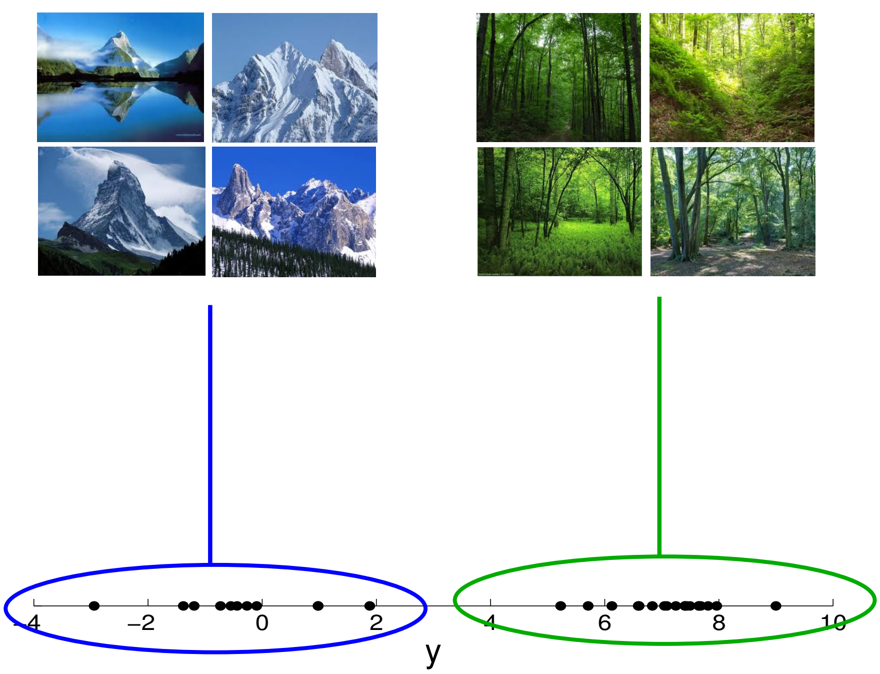
Simple clustering example

Paper 8, Easter term 2013

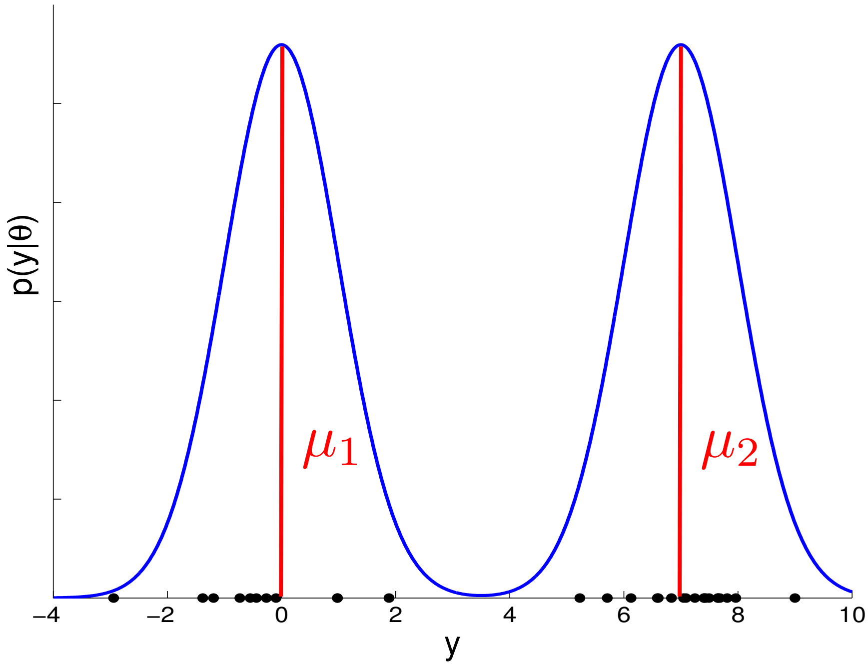
Richard E. Turner (ret26@cam.ac.uk)
Computational and Biological Learning Lab
University of Cambridge

