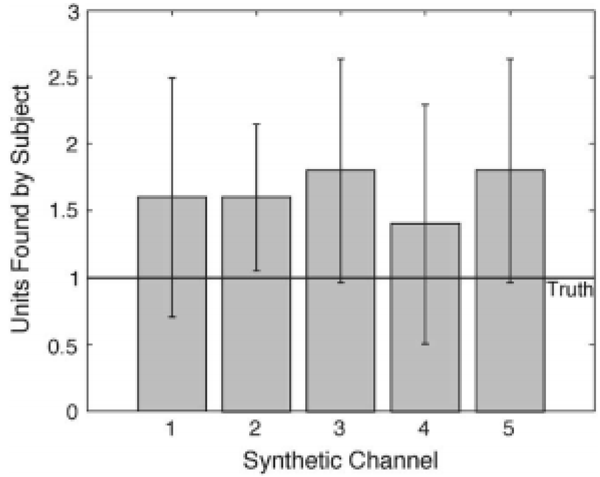
High error rates

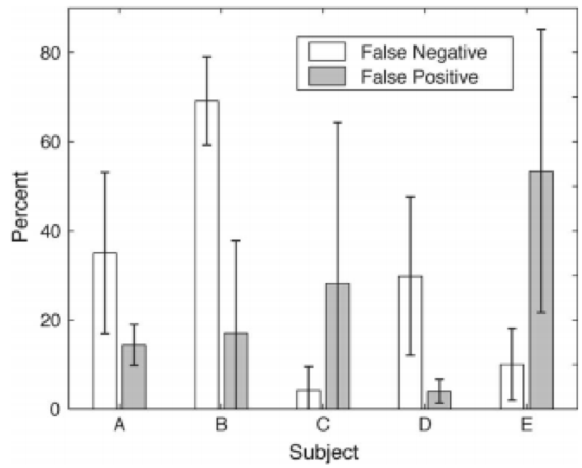
Suboptimal boundaries

Time consuming

Not reproducible

Not model based





Wood et al. 2004

AUTOMATIC CLUSTERING

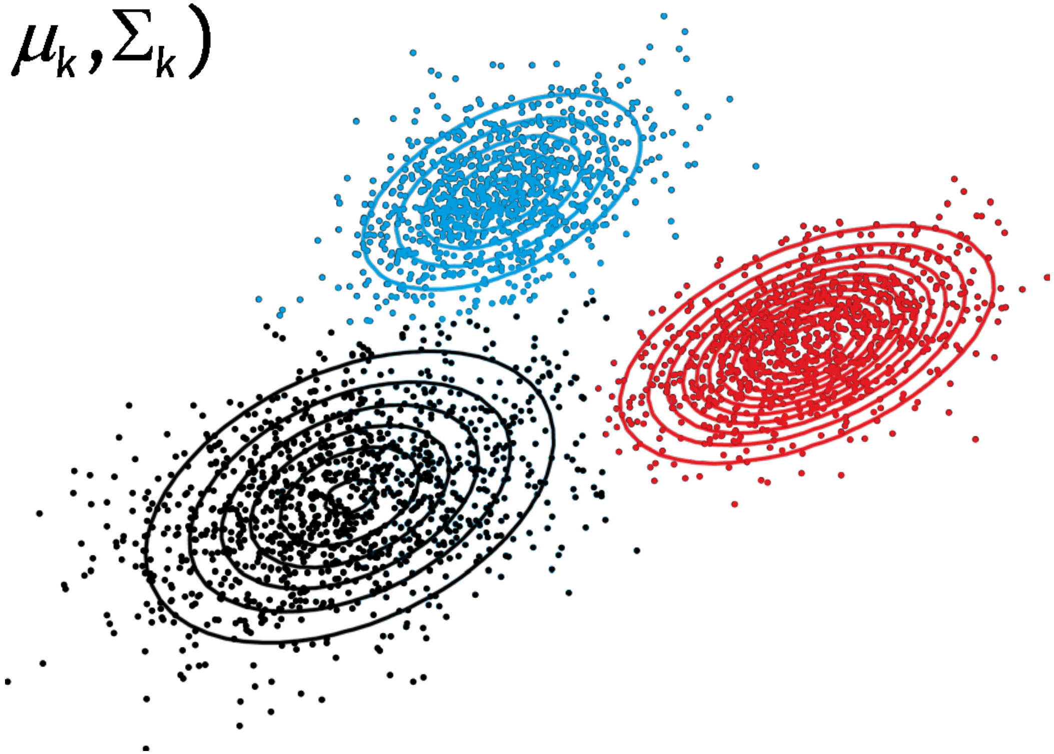
- **Objective**
- **Model-based**
- **Reproducible**
- **Quantifiable**

