# Clustering and Mixture Models

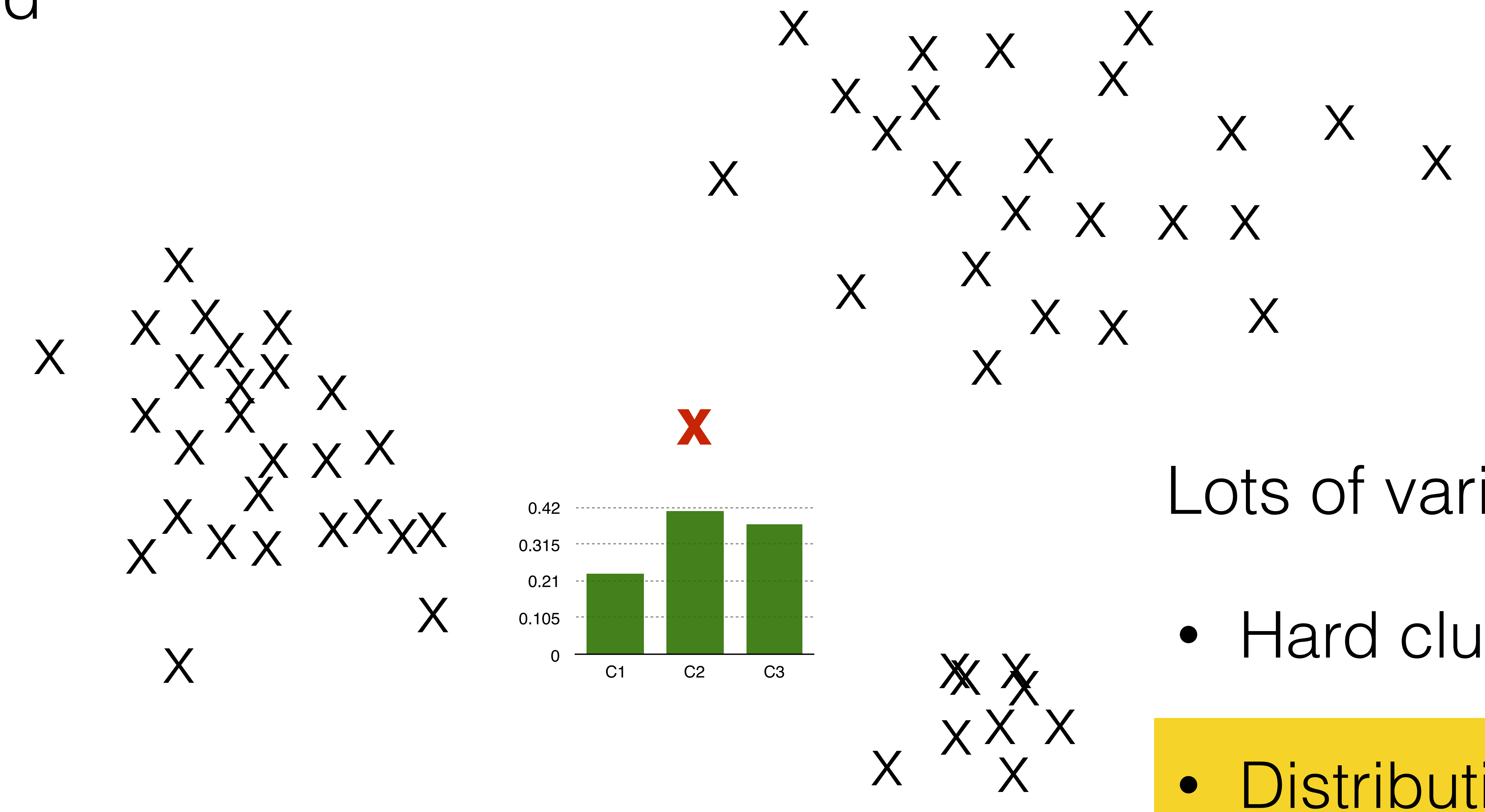
Machine Learning
CSx824/ECEx242
Bert Huang
Virginia Tech

# Outline

- Clustering intuition
- Mixture models
- Mixture of Gaussians
- Expectation maximization recipe

# Clustering

unsupervised




Lots of variants:

- Hard cluster assignment
- Distribution-based
- Hierarchical, etc.