Clustering



Mixture of Gaussians

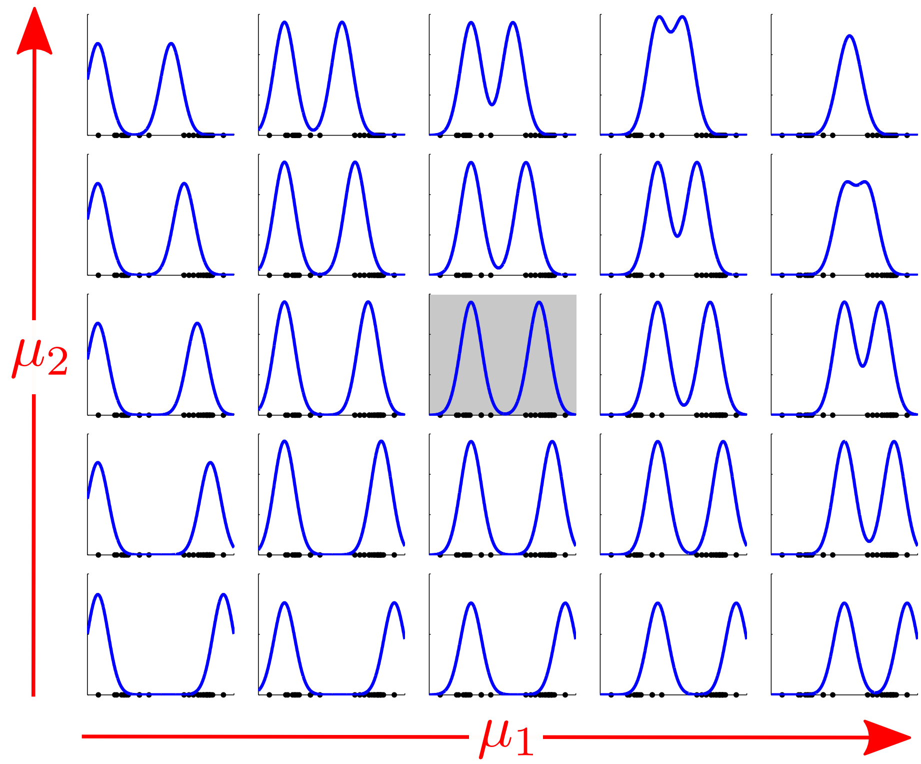
$$p(y|\theta) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-\mu_1)^2\right) + \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-\mu_2)^2\right)$$



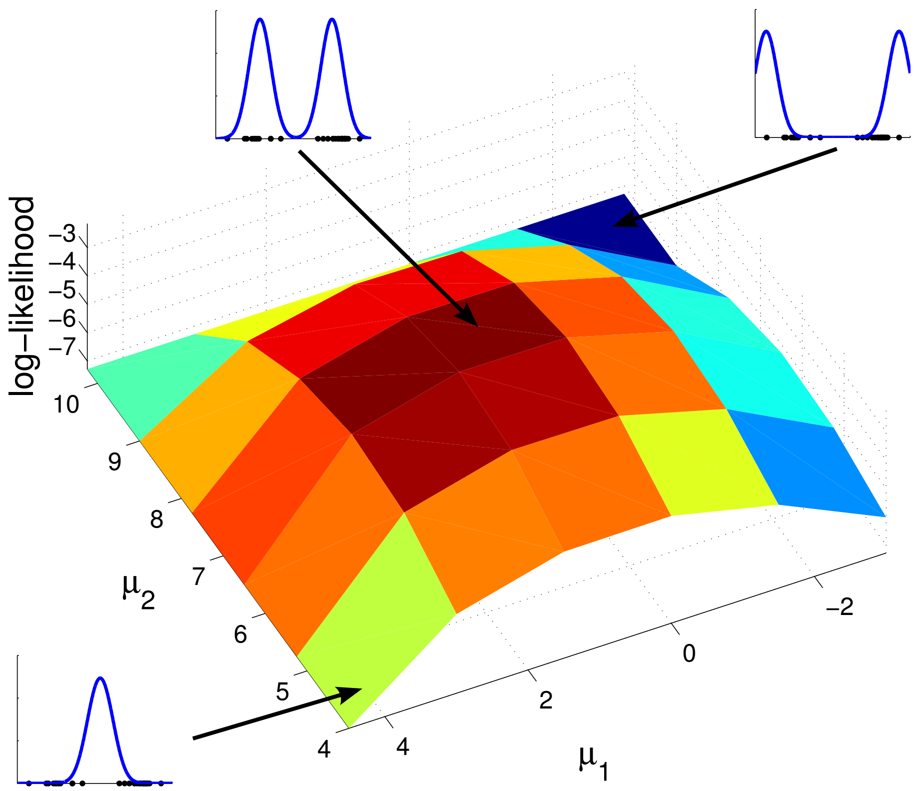
Maximum likelihood learning

$$\begin{aligned} p(y_1, y_2, \dots, y_N|\theta) &= \prod_{n=1}^{N} p(y_n|\theta) \\ &= (8\pi)^{-N/2} \prod_{n=1}^{N} \left[\exp\left(-\frac{1}{2}(y_n-\mu_1)^2\right) + \exp\left(-\frac{1}{2}(y_n-\mu_2)^2\right)\right] \end{aligned}$$

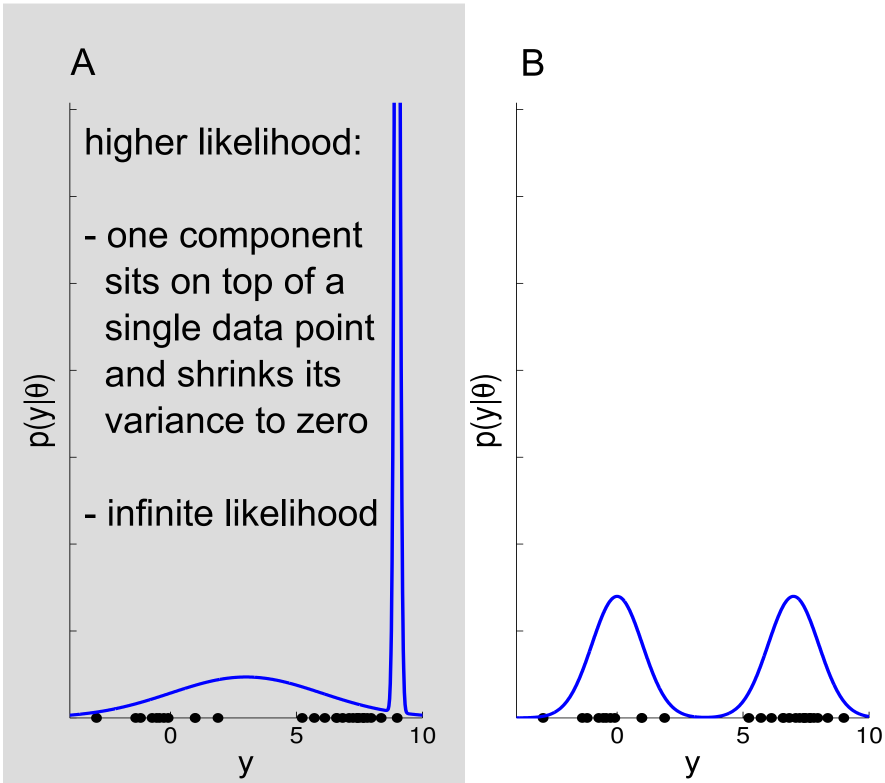
Which parameters have highest likelihood?



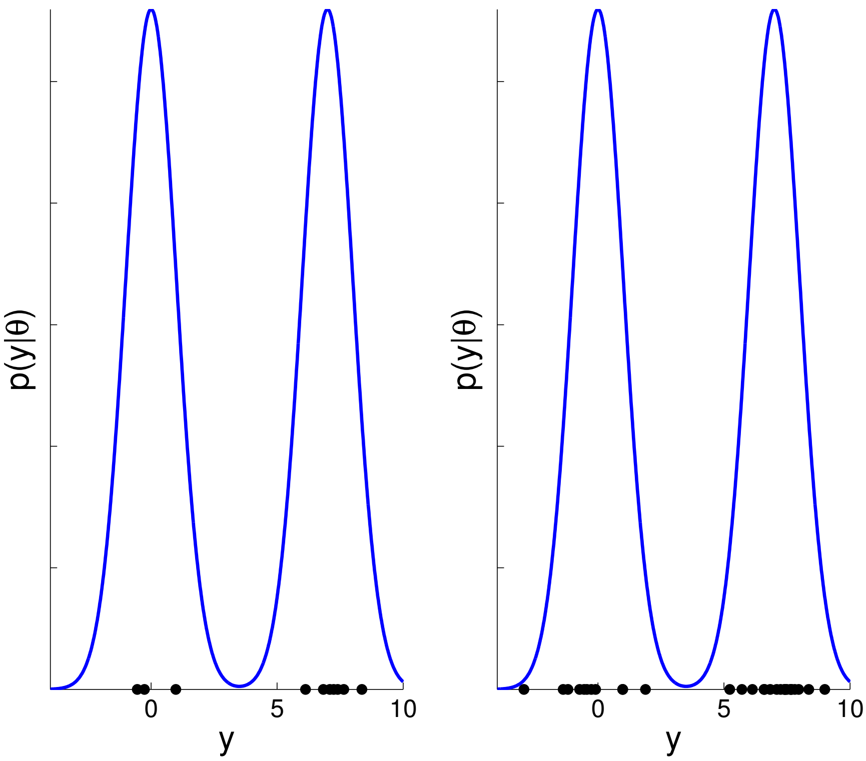
Log-likelihood for the different parameter settings



Which parameters have the highest likelihood?



