Modèles de melange (G.M.M)
Cours 2

1 Introduction

1.1 Clustering:

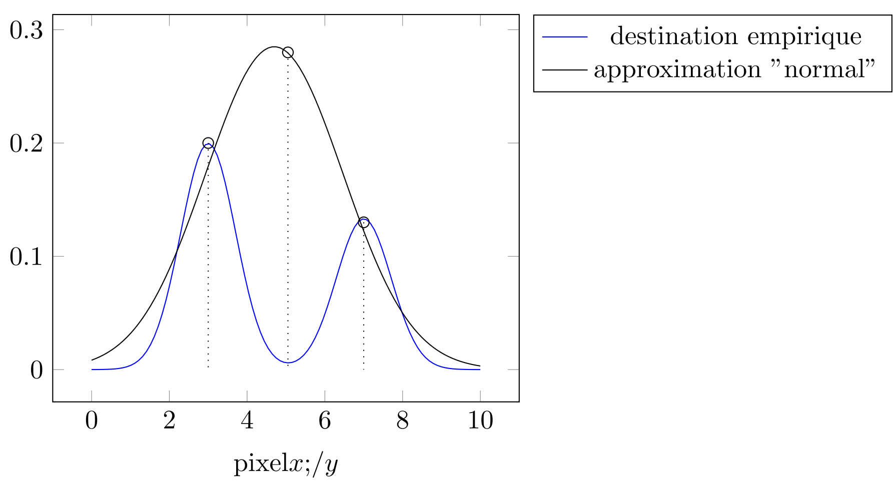
Apprentissage non-supervisé
$\mathcal{D} = (\underbrace{\vec{x_n}}_{\mathbb{R}^d})_{n=1}^{N}$, on fix K, un nombre de clusters

ex: K-means
Algo iteratif:

0) Choisir une distance *
1) Assignation
   pour chaque cluster: $\vec{m_k}$, $\vec{x_n} \in$ au cluster le plus proche (min distance $(\vec{x_n}, \vec{m_k})$)
2) Re-estimation
   recalculer les $(\vec{m_k})_{k=1}^{K}$

1.2 Estimation de distribution:

ex: classification (d'image)



→ augmenter la capacité du modèle
→ augmenter le nombre de paramètres

1.3 Mélange de Gaussienne (GMM):

K le nombre de Gaussienne / cluster

$$p(\vec{x_n}|\underbrace{\mathcal{O}}_{(\Pi_k,\vec{\mu_k},\Sigma_k)_{k=1}^K}) = \sum_{k=1}^{K} \underbrace{\Pi_k}_{\text{Poids du mélange}} \underbrace{\mathcal{N}(\vec{\mu_k},\Sigma_k)}_{\text{la gaussienne}}$$

<u>Estimer les paramètres du mélange</u>

$\rightarrow$ maximiser $\prod_{n=1}^{N} P(\vec{X} = \vec{x_n}|\mathcal{O})$
$\rightarrow$ maximiser $\log(\prod_{n=1}^{N} \underbrace{P(\vec{X} = \vec{x_n}|\mathcal{O})}_{\sum_{k=1}^{K}\Pi_k\mathcal{N}(\vec{\mu_k},\Sigma_k)})$

2 Algorithme E.M

•Algo iteratif qui cherche à maximiser

$$\log(P(\vec{X} = \vec{x_n}|\mathcal{O}))$$

•Introduire des variable <u>latentes</u> (cachéés):

$$\text{pour chaque } \vec{x} \rightarrow \vec{Z}$$

$$\vec{Z} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \vec{Z_k} = 1 \Leftrightarrow \vec{x} \in \text{ cluster k}$$

$\vec{Z}$: •pseudo affectation
•une vecteur latent
•inconnue $\Rightarrow \vec{Z}$ un vecteur aléatoire
•affectation "soft": un point peut appartenire à tous les clusters.

<u>Résumé des programes</u>
Introduction $\vec{Z}$ associé à $\vec{X}$
Si on souhaite maximiser

$$P(\vec{X}|\underline{\underline{\mathcal{O}}}) = \sum_{Z} P(\vec{X},\vec{Z}|\mathcal{O})$$

$$P(\vec{X}|\mathcal{O}) = \underbrace{\sum_{Z} P(\vec{X}|\vec{Z},\mathcal{O})}_{\mathcal{N}(\vec{\mu_k},\Sigma_k)} \underbrace{P(\vec{Z}|\mathcal{O})}_{\Pi_k}$$

si $\vec{Z}_k = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ \rightarrow k \\ \\ \\ \end{array}$