- **K-Means**
- **Mixture of Gaussian**
- **Others?**

WHICH ALGORITHM?

Fit a Gaussian mixture model

$$p(x) = \sum_k \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)$$



LOG LIKELIHOOD

$$\ln p(x|\mu, \Sigma, \pi) = \sum_n \ln \sum_k \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)$$

π_k	Mixing coefficient of cluster k
μ_k	Mean of cluster k
Σ_k	Covariance of cluster k

EXPECTATION MAXIMIZATION

Finds ML estimate of latent variable model

Parameters: θ
Latent variables: Z
Data: X

Alternate until convergence:

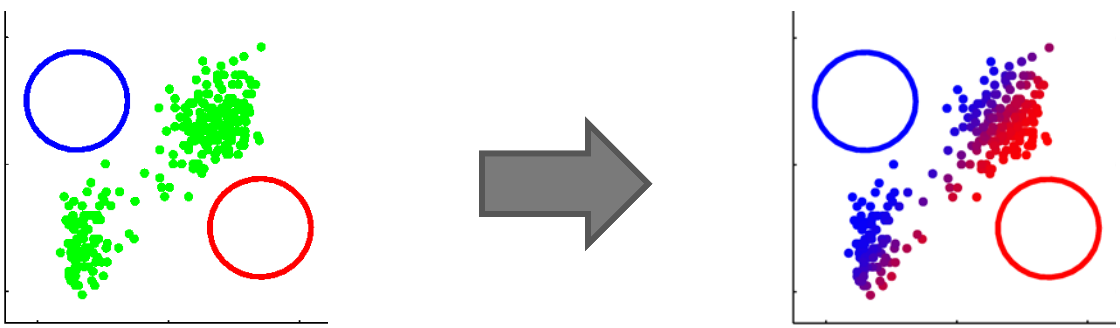
1. **E step: estimate latent variables Z given the current set of parameters θ**
2. **M step: update parameters θ to maximize the likelihood given current estimate of the latent variables Z**

E-STEP

Evaluate ‚responsibilities‘ with current parameter values:

(posterior probability of a data point to belong to cluster k)

$$\gamma_{k,n}^{new} = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}$$



M-STEP

Update means:

$$\mu_k^{new} = \frac{1}{N_k} \sum_n \gamma_{k,n} x_n$$

Update covariances:

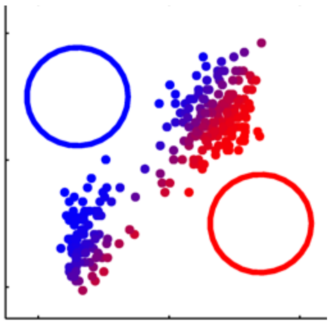
$$\Sigma_k^{new} = \frac{1}{N_k} \sum_n \gamma_{k,n} (x_n - \mu_k^{new})(x_n - \mu_k^{new})'$$

Update mixing coefficients:

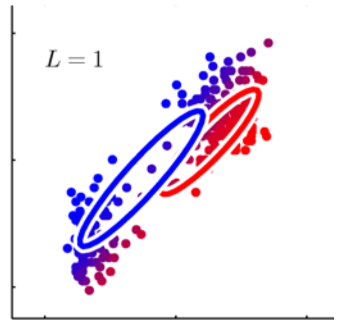
$$\pi_k^{new} = \frac{N_k}{N}$$

$$N_k = \sum_n \gamma_{k,n}$$

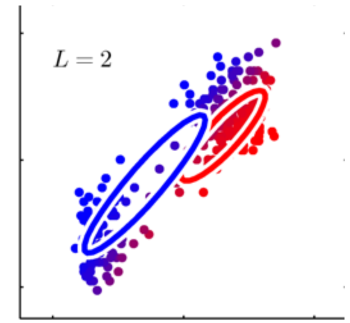
M-STEP



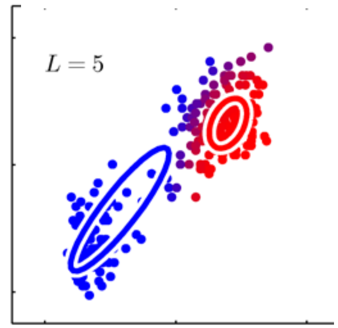
Initialization



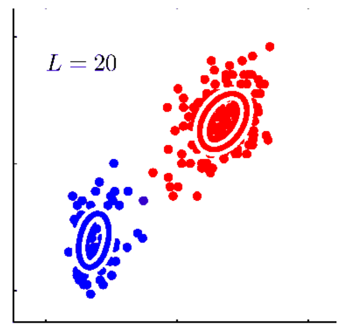
Step 1



Step 2

