Learning general mixtures of Gaussian

$$P(X = x|Y = i) = \frac{1}{\sqrt{(2\pi)^m |\Sigma_i|}} \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right)$$

- Marginal likelihood, for data $\{x^j \mid j = 1..n\}$:

$$\prod_{j=1}^{n} P(x^j) = \prod_{j=1}^{n} \sum_i P(X = x^j, Y = i) = \prod_{j=1}^{n} \sum_i P(X = x^j | Y = i) P(Y = i)$$

$$= \prod_{j=1}^{n} \sum_i \frac{1}{\sqrt{(2\pi)^m |\Sigma_i|}} \exp\left(-\frac{1}{2}(x^j-\mu_i)^T \Sigma_i^{-1} (x^j-\mu_i)\right) P(Y = i)$$

- Need to differentiate and solve for μ_i, Σ_i, and P(Y=i) for i=1..k
- There will be no closed for solution, gradient is complex, lots of local optimum
- Wouldn't it be nice if there was a better way!