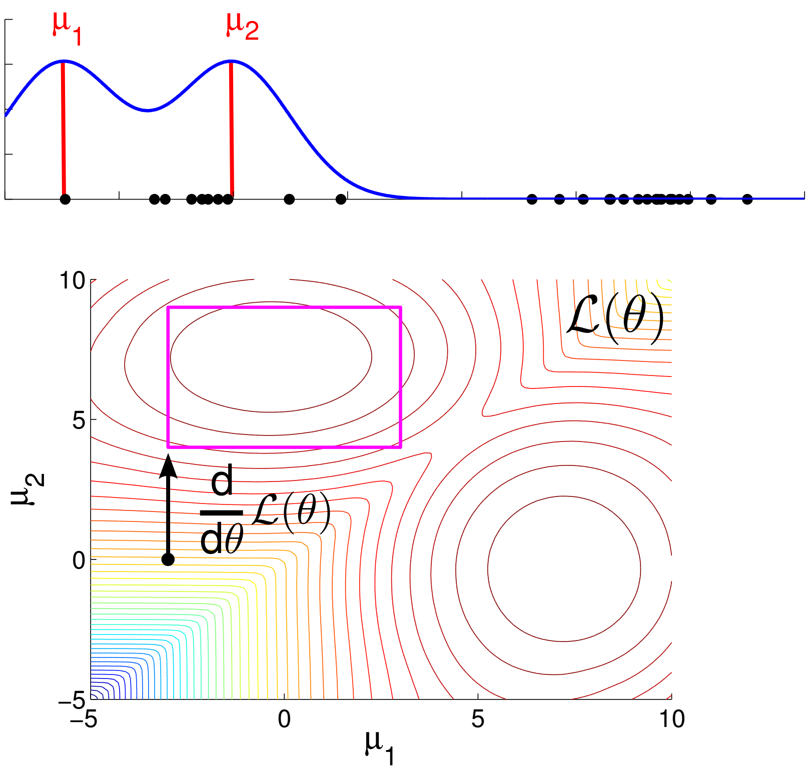
Which dataset leads to estimates with highest confidence?



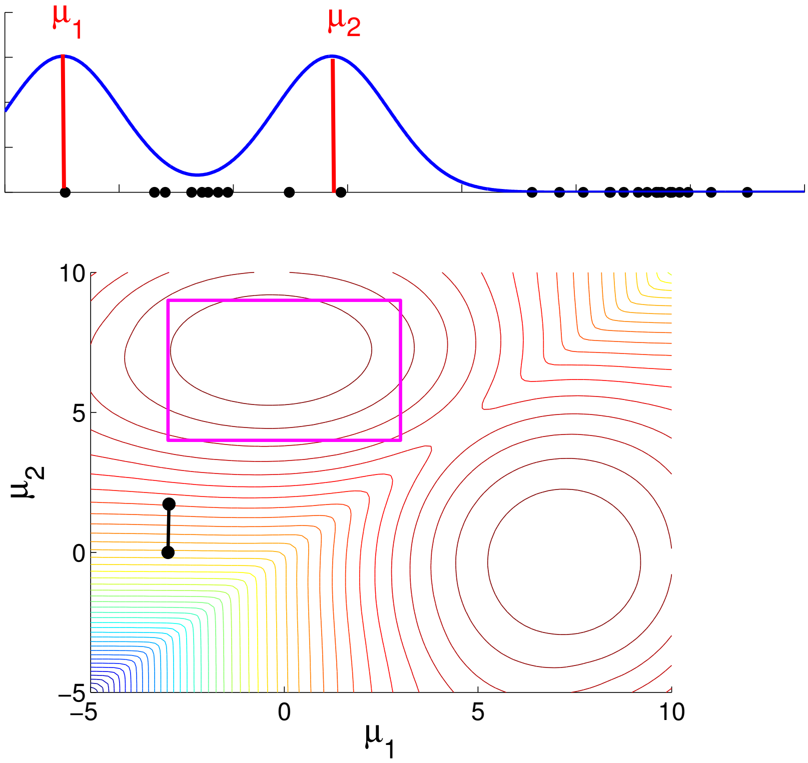
Summary

- Maximum likelihood can be used to estimate parameters from data
- But...
  - costly to grid up parameter space and evaluate likelihood for each parameter setting
  - maximum likelihood can "over fit"
  - unclear how to get back uncertainty estimates
  - unclear how to use it to solve questions like "how many clusters are in the dataset?"

What will happen on the first iteration?



Maximum-likelihood and gradient ascent



Maximum-likelihood and gradient ascent