Clustering example
Mixture of Gaussians

\[
\left[ \int \frac{\exp\left( \frac{1}{2}(y_n - \mu_1)^2\right)}{(2\pi)^{N/2}} \, dx \right] \prod_{N} \frac{1}{\sqrt{1 - \mu_8}} = (\theta|\mu_8) \prod_{N} = (\theta|N\mu, \cdots, \mu_8, \mu_8) d
\]

Maximum Likelihood Learning

\[
\left( \int \frac{\exp\left( \frac{1}{2}(y_n - \mu_1)^2\right)}{(2\pi)^{N/2}} \, dx \right) \frac{\mu_2}{\sqrt{1 - \mu_8}} + \left( \int \frac{\exp\left( \frac{1}{2}(y_n - \mu_2)^2\right)}{(2\pi)^{N/2}} \, dx \right) \frac{\mu_2}{\sqrt{1 - \mu_8}} = (\theta|\mu_8) d
\]

Mixture of Gaussians
Which parameters have highest likelihood?

Log-likelihood for the different parameter settings

Which parameters have highest likelihood?
Which parameters have the highest likelihood?

\[ p(y | \theta) \]

Higher likelihood:
- one component sits on top of a single data point and shrinks its variance to zero
- infinite likelihood

Which dataset leads to estimates with highest confidence?

Which parameters have the highest likelihood?
Maximum likelihood can be used to estimate parameters from data

- 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?

But...

Maximum likelihood can be used to estimate parameters from data

Summary
Maximum-likelihood and gradient ascent