# Mixture Models

$$X = \{x_1, \ldots, x_n\}$$

$$P(X) = \prod_{i=1}^{n} \sum_{c_i=1}^{K} p(c_i) p(x_i | c_i)$$

probability of $\mathbf{x_i}$ if **i** is in cluster $\mathbf{c_i}$

probability that example **i** is in cluster $\mathbf{c_i}$

**generative** process:

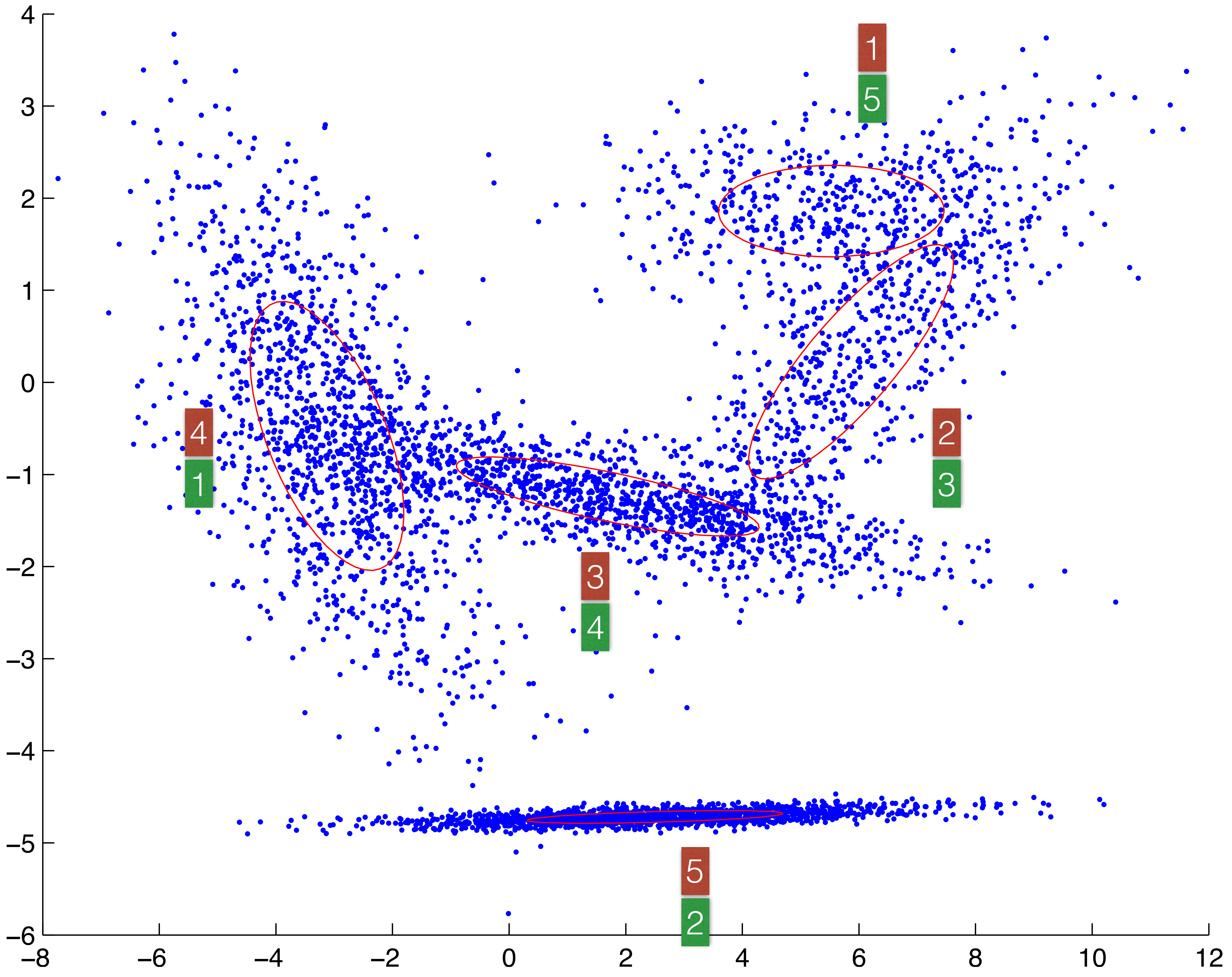
1. Sample cluster
2. Sample data example from cluster distribution

