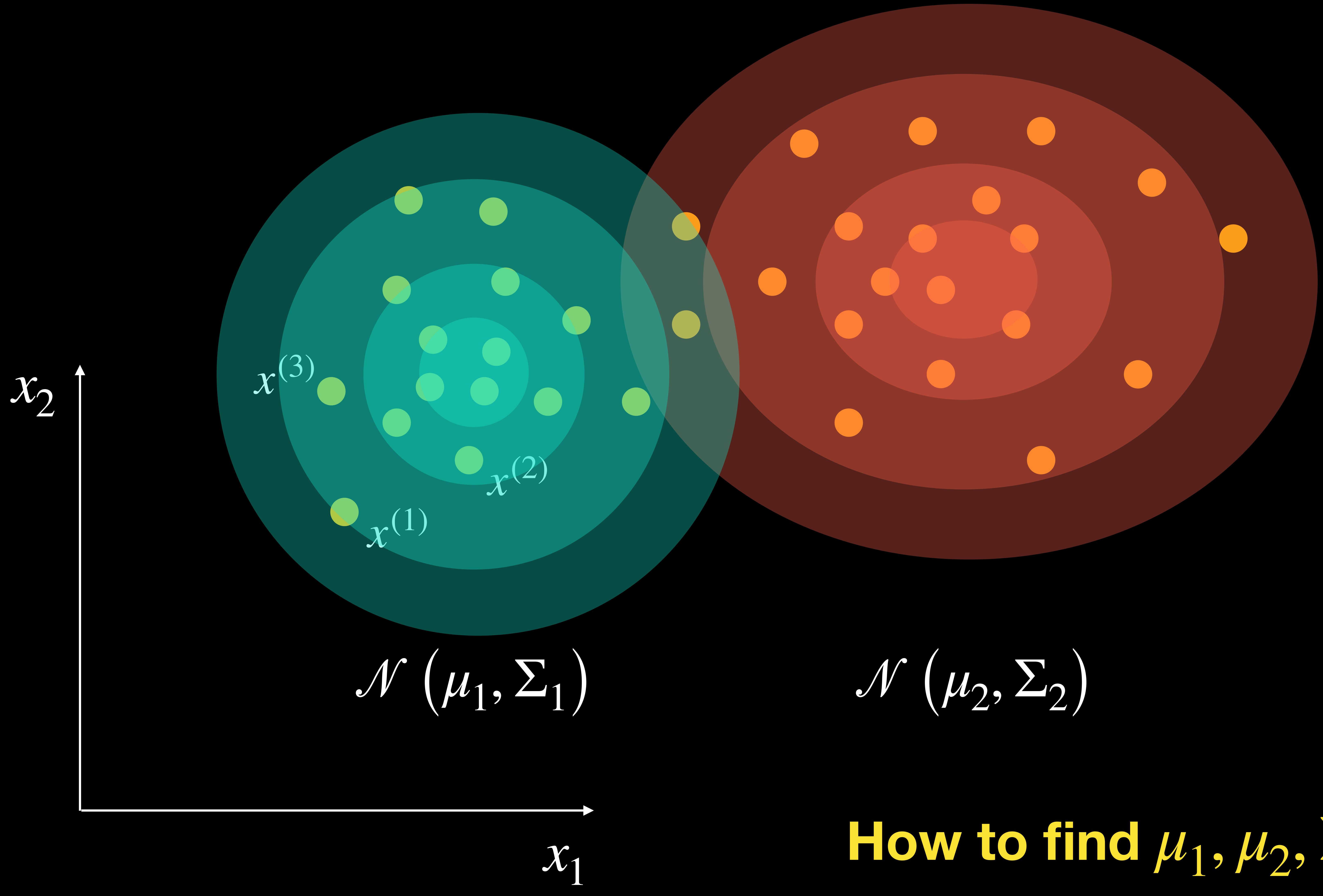
Gaussian Mixture Models

Prepared by: Joseph Bakarji

Modeling data as a Mixture of Gaussians

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...



