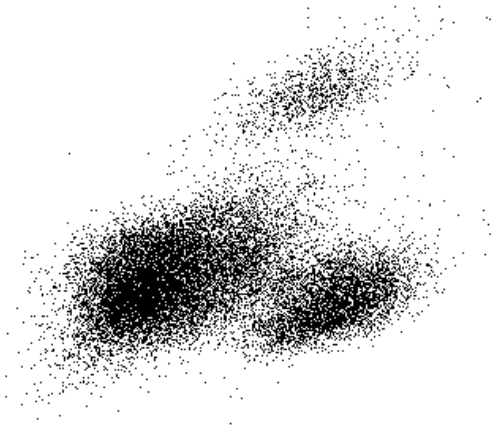
Graduate School of Neural Information Processing, Tübingen
Philipp Berens, Alexander Ecker, Matthias Bethge

Task 2: Spike sorting using mixture models

Due date: Monday, April 27, noon

Prerequisites

Depending on our feedback from task 1, use your extracted features or download the file with our features from Illias. The figure on the right shows two of the features we extracted (1 and 7, corresponding to the first principal component on channel 1 and 3).



Tasks

1. **Generate toy data**: Sample 1000 data points from a two dimensional mixture of Gaussian model with three clusters and the following parameters:

$$\mu_1 = \begin{pmatrix}0\\0\end{pmatrix}, \Sigma_1 = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}, \pi_1 = 0.3$$

$$\mu_2 = \begin{pmatrix}5\\1\end{pmatrix}, \Sigma_2 = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}, \pi_2 = 0.5$$

$$\mu_3 = \begin{pmatrix}0\\2\end{pmatrix}, \Sigma_3 = \begin{pmatrix}1 & -0.5\\-0.5 & 2\end{pmatrix}, \pi_3 = 0.2$$

   Figure 1: Plot the sampled data points and indicate in color the cluster each point came from. Plot the cluster means as well.

2. **Implement a Gaussian mixture model.** Implement the EM algorithm to fit a Gaussian mixture model in `sortSpikes()`. Sort the data points by inferring their class labels from your mixture model (by using maximum a-posteriori classification). Fix the seed of the random number generator to ensure deterministic and reproducible behavior by using `rng(seed)`. Test it on the toy dataset specifying the correct number of clusters and make sure the code works correctly.
   Figure 2: Plot the data points from the toy dataset and indicate in color the cluster each point was assigned to by your model. How does the assignment compare to ground truth?

3. **Model complexity.** A priori we do not how many neurons we recorded. Extend your algorithm with an automatic procedure to select the appropriate number of mixture components (clusters). Base your decision on the Bayesian Information Criterion

$$BIC = -2L + P \log N,$$

   where L is the log-likelihood of the data under the best model, P is the number of parameters of the model and N is the number of data points. You want to minimize the quantity.

Figure 3: Plot the BIC as a function of mixture components. What is the optimal number of clusters on the toy dataset?

4. **Spike sorting using MoG**. Run the full algorithm on your set of extracted features (including model complexity selection).
   Figure 4: (a) Plot the BIC as a function of mixture components. (b) Make scatter plots of the first PCs on all four channels (6 plots). Color-code each data point according to its class label in the model with the optimal number of clusters. In addition, indicate the position (mean) of the clusters in your plot.
5. **Local maxima [optional].** The EM algorithm tends to get stuck in suboptimal local maxima. Implement a procedure to circumvent this problem. You may use one of the following widely used approaches (or a combination thereof) or explore alternatives.
   a. The simplest approach to improve the results is to simply run the algorithm multiple times with different initializations and pick the best.
   b. A procedure called split & merge (e.g. Ueda et al. 2000, Neural Computation) has been shown to work well in helping the algorithm out of local maxima. It splits one cluster into two and merges two other clusters into one, keeping the number of clusters constant.
   c. Splitting or merging can also be used individually (therefore each changing the number of clusters) but in combination with a procedure to determine the model complexity (see above).

Tips

- Use a subset of the data for testing and debugging. But be careful not to make it too small, since the algorithm may fail to detect small clusters in this case.
- Split & merge: when splitting a cluster it is advisable to first run a couple of 'partial EM iterations' on only the data points assigned to that cluster before continuing with the full dataset. This both speeds up convergence and prevents 'component starvation', where a cluster becomes unused because no data points get assigned to it.