# Gaussian Mixture Model

$$P(x) = \sum_{c=1}^{K} p(c) \frac{1}{\sqrt{2\pi|\Sigma_c|}} \exp\left(-\frac{1}{2}(x-\mu_c)^\top \Sigma_c^{-1}(x-\mu_c)\right)$$

multinomial cluster membership

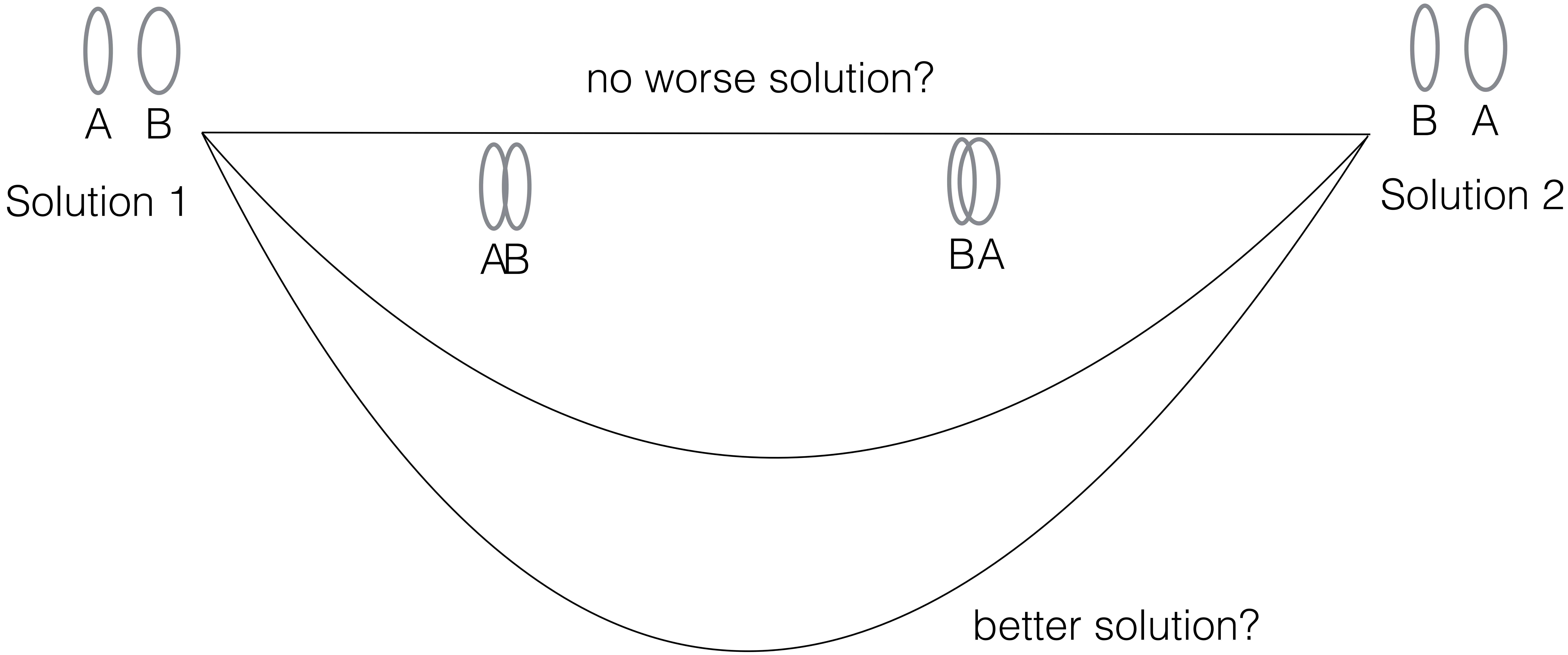
multivariate Gaussian data $\mathcal{N}(x|\mu_c, \Sigma_c)$