•$(\vec{X}, \vec{Z})$: données complètes

•$(\vec{X})$: données incomplètes

étape E(xpection):

→ connaître $\vec{Z}$ à $\mathcal{O}$ fixé

→ calcul la probabilité d'affectation: $P(\vec{Z}|\vec{X}, \mathcal{O})$

étape M(aximigation)

Les données sont complètes

⇒ calcule de $\mathcal{O}$, on "fixe" $\vec{Z}$

3 Optimisation variationelle

On souhaite maxmiser selon $\mathcal{O}$:

$$\log(P(\vec{X}|\mathcal{O}) = \sum_{\vec{Z}} q(\vec{Z}) \log \frac{P(\vec{X}, \vec{Z}|\mathcal{O})}{q(\vec{Z})} - \sum_{\vec{Z}} q(\vec{Z}) \log \frac{P(\vec{X}|\vec{Z}, \mathcal{O})}{q(\vec{Z})}$$

•Après l'introduction de $\vec{Z}$, on introduit un distribution auxiliaire sur $\vec{Z}$: $q(\vec{Z})$

rappel:

$$P(\vec{X}|O) = \frac{P(\vec{X}, \vec{Z}|\mathcal{O})}{P(\vec{X}|\vec{Z}, \mathcal{O})}$$

$$\Rightarrow \log P(\vec{X}|\mathcal{O}) = \log(P(\vec{X}, \vec{Z}|\mathcal{O}) - \log(P(\vec{X}|\vec{Z}, \mathcal{O})$$

c.à.d:

le 2eme terme

$$-\sum_{\vec{Z}} q(\vec{Z}) \log \left(\frac{P(\vec{Z}|\vec{X}, \mathcal{O})}{q(\vec{Z})} \right) = E_{\vec{Z}\ q(\vec{Z})} \left[\log \frac{P(\vec{Z}, \vec{X}|\mathcal{O})}{q(\vec{Z})} \right]$$

Divergence de Kullback-leiber (D_{KL})

$$D_{KL}(q(\vec{Z})||P(\vec{Z}|\vec{X}, \mathcal{O}))$$

deux distribution sur $\vec{Z}$

Divergence ≠ distance (asymétrique)

•1er terme:

$$E_{\vec{Z}\ q(\vec{Z})} \left[\log \frac{P(\vec{Z}, \vec{X}|\mathcal{O})}{q(\vec{Z})} \right]$$

→ ELBO (Evidence Lower Bound)

$$\log(P(\vec{X}|\mathcal{O})) = \underbrace{\mathcal{L}(\mathcal{O},q)}_{\text{ELBO } \geq \underbrace{\mathcal{L}(\mathcal{O},q)}_{\text{borne inf}}} + D_{KL}(q(\vec{Z})||P(\vec{Z}|\vec{X},\mathcal{O}))$$

3.1 Etape E

•Les paramètres sont fixés: $\mathcal{O} = \mathcal{O}^{\text{ old}}$
•Maximiser $\mathcal{L}(\mathcal{O}^{\text{ old}}, q)$

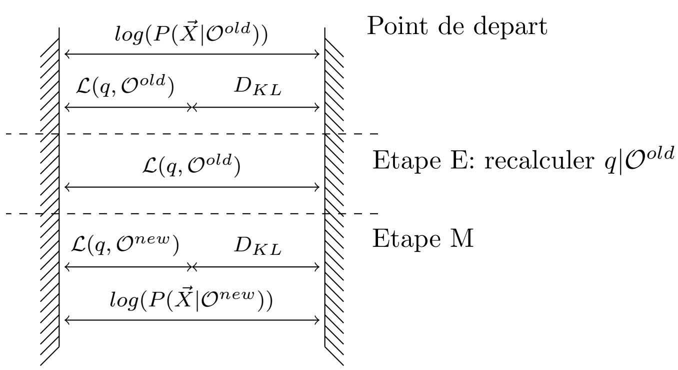
$$\rightarrow \mathcal{L}(\mathcal{O}^{\text{ old}}, q) = -D_{KL}(q(\vec{Z})||P(\vec{Z}|\vec{X},\mathcal{O}^{\text{ old}})) + \underbrace{\cancel{\log(P(\vec{X}|\mathcal{O}^{\text{ old}}))}}_{cte}$$

$$\Rightarrow q(\vec{Z}) = P(\vec{Z}|\vec{X},\mathcal{O}^{\text{ old}})$$

3.2 Etape M

Maximiser $\mathcal{L}$ selon $\mathcal{O}$ avec q fixé

Illustration:



https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm