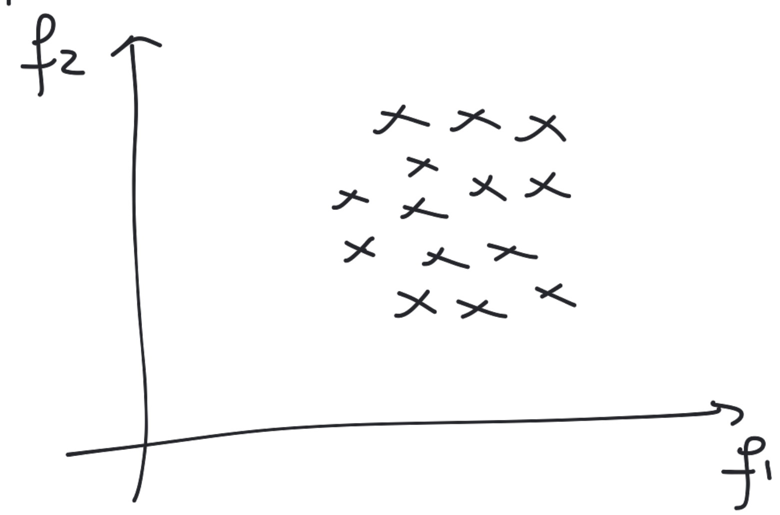
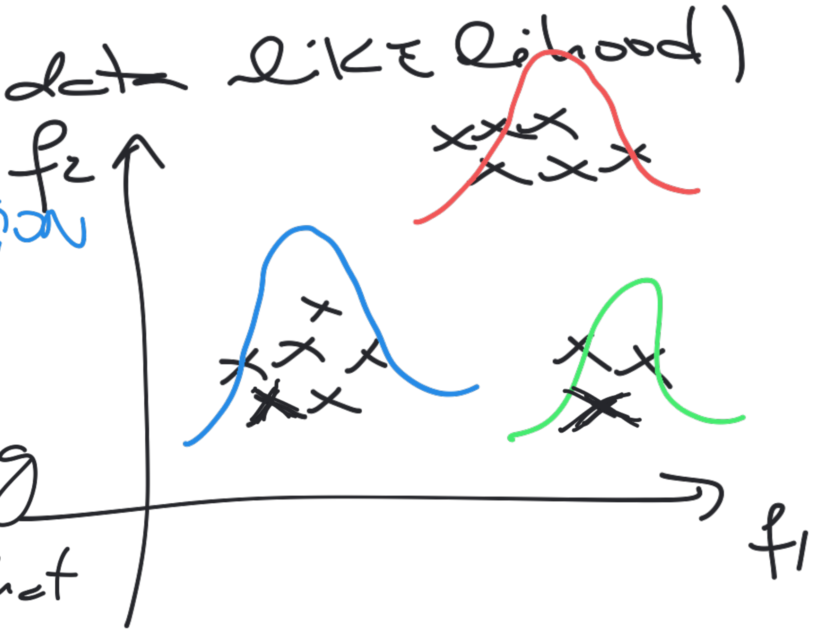
In previous problems, we worked data:



what if data samples (data likelihood) look like this? ESTIMATION DATA likelihood



- what do we "care" about having a distribution that fits the data?

- clustering
  find clusters/groups in the data

Gaussian Mixture Model (GMM)

$$X = \{x_i\}_{i=1}^{N}$$

$$\underline{\underline{p(x)}} = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

weight of each Gaussian component

$$0 \leq \pi_k \leq 1 \,, \quad \sum_{k=1}^{K} \pi_k = 1$$

1. Write down observed data likelihood

$$\mathcal{L}^{o} = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

2. Describe log-likelihood:

$$\mathcal{L} = \ln \mathcal{L}^{o}$$
$$= \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k) \right)$$

3. Optimize the parameters $\theta = \{\pi_k, \mu_k \Sigma_k\}_{k=1}^{K}$

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = 0, \quad \frac{\partial \mathcal{L}}{\partial \mu_k} = 0, \quad \frac{\partial \mathcal{L}}{\partial \Sigma_k} = 0$$

Expectation-Maximization (EM)

It helps in solving/optimizing a "difficult" log-likelihood function

→ Introduces Hidden Latent Variables

→ In GMM problem:

z_i = label of the Gaussian component which x_i was drawn from

$z_i = \{1, 2, 3, \dots, K\}$

Hidden variable

$$\mathcal{L}^{c} = \prod_{i=1}^{N} \pi_{z_i} \cdot \mathcal{N}(x_i \mid \mu_{z_i}, \Sigma_{z_i})$$

complete data likelihood

WE don't know z's!

EM algorithm poses the optimization function:

$$\boxed{Q(\theta, \theta^t) = E[\underbrace{\ln \mathcal{L}^e}\,|\, X, \theta^t]}$$

$\theta^t \equiv$ parameters at time t

FOR some initialization of parameters: θ^0

E-STEP: Find the z's, while holding the parameters values fixed

M-STEP: Find the best set of parameters θ while holding z's fixed

It stops until convergence criteria is met

- \# iterations
- ERROR MEASURE

Expected value of discrete R.V. X:

$$E[X] = \sum_x \underline{x} \cdot \underbrace{p_x(x)}_{\text{PMF}}$$

$$Q(\theta, \theta^t) = E[\ln \mathcal{L}^c \mid X, \theta^t]$$

$$= \sum_z \ln \mathcal{L}^c \cdot P(z \mid X, \theta^t)$$

Expected value on z's

$$= \sum_{z_i=1}^{K} \ln \mathcal{L}^c \cdot \underbrace{P(z_i \mid x_i, \theta^t)}_{\text{E-STEP}}$$

optimization

$$P(z_i \mid x_i, \theta^t) = \frac{\boxed{P(x_i \mid z_i, \theta^t)} \cdot \boxed{P(z_i \mid \theta^t)}}{\boxed{P(x_i \mid \theta^t)}}$$

Bayes' Rule

$$= \frac{P(x_i \mid \mu^t_{z_i}, \Sigma^t_{z_i}) \cdot \pi^t_{z_i}}{\sum_{z_i=1}^{K} P(x_i \mid \mu^t_{z_i}, \Sigma^t_{z_i}) \cdot \pi^t_{z_i}}$$

Law of Total Prob.

$$= C_{ik} \longleftarrow$$ MEMBERSHIPS OR RESPONSABILITIES

C_{ik} is a $N \times K$ matrix

$$\begin{array}{c|cccc} & 1 & 2 & \cdots & K \\ \hline x_1 & \circ & \circ & \cdots & \circ \\ x_2 & 0.1 & 0.8 & \cdots & 0.1 \\ x_3 & & & & \\ \vdots & & & & \\ x_N & & & & \end{array}$$

$$\boxed{\sum_K C_{ik} = 1}$$

This finishes the E-STEP.

M - STEP

Hold C_{ik} matrix fixed

$$Q(\theta, \theta^t) = \sum_{z_i=1}^{K} \ln \mathcal{L}^c \cdot C_{ik} \quad (z_i = k)$$

$$= \sum_{k=1}^{K} \ln \mathcal{L}^c \cdot C_{ik}$$

$$\ln \mathcal{L}^c = \ln \left(\prod_{i=1}^{N} \pi_{z_i} \cdot \mathcal{N}(x_i \mid \mu_{z_i}, \Sigma_{z_i}) \right)$$

$$= \sum_{i=1}^{N} \left(\ln(\pi_{z_i}) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sigma_{z_i}^2\right) - \frac{1}{2\sigma_{z_i}^2} \cdot \|x_i - \mu_{z_i}\|_2^2 \right)$$

$$Q(\theta, \theta^t) = \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\ln \pi_k - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\underline{\ln(\sigma_k^2)} - \underline{\frac{1}{2\sigma_k^2}\|x_i - \mu_k\|_2^2} \right) \cdot C_{ik}$$

assume

$$\Sigma_k = \sigma_k^2 \cdot I$$

$$\frac{\partial Q(\theta,\theta^t)}{\partial \mu_k} = 0$$

$$\Longleftrightarrow \sum_{i=1}^{N} \frac{1}{\sigma_k^2}(x_i - \mu_k) \cdot C_{ik} = 0$$

$$\Longleftrightarrow \sum_{i=1}^{N} (x_i - \mu_k) \cdot C_{ik} = 0$$

$$\Longleftrightarrow \sum_{i=1}^{N} x_i \cdot C_{ik} - \sum_{i=1}^{N} \mu_k \cdot C_{ik} = 0$$

$$\Longleftrightarrow \mu_k = \frac{\sum_{i=1}^{N} x_i \cdot C_{ik}}{\sum_{i=1}^{N} C_{ik}}$$

"weighted" mean of component k.

$$\frac{\partial Q(\theta, \theta^t)}{\partial \sigma_k^2} = 0$$

$$\Longleftrightarrow \sum_{i=1}^{N} \left(-\frac{1}{2\cancel{\sigma_k^2}} + \frac{2}{(2\sigma_k^2)^{\cancel{2}}} \cdot \|x_i - \mu_k\|_2^2 \right) \cdot c_{ik} = 0$$

$$\Longleftrightarrow \sum_{i=1}^{N} - \underline{\sigma_k^2} \cdot c_{ik} + \sum_{i=1}^{N} \|x_i - \mu_k\|_2^2 \cdot c_{ik} = 0$$

$$\Longleftrightarrow \sigma_k^2 \cdot \sum_{i=1}^{N} c_{ik} = \sum_{i=1}^{N} \|x_i - \mu_k\|_2^2 \cdot c_{ik}$$

$$\Longleftrightarrow \sigma_k^2 = \frac{\sum_{i=1}^{N} \|x_i - \mu_k\|_2^2 \cdot \underline{c_{ik}}}{\sum_{i=1}^{N} c_{ik}}$$

weighted variance for each component

$$\frac{\partial Q(\theta, \theta^t)}{\partial \pi_k} = 0$$

TO BE continued...