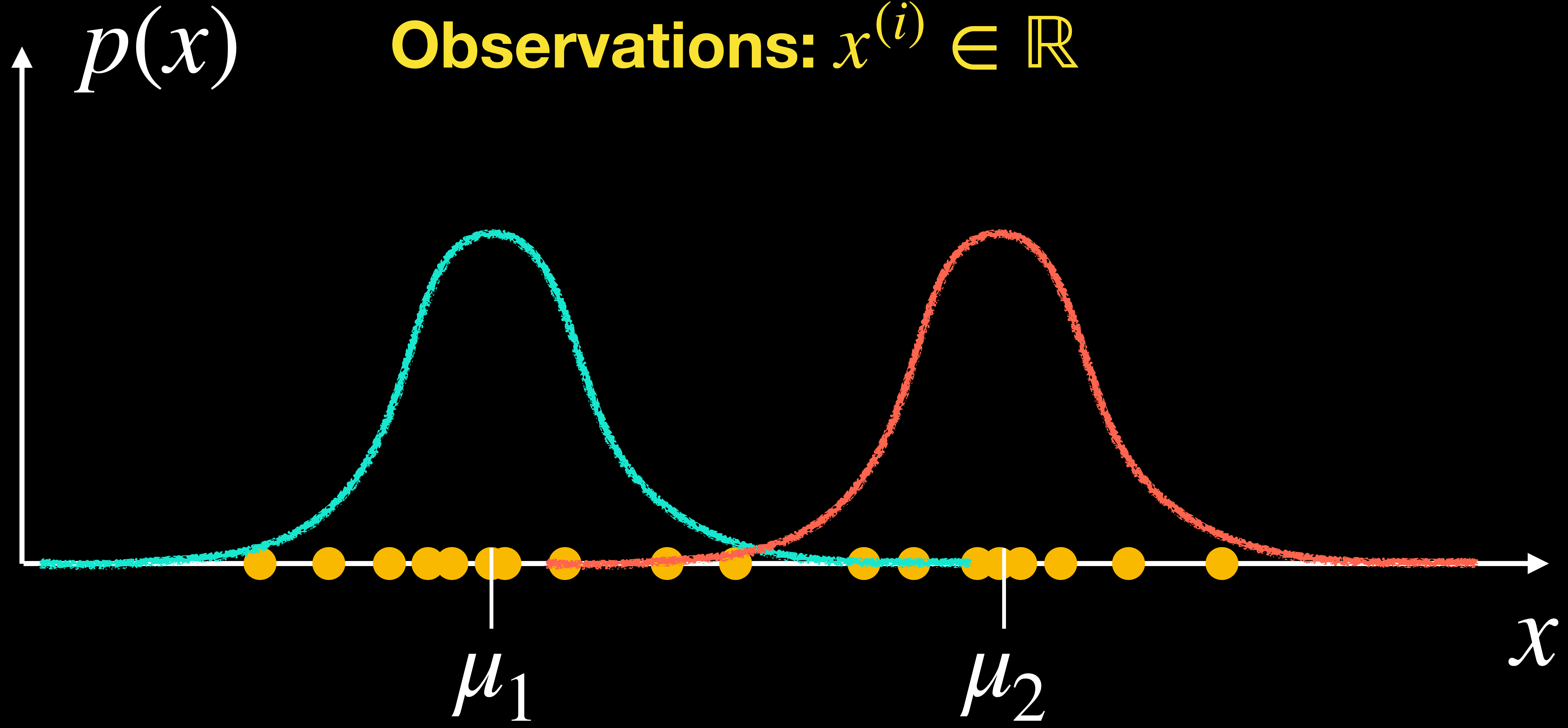
How to find $\mu_1, \mu_2, \Sigma_1, \Sigma_2$?

Modeling data as a Mixture of Gaussians

Observations: $x^{(i)} \in \mathbb{R}$

Dataset

x
1.2
3.2
4.3
3.2
…



Assume:

- **Data is modeled by mixture of gaussians:** $\mathcal{N}\left(\mu_i, \sigma_i^2\right)$
- **We know how many gaussians we need**