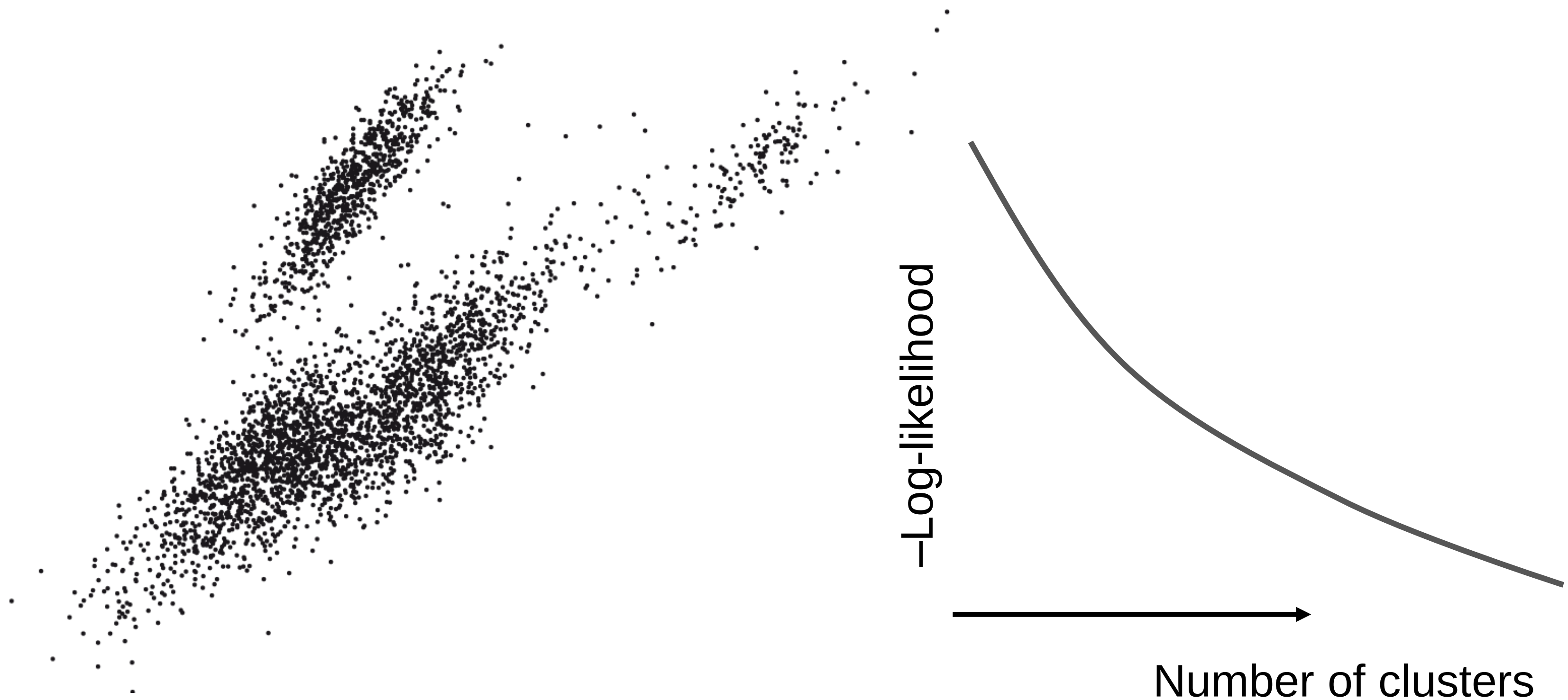


Step 5



Step 20

HOW DO WE DETERMINE THE CLUSTER NUMBER?

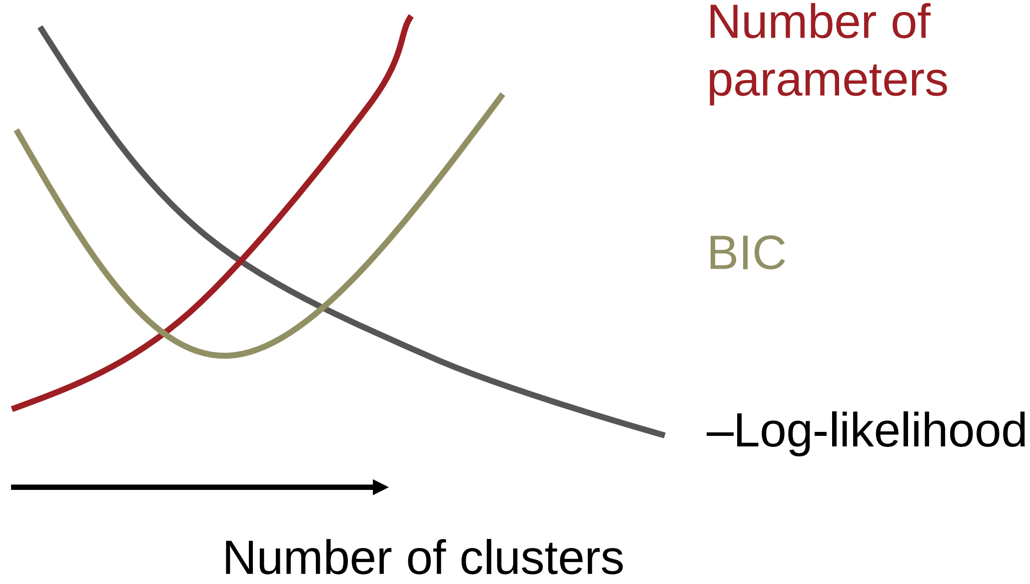


DEALING WITH MODEL COMPLEXITY

Penalize number of parameters!

Use information criteria like BIC (or modified versions)

