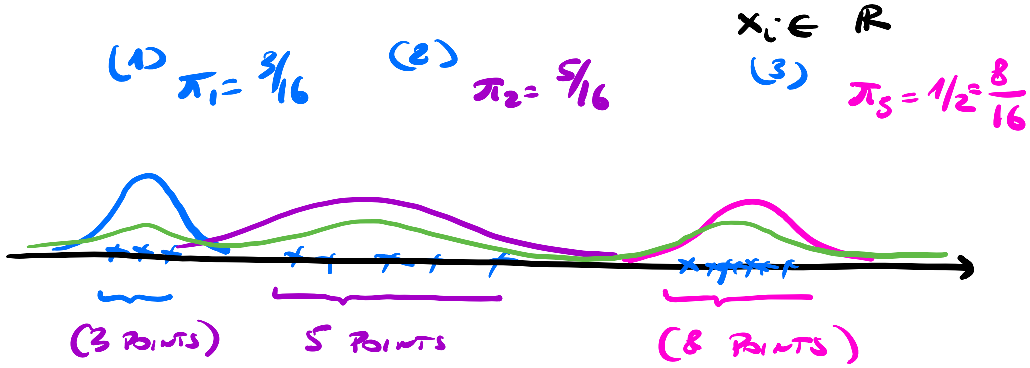
The Gaussian Mixture Model (GMM)

For clustering



Generative probabilistic model for the data $x_i \in \mathbb{R}^d$

$z_i \in \{1, \dots, K\}$ $\quad K \equiv$ # of clusters.

$$P(z_i = j) = \pi_j \qquad j \in \{1 \dots K\}$$

$$P(x_i \mid z_i = j) \sim \mathcal{N}(\mu_j, \Sigma_j)$$

Generative procedure for x_i:

1. Sample $z_i \in \{1 .. K\}$
2. Sample x_i from $p(x_i \mid z_i)$

LET'S COMPUTE THE PROBABILISTIC DESCRIPTION OF x_i.

$$p(x_i) = \sum_{j=1}^{K} \underbrace{p(x_i \mid z_i = j)\, p(z_i = j)}_{p(x_i,\, z_i = j)}$$

$$= \sum_{j=1}^{K} \underbrace{p(x_i \mid z_i = j)}_{\mathcal{N}(\mu_j, \Sigma_j)} \pi_j$$

- $0 \leq \pi_j \leq 1$
- $\sum_{j=1}^{K} \pi_j = 1$

ASSUMPTION:

$\{x_i\}_{i \in [m]}$ are i.i.d samples FROM THE UNKNOWN DISTRIBUTION

$$p(x_i) = \sum_{j=1}^{K} p(x_i \mid z_i = j)\, \pi_j \qquad (1)$$

μ_j, Σ_j

PROBLEM # 1

GIVEN DATA $\{x_i\}_{i \in [m]}$ ESTIMATE THE PARAMETERS (UNKNOWN) OF MODEL (1)

THAT IS $\theta = \{\mu_j, \Sigma_j, \pi_j, j = 1 \dots K\}$

PROBLEM # 2

ASSUME θ IS FIXED AND KNOWN.

AND GIVEN A NEW DATA POINT x

"DECIDE" TO WHICH CLUSTER x BELONGS TO

↓

GIVEN θ WE KNOW:

$$p(x \mid z = j) \quad \text{AND} \quad p(z = j) = \pi_j$$

! CAN COMPUTE

$$p(z = j \mid x) = \frac{p(x \mid z = j)\, p(z = j)}{\underbrace{p(x)}_{\text{CAN BE COMPUTED}}}$$

FOR ANY X
WHEN θ IS KNOWN.

BACK TO PROBLEM 1

SOLVE BY MAXIMUM LIKELIHOOD.

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \underbrace{P_\theta(x_1 x_2 \ldots x_m)}_{\prod_{i=1}^{m} \underbrace{P_\theta(x_i)}_{=(1)}}$$

WE WANT TO INTRODUCE AN ITERATIVE PROCEDURE THAT PRODUCES A SEQUENCE OF ESTIMATORS $\hat{\theta}^{(k)}$ $k=1,2,\ldots$ SO THAT (HOPEFULLY)

$$\lim_{k \to \infty} \hat{\theta}^{(k)} = \hat{\theta}_{ML}$$

WARNING: WE WILL ONLY BE ABLE TO PRODUCE A SEQUENCE

$\hat{\Theta}^{(k)}$ THAT CONVERGES TO

A LOCAL MAXIMUM OF $P_\Theta(x_1, x_2, \dots x_m)$

THE PROCEDURE WE INTRODUCE IS

CALLED THE EM - ALGORITHM.

