Soft clustering:

- So far, clustering was done using **hard assignments** (1 point -> 1 cluster)
- Sometimes this doesn't accurately represent the data: it seems reasonable to have overlapping clusters.
- In this case, we can use **soft assignment** to assign points to every cluster with a certain probability.
    - Example:
        - Generate data using $\mathbf{N(\mu_1, \sigma_1)}$ and $\mathbf{N(\mu_2, \sigma_2)}$
        - we are given the weights of animals. Unknown to us these are weights from two different species.
        - Can we determine the species (group / assignment) from the height?
            - Any of these points could technically have been generated from either curve
            - We can create soft assignments based on these probabilities
- Mixture model:
    - X comes from a mixture model with k mixture components if the probability distribution of X is:
        - $$P(X = x) = \sum_{j=1}^{k} P(C_j) P(X = x | C_j)$$
- Gaussian mixture model:
    - $$P(X = x | C_i) \sim N(\mu, \sigma)$$
    - $$P(X = x | C_i) \sim N(\mu, \sigma)$$
- GMM Clustering:
    - **Goal**: Find the GMM that maximizes the probability of seeing the data we have.
    - The probability of seeing the data we saw is (assuming each data point was sampled independently) the product of the probabilities of observing each data point.
    - Finding the GMM means finding the parameters that uniquely characterize it. What are these parameters?
        - $\mathbf{P(C_i)}$ & $\boldsymbol{\mu_i}$ & $\boldsymbol{\sigma_i}$ for all **k** components.
        - Lets call $\boldsymbol{\Theta}$ = $\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_k, P(C_1), \ldots, P(C_k)\}$
    - Goal:
        - $$\theta^* = \arg\max_{\theta} \prod_{i=1}^{n} \sum_{j=1}^{k} P(C_j) P(X_i \mid C_j)$$
            - Where $\boldsymbol{\Theta}$ = $\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_k, P(C_1), \ldots, P(C_k)\}$
            - Joint probability distribution of our data
            - Assuming our data are independent
- How do we find the critical points of this function?

- Notice: taking the log-transform does not change the critical points
- Define :
  - $$\begin{aligned} l(\theta) &= \log(L(\theta)) \\ &= \sum_{i=1}^{n} \log(\sum_{j=1}^{k} P(C_j)P(X_i \mid C_j)) \end{aligned}$$

Expectation Maximization Algorithm:

- 1.Start with random **θ**
- 2.Compute **P(C_j | X_l)** for all **X_i** by using **θ**
- 3.Compute / Update **θ** from **P(C_j | X_l)**
- 4.Repeat 2 & 3 until convergence