“clouds” can overlap

no identity for clusters

# Non-Convexity of GMM NLL



# Expectation Maximization Recipe

Input: $x_i \qquad i \in \{1, \ldots, n\}$

GMM parameters:

$$p(c) \quad \mu_c \quad \Sigma_c \qquad c \in \{1, \ldots, K\}$$

Latent variables:

$$z_i \in \{1, \ldots, K\}$$

Latent variable probabilities: $p(z_i)$

$$\sum_{c=1}^{K} p(c) = \sum_{c=1}^{K} p(z_i = c) = 1$$

E-step: fit latent variable probabilities

$$p(z_i = c) \leftarrow \frac{p(c)\mathcal{N}(x_i | \mu_c, \Sigma_c)}{\sum_{c'=1}^{K} p(c')\mathcal{N}(x_i | \mu_{c'}, \Sigma_{c'})}$$

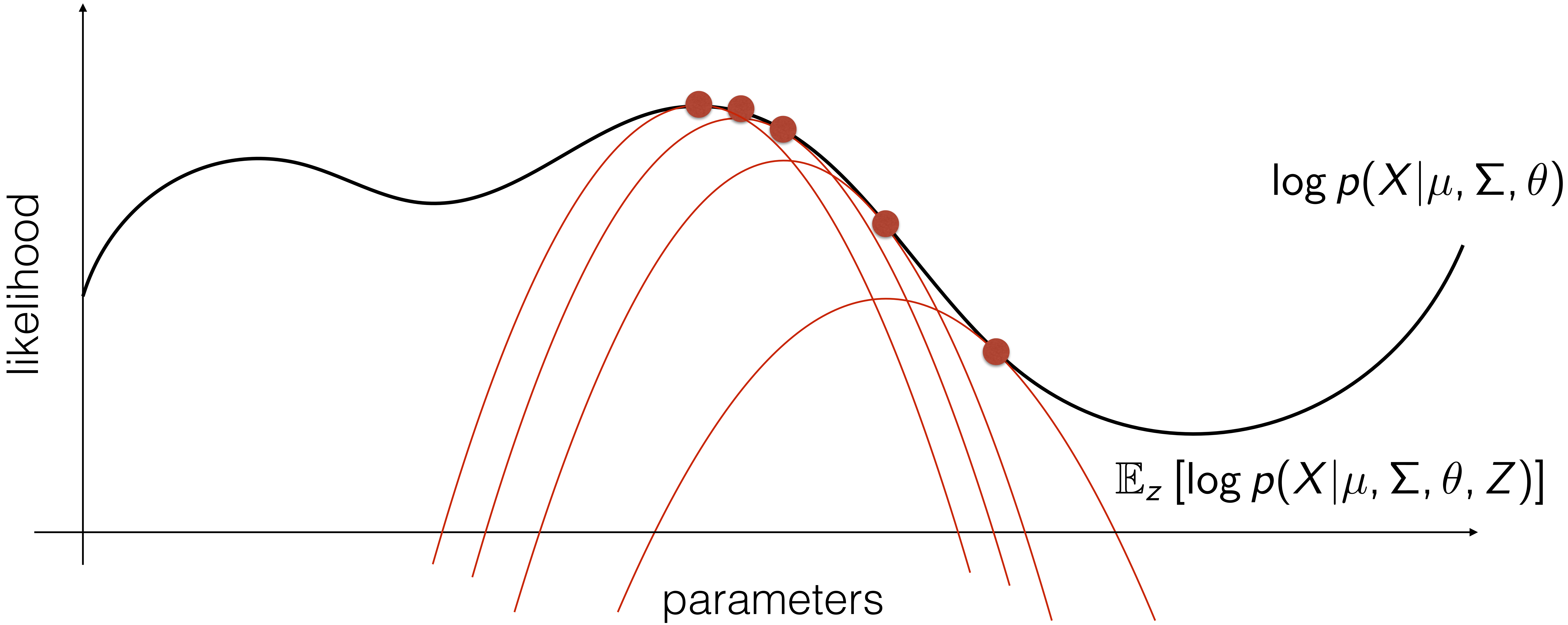
M-step: fit GMM parameters using expected likelihood

$$p(c) \leftarrow \frac{1}{n} \sum_{i=1}^{n} p(z_i = c)$$

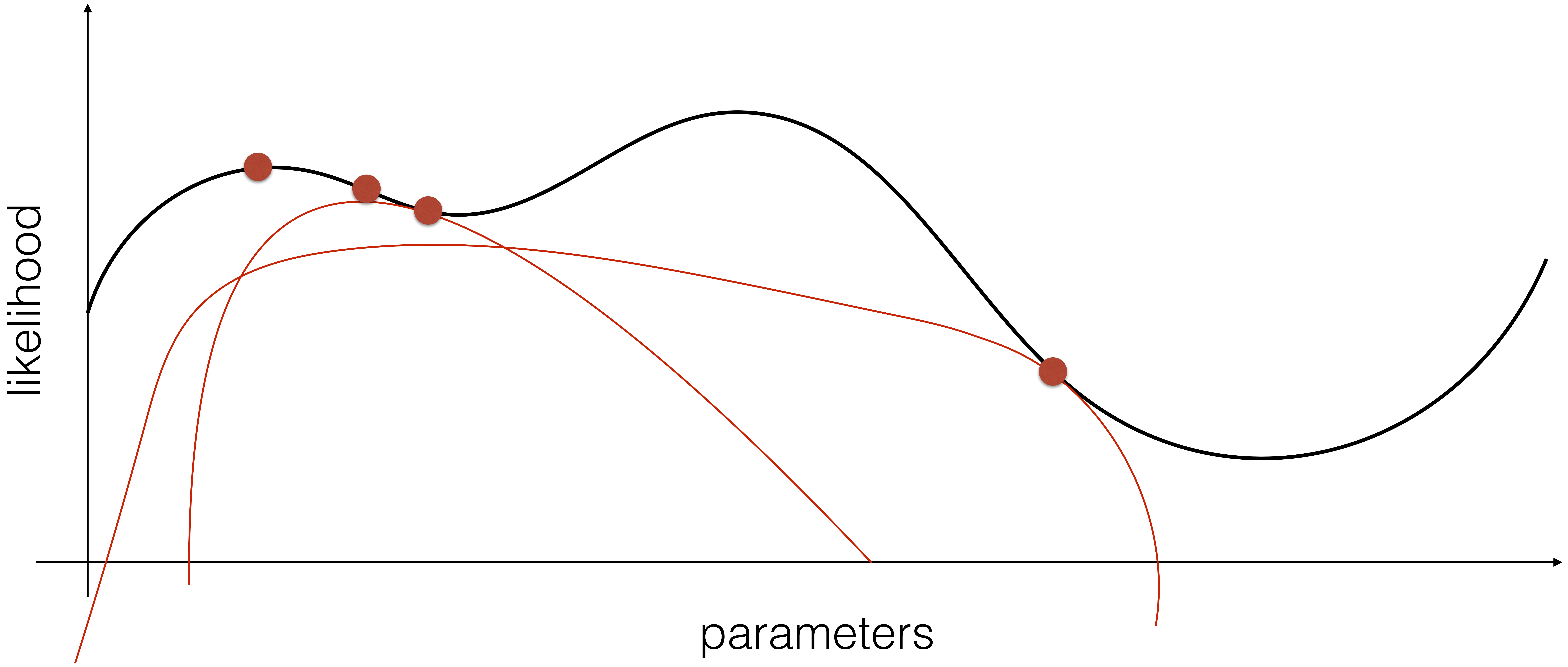
$$\mu_c \leftarrow \frac{\sum_{i=1}^{n} p(z_i = c) x_i}{\sum_{i=1}^{n} p(z_i = c)}$$

$$\Sigma_c \leftarrow \frac{\sum_{i=1}^{n} p(z_i = c)(x_i - \mu_c)(x_i - \mu_c)^\top}{\sum_{i=1}^{n} p(z_i = c)}$$

# EM as Maximizing Lower Bound



# EM as Maximizing Lower Bound



# EM Likelihood Landscape

# Initialization

- Some heuristics:
  - Completely random
  - Fit a single Gaussian to all data; randomly perturb K copies
  - Randomly initialize cluster memberships. Start with M-step

# Next Time

- Variational expectation maximization
- Hard EM and K-means