Find μ_1, μ_2

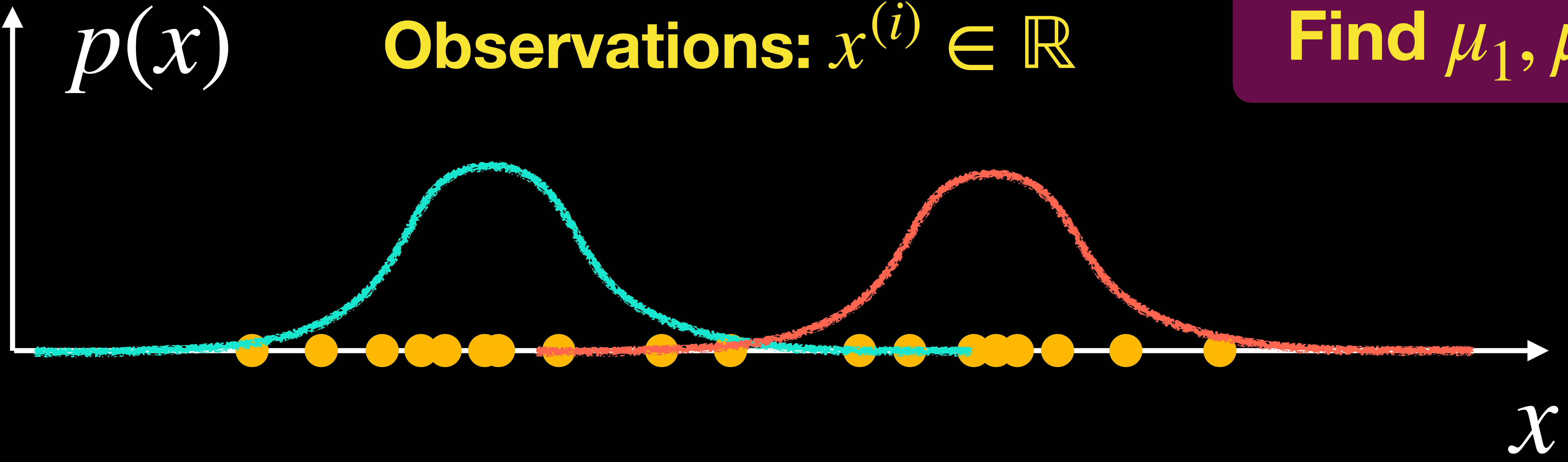
Modeling data as a Mixture of Gaussians

Dataset

x
1.2
3.2
4.3
3.2
…

Observations: $x^{(i)} \in \mathbb{R}$

Find μ_1, μ_2



Given: $x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(n)}$, **and** $k > 0$**: number of clusters**

Find $P\left(z^{(i)} = j\right)$ **Soft Assignment**

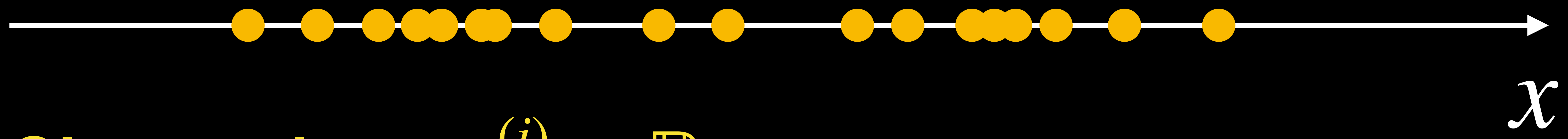
Probability that point i **belongs to cluster** j

$z^{(i)}$ **is not directly observed: Latent variable**

Gaussian Mixture Model

Dataset

x
1.2
3.2
4.3
3.2
…



Observations: $x^{(i)} \in \mathbb{R}$

Num. of Clusters: $k > 0$ **clusters**

Soft assignments: $z^{(i)}$

$$P\left(x^{(i)}, z^{(i)}\right) = P\left(x^{(i)} \,|\, z^{(i)}\right) P\left(z^{(i)}\right)$$

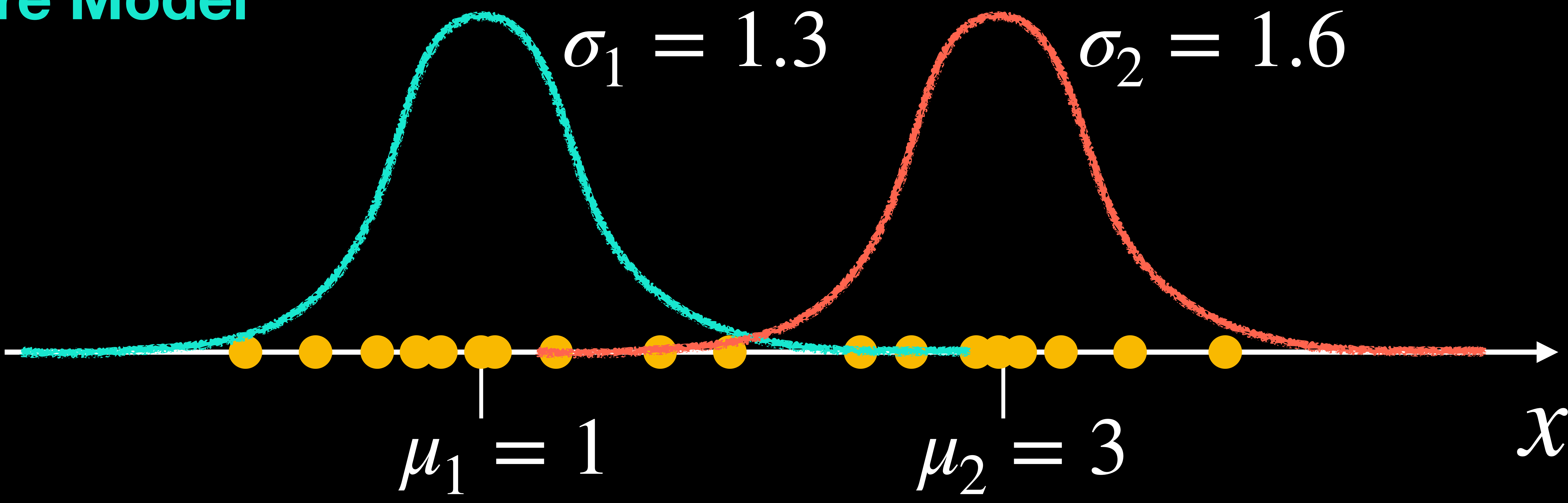
$$P(z^{(i)}) = \text{Multinomial}\left(\phi\right) \qquad \sum_{i=1}^{k} \phi_k = 1$$

