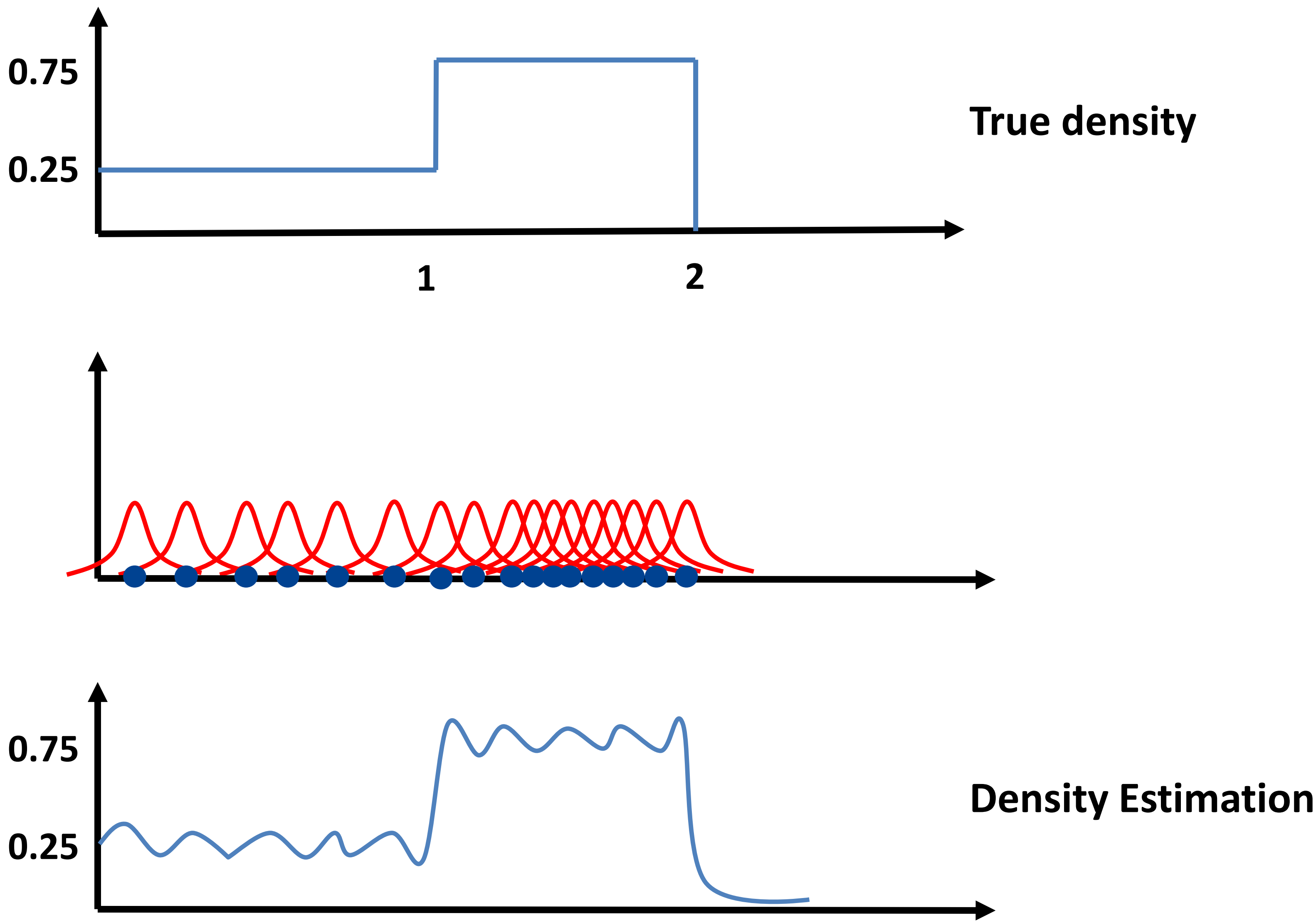
Pattern Classification

08. Gaussian Mixture Model

AbdElMoniem Bayoumi, PhD

Fall 2021

Recap: Kernel Density Estimator



Gaussian Mixture Model (GMM)

- Assume we have a small data set → not possible to estimate class conditionals using kernel density estimator

- Instead, we model each class conditional as a sum of multivariate Gaussian densities

Gaussian Mixture Model (GMM)

- The parameters, i.e., the mean vectors & covariance matrices, are determined so that this sum approximates as good as possible the given class conditional density

$$\hat{P}(\underline{X}) = \sum_{j=1}^{K} w_j \frac{e^{-\frac{1}{2}(\underline{X}-\underline{\mu}_j)^T \Sigma_j^{-1} (\underline{X}-\underline{\mu}_j)}}{(2\pi)^{\frac{N}{2}} det^{\frac{1}{2}}(\Sigma_j)}$$

$$= \sum_{j=1}^{K} w_j N(\underline{X}, \underline{\mu}_j, \Sigma_j)$$

$w_j \equiv$ represents the probability of each mixture component

$N(\underline{X}, \underline{\mu}_j, \Sigma_j) \equiv$ multi-variate Gaussian density with mean $\underline{\mu}_j$ and covariance Σ_j

