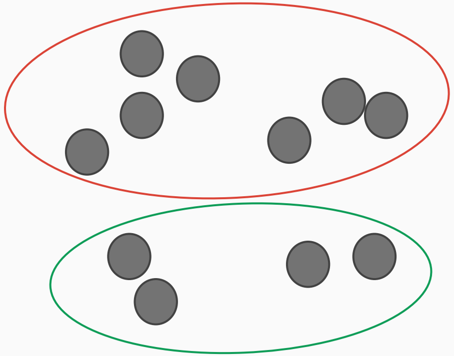
Mixture Models & Expectation Maximization

A brief introduction

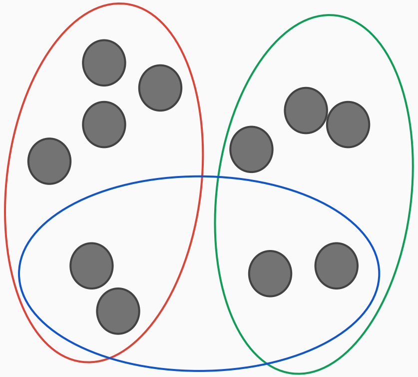
Cezar Sas

Clustering

HARD

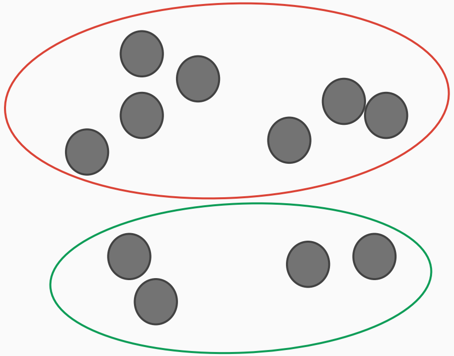


SOFT



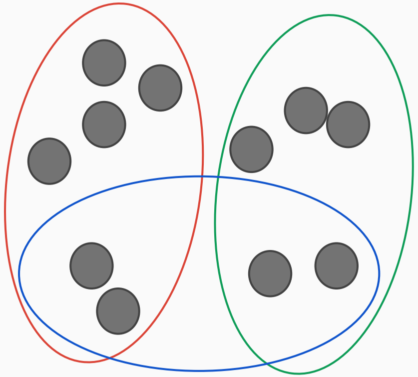
Clustering

HARD



Parameters: Centroids

SOFT

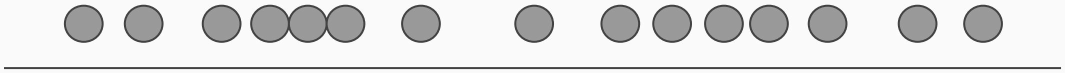


Parameters: Distributions Parameters (eg μ, σ^2)

Mixture Models

Problem:
- Find the distributions that produced the data

Issue:
- Unknown labels



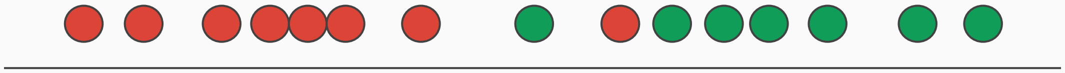
Mixture Models

Problem:

- Find the distributions that produced the data

Issue:

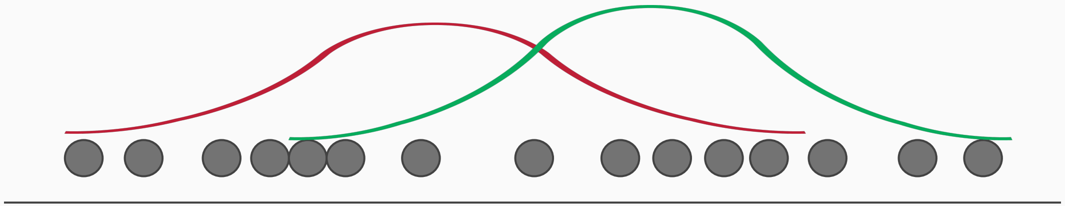
- Unknown labels



Mixture Models

Solution:

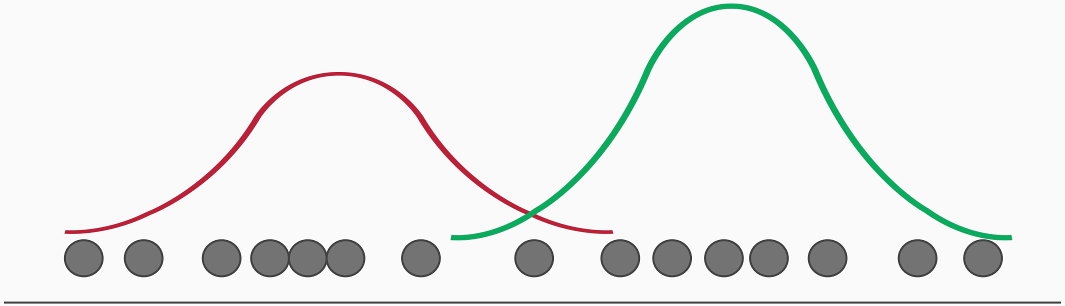
- Iteratively fit the data to find the best values of σ and μ for each distribution