$$P\left(x^{(i)} \,|\, z^{(i)} = j\right) = \mathcal{N}\left(\mu_j, \sigma_j^2\right)$$

Gaussian Mixture Model

Dataset

x
1.2
3.2
4.3
3.2
…



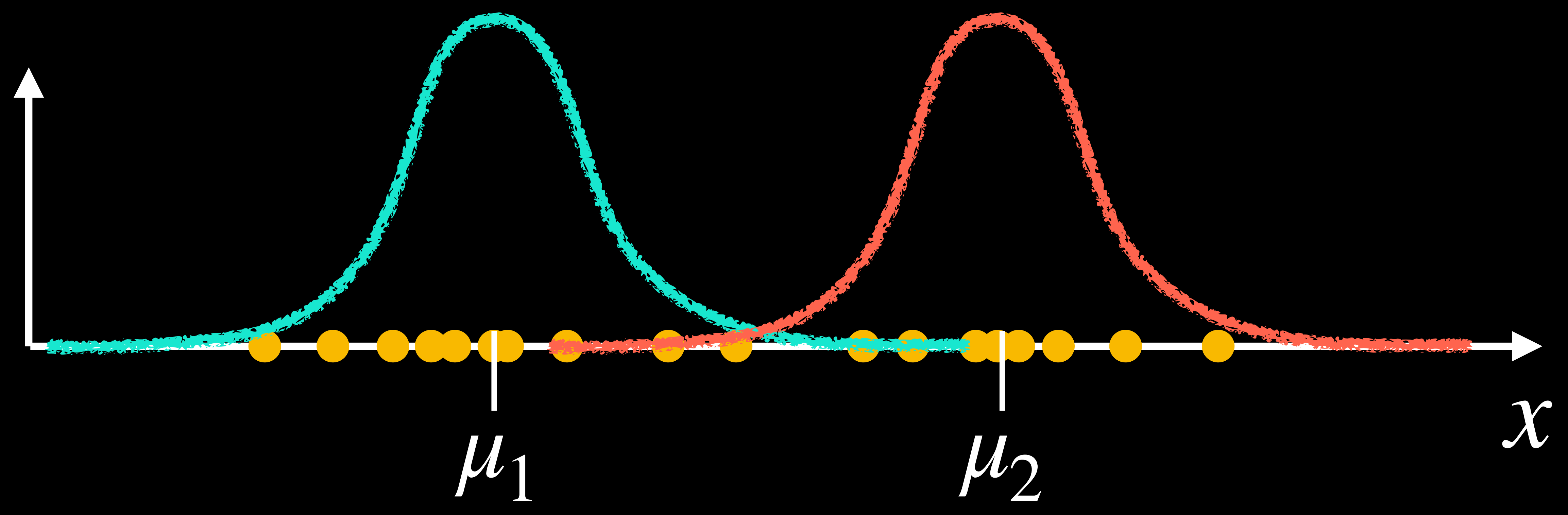
Intuition: if one were to 'create' the data

Sample

$\phi_1 = 0.3 \longrightarrow$ **Sample** $\mathcal{N}\left(\mu_1, \sigma_1^2\right) \longrightarrow x^{(i)}$

$\phi_2 = 0.7 \longrightarrow$ **Sample** $\mathcal{N}\left(\mu_2, \sigma_2^2\right) \longrightarrow x^{(i)}$

Expectation-Maximization (EM) Algorithm



E-Step: Guess $z^{(i)}$ given ϕ, μ, σ. Compute $P\left(z^{(i)} \,|\, x^{(i)}; \phi, \mu, \sigma\right)$

