2] Maximum Entropy formulation of GMM → stat. Phys.

- Let's define an energy cost associated with assigning data x_i to the cluster ℓ: $E_\ell(x_i)$

In average, the total cost is: $\langle E \rangle = \sum_i \sum_\ell E_\ell(x_i) \underbrace{p_r(x_i \in \text{clus } \ell)}_{p_\ell}$

~ clustering measure

→ Find a distribution: • it fixes in averages the energy cost
• with the highest possible entropy

$$S(\lambda, \beta) = -\sum_\ell p_\ell \log p_\ell + \underbrace{\beta\left(\sum_\ell p_\ell E_\ell - cte\right)} + \underbrace{\lambda\left(\sum_\ell p_\ell - 1\right)}$$

the entropy

imposing fix av. E with the lagrange parameter β

imposing normalis of p_ℓ with lagr. param. λ

- $\frac{\partial S}{\partial p_h} = -p_h - 1 + \beta E_\ell + \lambda = 0$
$\Leftrightarrow p_h = e^{1-\lambda} e^{-\beta E_h}$

- $\frac{\partial S}{\partial \lambda} = 0 \Rightarrow \sum_h \underbrace{e^{(1-\lambda)} e^{-\beta E_h}}_{p_h} = 1$

it fixes $e^{(1-\lambda)} = \frac{1}{\sum_h e^{-\beta E_h}}$

$$p_h(x) = \frac{e^{-\beta E_h(x)}}{\sum_\ell e^{-\beta E_\ell(x)}}$$

If we take: $E_h(x) = (x - \mu_h)^2$: $p_h(x) = \frac{e^{-\beta(x-\mu_h)^2}}{\sum_\ell e^{-\beta(x-\mu_\ell)^2}}$: the resp. (without ρ)

→ $\beta \leftrightarrow \frac{1}{2\sigma^2}$: fixed ∀ clusters.

If we max the free-energy: $-\beta F = \log(Z) = \log\left(\prod_i Z_i\right)$

$$= \sum_i \log\left(\sum_h e^{-\beta(x-\mu_h)^2}\right)$$

$$0 = \frac{\partial -\beta F}{\partial \mu_k} \; ? \; : \; \boxed{\sum_i \frac{(x_i - \mu_k)\, e^{-\beta (x_i - \mu_k)^2}}{\sum_\ell \exp\left[-\beta (x_i - \mu_\ell)^2\right]} = 0} \iff$$

max likelihood for GMM

$(\square)$

Phase transition in Learning

- when β is very small (high T, high variance)

the probability to belong to any cluster is $\sim \frac{1}{K}$ where K is the total number of cluster

$\Longrightarrow$ all $\vec{\mu}_k \sim$ center of mass $= \vec{0}$ $\left\{ \begin{array}{l} \text{] center the dataset} \\ \sum_i \vec{x}_i = 0 \end{array} \right.$

$\vec{\mu}_k = 0$ should be a solution of $(\square)$: $\underbrace{\sum_i \frac{\vec{x}_i \, \cancel{e^{-\beta \|\vec{x}_i\|^2}}}{\sum_\ell \cancel{e^{-\beta\|\vec{x}_i\|^2}}}}_{\frac{1}{K}\sum_i \vec{x}_i = 0} = 0$

Is it stable? (for small β)

Linear perturbation analysis : we add a small perturbation to μ: $\vec{\mu}_k = \vec{0} + \vec{\varepsilon}_k$

$\oplus \; \sum_k \vec{\varepsilon}_k = 0$

$(\square)$: $$\sum_i \frac{(\vec{x}_i - \vec{\varepsilon}_h) e^{-\beta\|\vec{x}_i - \vec{\varepsilon}_h\|^2}}{\sum_\ell e^{-\beta\|\vec{x}_i - \vec{\varepsilon}_\ell\|^2}} = 0$$

$$\Longleftrightarrow \quad \vec{\varepsilon}_h^{\,(t+1)} = \frac{\sum_i \vec{x}_i \left(e^{-\beta\|\vec{x}_i - \vec{\varepsilon}_h\|^2} \big/ \sum_\ell e^{-\beta\|\vec{x}_i - \vec{\varepsilon}_\ell\|^2} \right)}{\sum_j e^{-\beta\|\vec{x}_j - \vec{\varepsilon}_h\|^2} \big/ \sum_\ell e^{-\beta\|\vec{x}_j - \vec{\varepsilon}_\ell\|^2}} = \frac{\sum_i x_i \, p_h(\vec{x}_i)}{\sum_i p_h(\vec{x}_i)}$$

$p_h(\vec{x}_i)$

$$p_h(\vec{x}_i) = \frac{e^{-\beta\|\vec{x}_i - \vec{\varepsilon}_h\|^2}}{\sum_\ell e^{-\beta\|\vec{x}_i - \vec{\varepsilon}_\ell\|^2}} = \frac{\cancel{e^{-\beta\|\vec{x}_i\|^2}} \, e^{2\beta \vec{x}_i^T \vec{\varepsilon}_h - \beta\|\varepsilon_h\|^2}}{\sum_\ell \cancel{e^{-\beta\|\vec{x}_i\|^2}} \, e^{2\beta \vec{x}_i^T \vec{\varepsilon}_\ell - \beta\|\varepsilon_\ell\|^2}}$$

$\beta\|\varepsilon_h\|^2 \sim O(\varepsilon^2)$, $\beta\|\varepsilon_\ell\|^2 \sim O(\varepsilon^2)$

$$\underset{\text{small } \varepsilon}{\simeq} \frac{1 + 2\beta \vec{x}_i^T \cdot \vec{\varepsilon}_h}{\sum_\ell (1 - 2\beta \vec{x}_i^T \cdot \vec{\varepsilon}_\ell)} = \frac{1}{k}\left(1 + 2\beta \vec{x}_i^T \cdot \vec{\varepsilon}_h\right)$$

$(\square)$: $$\vec{\varepsilon}_h \simeq \frac{\sum_i \vec{x}_i (1 + 2\beta \vec{x}_i^T \cdot \vec{\varepsilon}_h)}{\sum_j (1 + 2\beta \vec{x}_j^T \vec{\varepsilon}_h)} = \frac{2\beta \sum_i \vec{x}_i \cdot \vec{x}_i^T \cdot \vec{\varepsilon}_h}{N\left(1 + 2\beta \frac{\sum_j \vec{x}_j^T \vec{\varepsilon}_h}{N}\right)} \simeq 2\beta \boxed{\frac{\sum_i \vec{x}_i \vec{x}_i^T}{N}} \vec{\varepsilon}_h$$

covariance matrix $=: C$

$$\vec{\varepsilon}_h^{\,(t+1)} \simeq 2\beta C \, \vec{\varepsilon}_h^{\,(t)}$$

- if "$2\beta C$" > 1 : the solution is unstable
- if "$2\beta C$" < 1 : ——— stable

The stability depends on the eigenvalues of C

I can always write: $\vec{\varepsilon}_h = \sum_i d_i \vec{u}_i$ where the $\vec{u}_i$ are eigenvectors of C

$$\vec{\varepsilon}_h^{\,(t+1)} = 2\beta \sum_i d_i \cdot c_i \cdot \vec{u}_i$$ where c_i are eigenvalues of C

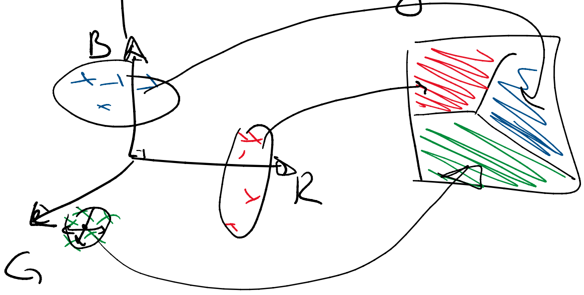
we have the threshold: $\boxed{2\beta_c \cdot c_{max} = 1}$ $\quad \beta_c = \frac{1}{2 c_{max}}$

$c_{max} = \max_i \{c_i\}$: max. eigenvalue of C

2 applications of GMM :

i) Segmented an image

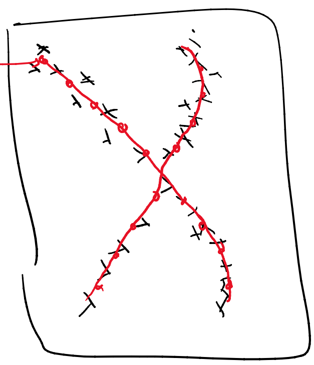
↳ compressed an image



ii) Principal graph

find the red curve from the datapt.

GMM centers



add a prior : $p(\Theta) \to p(\{\vec{\mu}_k\}_k) = e^{-\lambda \sum_{a,b} \boxed{b_{ab}} \overbrace{\|\vec{\mu}_a - \vec{\mu}_b\|^2}^{\text{altractive interact.}}}$

adjacency matrix

For the graph structure:

→ construct the minimum spanning tree

$b_{ab} : \begin{cases} 1 & \text{if } a \text{ is connected to } b \\ 0 & \text{otherwise} \end{cases}$