$$BIC = -2 \ln p(x | \mu_{opt}, \Sigma_{\text{opt}}, \pi_{opt}) + P \ln N$$

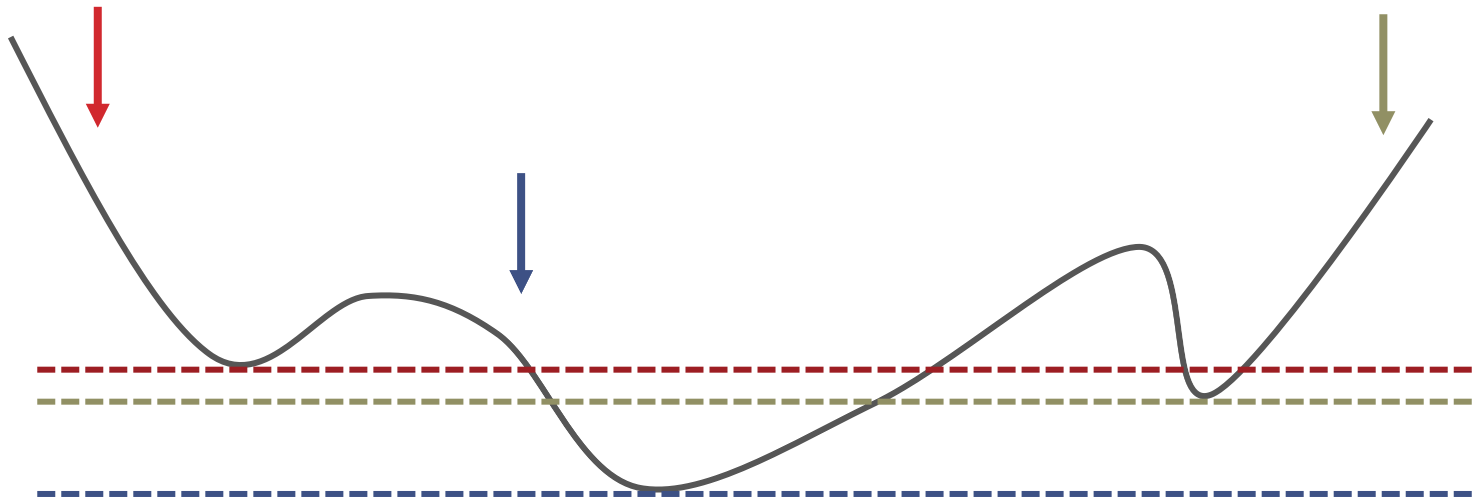


DEALING WITH LOCAL MAXIMA

EM can get stuck in local maxima. What do we do?

1. Restart the algorithm multiple times and pick the result with the best likelihood.
2. Split-and-merge

RESTARTING TO AVOID LOCAL MAXIMA



SPLIT AND MERGE

After EM has converged, try alternately

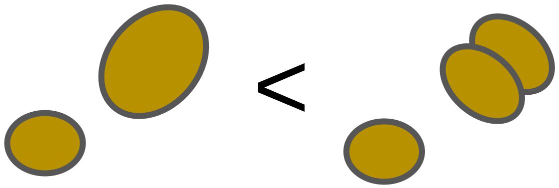
- **Splitting a cluster**

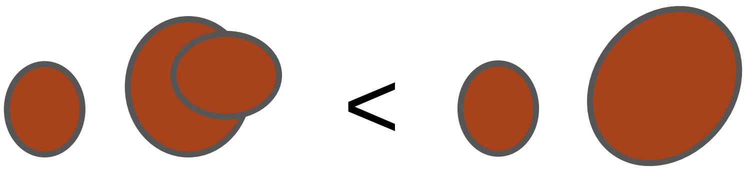
  For a candidate, run ‚partial EM‘ with two components instead of one, keeping all other components fixed

  Evaluate whether BIC decreased

- **Merging two clusters**

  Compute new model with 2 clusters combined into one and evaluate BIC





Modified from Ueda et al. 2002, Neural Computation

GRADING

1 person responsible for report, 1 for presentation

30% - 10% for grade

PRESENTATIONS

Calabrese & Paninski: Spike sorting of non-stationary data with Kalman filter

Pillow et al: A model based spike sorting algorithm

Diego & Hamprecht: Sparse space-time deconvolution for Calcium image analysis

Theis et al: Supervised spike inference from calcium signals

Archer et al: Bayesian entropy estimation

PRESENTATIONS

Macke et al: Empirical models of population activity

Park & Pillow: Bayesian estimation of low rank RFs

Cunningham et al: Gaussian processes rate estimation

Okun et al: Diverse coupling to neural populations

Semedo et al: Extracting latent structure from multiple populations

Fletcher and Randan: Neuronal connectivity from calcium imaging