M-Step: Estimate ϕ, μ, σ using Maximum Likelihood Estimation

Expectation Step

E-Step: Guess $z^{(i)}$ given ϕ, μ, σ

$$w_j^{(i)} = P\left(z^{(i)} = j \,|\, x^{(i)}; \phi, \mu, \sigma\right) = \frac{P\left(z^{(i)} = j, x^{(i)}\right)}{P\left(x^{(i)}\right)}$$

$$= \frac{P\left(x^{(i)} \,|\, z^{(i)} = j\right) P\left(z^{(i)} = j\right)}{P\left(x^{(i)}\right)}$$

$$= \frac{P\left(x^{(i)} \,|\, z^{(i)} = j\right) P\left(z^{(i)} = j\right)}{\sum_{s=1}^{k} P\left(x^{(i)} \,|\, z^{(i)} = s\right) P\left(z^{(i)} = s\right)}$$

Expectation Step

E-Step: Guess $z^{(i)}$ given ϕ, μ, σ

$$w_j^{(i)} = P\left(z^{(i)} = j \mid x^{(i)}; \phi, \mu, \sigma\right)$$

$$= \frac{P\left(x^{(i)} \mid z^{(i)} = j\right) P\left(z^{(i)} = j\right)}{\sum_{s=1}^{k} P\left(x^{(i)} \mid z^{(i)} = s\right) P\left(z^{(i)} = s\right)}$$

$\mathcal{N}(\mu_j, \sigma_j^2)$

ϕ_j

ϕ_s

$\mathcal{N}\left(\mu_s, \sigma_s^2\right)$

Maximization Step

M-Step: Given $P\left(z^{(i)} = j\right) \equiv w_j^{(i)}$ **estimate** ϕ, μ, σ

Maximum Likelihood Estimation

$$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^{n} \log P(x^{(i)}; \phi, \mu, \sigma)$$

$$= \sum_{i=1}^{n} \log \sum_{z^{(i)}=1}^{k} P(x^{(i)} \,|\, z^{(i)}; \mu, \sigma)\, P(z^{(i)}; \phi)\,.$$

$$= \sum_{i=1}^{n} \log p(x^{(i)} \,|\, z^{(i)}; \mu, \sigma) + \log p(z^{(i)}; \phi)$$

if $z^{(i)}$ is known

Maximization Step

M-Step: Given $P\left(z^{(i)}=j\right) \equiv w_j^{(i)}$ **estimate** ϕ, μ, σ

If $z^{(i)}$ were known → **Gaussian Discriminant Analysis**

$$\phi_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{z^{(i)}=j\}$$

$$\mu_j = \frac{\sum_{i=1}^{n} \mathbf{1}\{z^{(i)}=j\}x^{(i)}}{\sum_{i=1}^{n} \mathbf{1}\{z^{(i)}=j\}}$$

$$\Sigma_j = \frac{\sum_{i=1}^{n} \mathbf{1}\{z^{(i)}=j\}(x^{(i)}-\mu_j)(x^{(i)}-\mu_j)^T}{\sum_{i=1}^{n} \mathbf{1}\{z^{(i)}=j\}}$$

Maximization Step

M-Step: Given $P\left(z^{(i)} = j\right) \equiv w_j^{(i)}$ **estimate** ϕ, μ, σ

Since $z^{(i)}$ is not known, we use **soft assignments** instead:

$$\phi_j = \frac{1}{n} \sum_{i=1}^{n} w_j^{(i)}$$

$$\mu_j = \frac{\sum_{i=1}^{n} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{n} w_j^{(i)}}$$

$$\Sigma_j = \frac{\sum_{i=1}^{n} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{n} w_j^{(i)}}$$

EM Algorithm Summary

Iterate

E-Step: Estimate $w_j^{(i)}$

$$w_j^{(i)} = \frac{\overbrace{P\left(x^{(i)} \mid z^{(i)} = j\right)}^{\mathcal{N}(\mu_j, \Sigma_j)} \overbrace{P\left(z^{(i)} = j\right)}^{\phi_j}}{\sum_{s=1}^{k} \underbrace{P\left(x^{(i)} \mid z^{(i)} = s\right)}_{\mathcal{N}\left(\mu_s, \Sigma_s\right)} \underbrace{P\left(z^{(i)} = s\right)}_{\phi_s}}$$

M-Step: Estimate ϕ, μ, Σ

$$\phi_j = \frac{1}{n} \sum_{i=1}^{n} w_j^{(i)}$$

$$\mu_j = \frac{\sum_{i=1}^{n} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{n} w_j^{(i)}}$$

$$\Sigma_j = \frac{\sum_{i=1}^{n} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{n} w_j^{(i)}}$$