EM = EXPECTATION

MAXIMIZATION

<u>REMARK:</u> IF WE ALSO KNEW ALL

THE (<u>HIDDEN</u>) VARIABLES

$z_1 \dots z_m$ THEN THE PROBLEM.

$$\hat{\Theta} = \arg\max \log P_\Theta(x_1, z_1, x_2 z_2, \dots x_m, z_m)$$

<u>IDEA:</u> CONSTRUCT AN ESTIMATE

OF $\log P_\Theta(x_1 z_1, \dots x_m, z_m)$

$\tilde{z}_i$

(2)

$$Q(\theta, \hat\theta^{(k)}) = \underbrace{\mathbb{E}_{p_{\hat\theta^{(k)}}(\underline{z}|\underline{x})} \log p_\theta(x_1, z_1, \dots, x_m, z_m)}_{= \sum_{z_1=1}^{K} \cdots \sum_{z_m=1}^{K} \log p_\theta(x_1 z_1 \dots x_m, z_m)\, p_{\hat\theta^{(k)}}(z_1|x_1) \dots p_{\hat\theta^{(k)}}(z_m|x_m)} + c$$

THAT DOES NOT DEPEND ON θ

Numbers that we can compute.

$z_i \in \{1 \dots K\}$

THE STEPS OF EM ARE:

INITIALIZATION FIX $\hat\theta^{(0)}$

FOR $k = 1, 2, \dots$

(1) COMPUTE

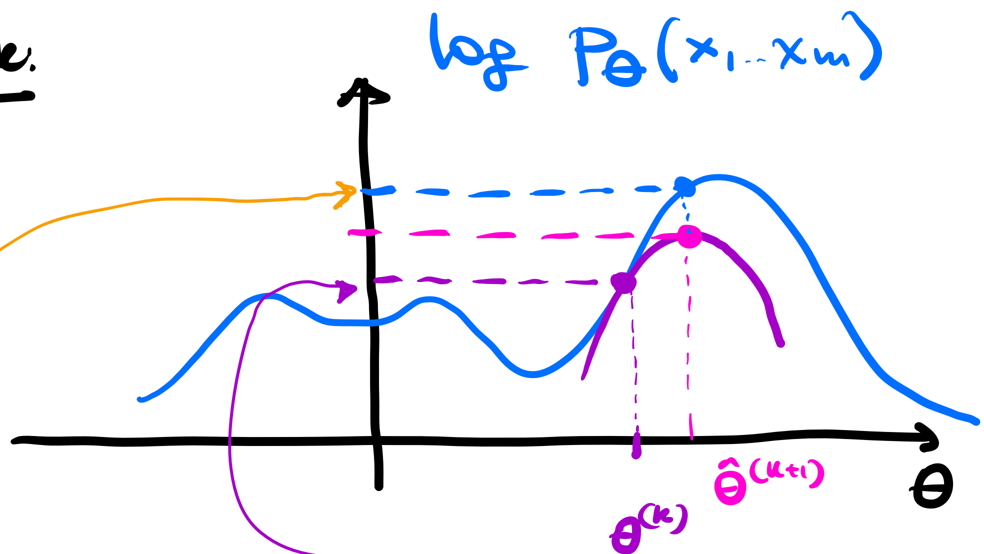
$Q(\theta, \hat\theta^{(k-1)})$ AS IN (2)

(EXPECTATION STEP)

(2) $\hat{\theta}^{(k)} = \arg\max Q(\theta, \hat{\theta}^{(k-1)})$

<u>MAXIMIZATION STEP</u>

<u>Remark:</u>



$$Q(\theta, \hat{\theta}^{(k)}) \leq \log P_\theta(x_1 \dots x_m) + c$$

$$Q(\hat{\theta}^{(k)}, \hat{\theta}^{(k)}) = \log P_{\hat{\theta}^{(k)}}(x_1 \dots x_m) + c$$

$$\Downarrow$$

$$\boxed{\log P_{\hat{\theta}^{(k+1)}}(x_1 \dots x_m) \geq \log P_{\hat{\theta}^{k}}(x_1 \dots x_m)}$$