Gaussian Mixture Model (GMM)

$$\hat{P}(\underline{X}) = \sum_{j=1}^{K} w_j N(\underline{X}, \underline{\mu}_j, \Sigma_j)$$

Condition:

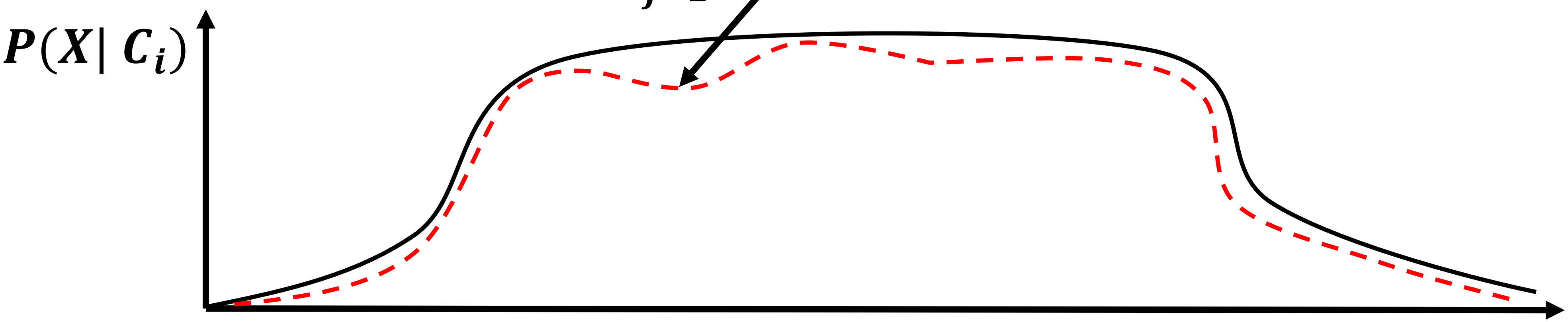
$$\sum_{j=1}^{K} w_j = 1$$

Because we need:

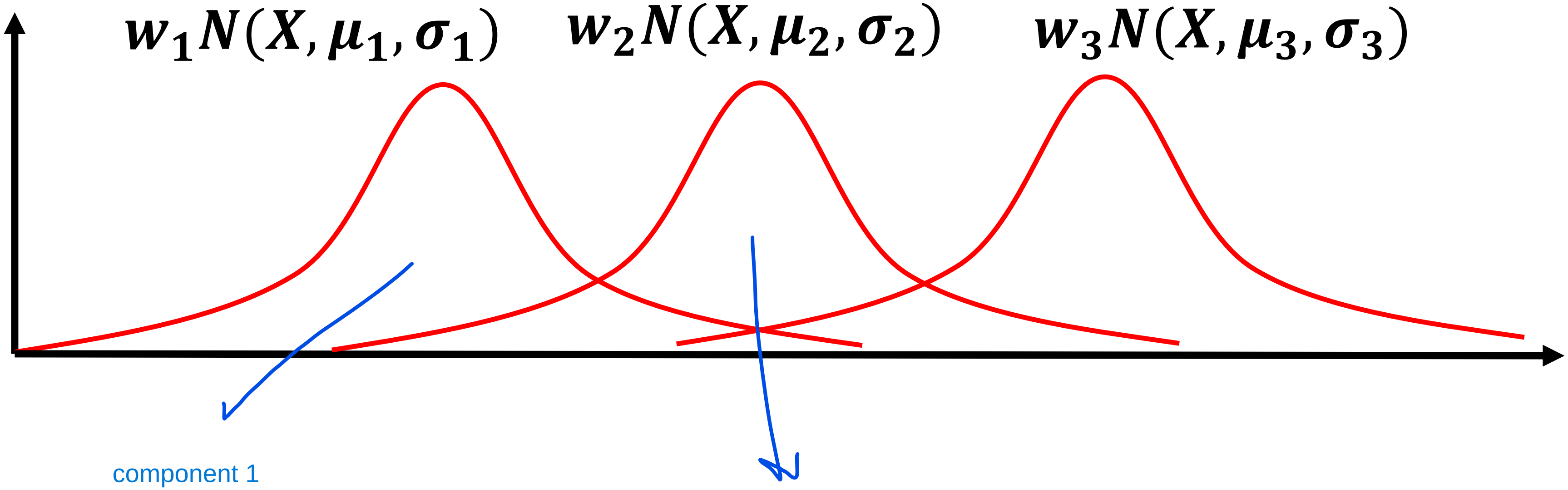
$$\int_{-\infty}^{\infty} \hat{P}(\underline{X}) = \sum_{j=1}^{K} w_j \underbrace{\int_{-\infty}^{\infty} N(\underline{X}, \underline{\mu}_j, \Sigma_j)}_{=1} = 1$$

1-D Example

$$\sum_{j=1}^{K} w_j N(X, \mu_j, \sigma_j) \rightarrow \text{approx. } P(X|\, C_i)$$



$w_1 N(X, \mu_1, \sigma_1)$ $w_2 N(X, \mu_2, \sigma_2)$ $w_3 N(X, \mu_3, \sigma_3)$



component 1

component2

Expectation–Maximization (EM)

- Apply EM algorithm, which is an iterative algorithm, to estimate the parameters of the GMM components

- For simplicity assume 2-component case ,i.e.,
  $\hat{P}\left(\underline{X}|C_i\right) = w\, N\left(\underline{X}, \underline{\mu}_1, \Sigma_1\right) + (1-w)\, N(\underline{X}, \underline{\mu}_2, \Sigma_2)$

Expectation–Maximization (EM)

1. Take initial guesses for the parameters: $w, \underline{\mu}_1, \Sigma_1, \underline{\mu}_2$ and Σ_2

2. **Expectation step:** compute the responsibilities:

$$\hat{\gamma}_m = \frac{\hat{w} N\left(\underline{X}(m), \underline{\hat{\mu}}_1, \hat{\Sigma}_1\right)}{\hat{w}\ N\left(\underline{X}(m), \underline{\hat{\mu}}_1, \hat{\Sigma}_1\right) + (1-\hat{w})\ N\left(\underline{X}(m), \underline{\hat{\mu}}_2, \hat{\Sigma}_2\right)}$$

$\hat{\gamma}_m$ represents the probability that $\underline{X}(m)$ is generated from component 1

3. **Maximization step:** compute the weighted means & covariance matrices:

$$\underline{\hat{\mu}}_1 = \frac{\sum_{m=1}^{M} \hat{\gamma}_m \underline{X}(m)}{\sum_{m=1}^{M} \hat{\gamma}_m},\ \underline{\hat{\mu}}_2 = \frac{\sum_{m=1}^{M} (1-\hat{\gamma}_m)\ \underline{X}(m)}{\sum_{m=1}^{M} (1-\hat{\gamma}_m)}$$

$$\hat{\Sigma}_1 = \frac{\sum_{m=1}^{M} \hat{\gamma}_m \left(\underline{X}(m)-\underline{\hat{\mu}}_1\right)\left(\underline{X}(m)-\underline{\hat{\mu}}_1\right)^T}{\sum_{m=1}^{M} \hat{\gamma}_m}\ ,\ \hat{\Sigma}_2 = \frac{\sum_{m=1}^{M} (1-\hat{\gamma}_m)\left(\underline{X}(m)-\underline{\hat{\mu}}_2\right)\left(\underline{X}(m)-\underline{\hat{\mu}}_2\right)^T}{\sum_{m=1}^{M} (1-\hat{\gamma}_m)}$$

$$\hat{w} = \frac{\sum_{m=1}^{M} \hat{\gamma}_m}{M}$$

4. Iterate steps 2 & 3 until convergence

Expectation–Maximization (EM)

$$\gamma_m = P\left(\underline{X}(m) \in \text{component } 1\right)$$

$$= \frac{P(comp.\,1)\ P\left(\underline{X}(m)|comp.\,1\right)}{P\left(\underline{X}(m)\right)}$$

apply Bayes rule

$$= \frac{P(comp.\,1)\ P\left(\underline{X}(m)|comp.\,1\right)}{P(comp.\,1)\ P\left(\underline{X}(m)|comp.\,1\right) + P(comp.\,2)\ P\left(\underline{X}(m)|comp.\,2\right)}$$

$$\equiv \frac{w\ N\left(\underline{X}(m),\underline{\mu}_1,\Sigma_1\right)}{w\ N\left(\underline{X}(m),\underline{\mu}_1,\Sigma_1\right)+(1-w)\ N\left(\underline{X}(m),\underline{\mu}_2,\Sigma_2\right)}$$

Issues with GMM

- Initialization:
  - EM is an iterative algorithm which is very sensitive to initial conditions:
  - Start from trash $\rightarrow$ end up with trash
  - Usually, we use the K-Means to get a good initialization

- Number of Gaussian Components:
  - Try different number of Gaussian components and choose the best based on validation set.

Midterm

- Midterm will be up to this slide, i.e., GMM is included

Acknowledgment

- These slides have been created relying on lecture notes of Amir Atiya and Mohand Saïd Allili