Mixture Models

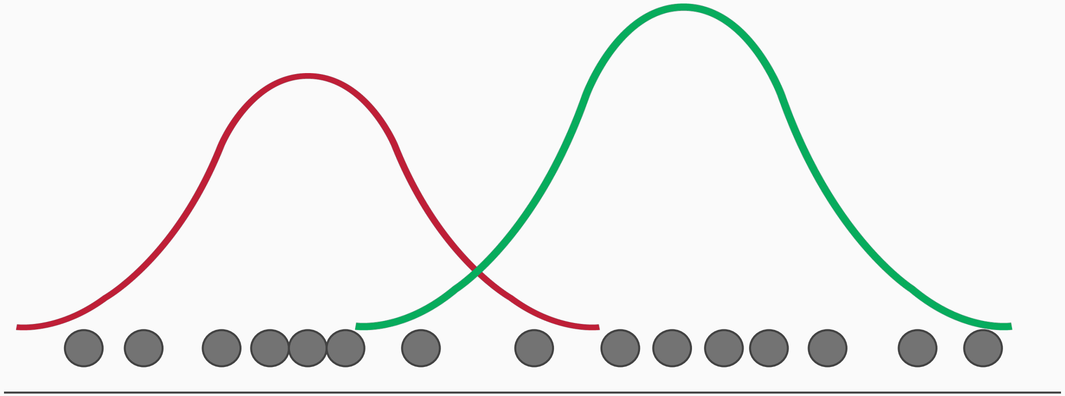
Solution:

- Iteratively fit the data to find the best values of σ and μ for each distribution



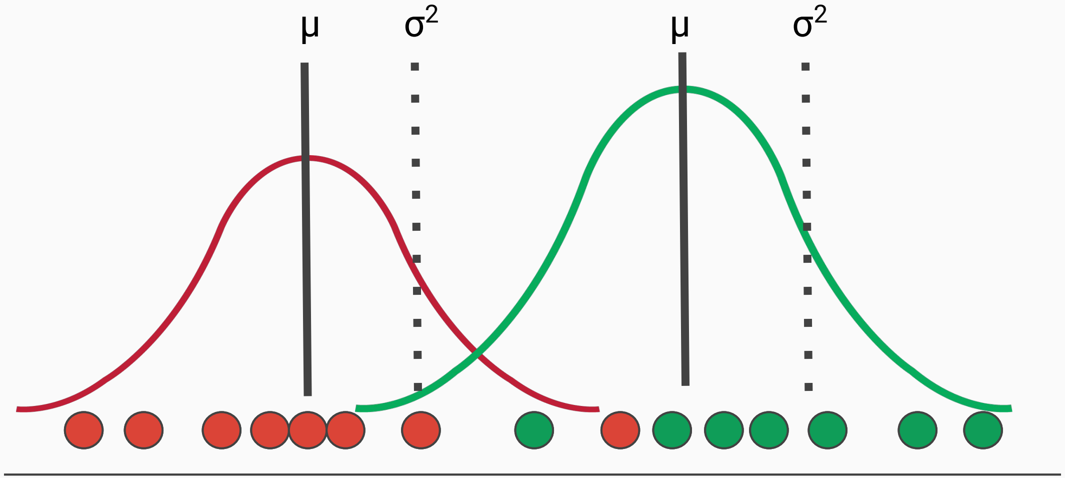
Mixture Models

Stop when no more changes are needed



Mixture Models

Stop when no more changes are needed



Expectation Maximization

EM

- Expectation

Calculate weights for each data point indicating the weighted probability of the point belonging to a distribution based on the likelihood of it being produced by a parameter.

- Maximization

Compute a better estimate for the parameters using the weight-adjusted data.

1. Initialize parameters μ, **Σ**, π
2. Expectation

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

3. Maximization

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \qquad N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right)\left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right)^{\mathrm{T}}$$

$$\pi_k^{\text{new}} = \frac{N_k}{N}$$

4. Maximize log likelihood

$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

* Equations from PRML book

Implementation

- Python
- Numpy
- Scipy
- Pandas*
- Plotting libraries

* or your favorite file reader method

Tips

1. Initialize the covariance matrix (NOT as Identity)
2. Wild singular matrices may appear
3. Comment your code
4. You are NOT required to implement the Distribution function and the PDF (use pre-existing functions)