IAA EXAMEN 2020

EXERCICE 1

(1.) Il s'agit du pb k-means, provenant du domaine de la classification non-supervisée, plus précisément du clustering

ALGO :

INPUT : $T \in \mathbb{N}$ (nombre d'itérations), $\{x_i\}_{1 \leq i \leq n}$ (observations), k

INITIALISER : $\hat{\mu}_j \in \mathcal{X}$ (aléatoirement) $\forall j \in [k]$

FOR $t = 1$ TO T :

compute Voronoi partition $(C_1, \ldots, C_k)$ based on $(\hat{\mu}_1, \ldots, \hat{\mu}_k)$

$$\hat{C}_j \leftarrow \{X_1, \ldots, X_n\} \cap C_j \quad \forall j \in [k]$$

$$\hat{\mu}_j \leftarrow \frac{1}{|\hat{C}_j|} \sum_{x \in \hat{C}_j} x$$

OUT : $(C_1, \ldots, C_k)$

(2) Par déf. du partitionnement de Voronoi, on a $\forall j \in [k]$ et $\forall x \in C_j$:

$$\|x - \mu_j\|_{\ell_2}^2 \leq \|x - \mu_\ell\|_{\ell_2}^2 \quad \forall \ell \in [k]$$

Ainsi, $\forall x \in \mathbb{R}^d$,

$$\sum_{j=1}^{k} \|x - \mu_j\|_{\ell_2}^2 \mathbb{1}_{x \in C_j} = \|x - \mu_j\|_{\ell_2}^2 \overset{\forall \ell \in [k]}{\leq} \|x - \mu_\ell\|_{\ell_2}^2$$
$$= \sum_{\ell=1}^{k} \|x - \mu_\ell\|_{\ell_2}^2 \mathbb{1}_{x \in C_\ell}$$

En sommant sur les x_i, $i \in [n]$:

$$\sum_{i=1}^{n} \sum_{j=1}^{k} \|x_i - \mu_j\|_{\ell_2}^2 \mathbb{1}_{x \in C_j} \leq \sum_{i=1}^{n} \sum_{j=1}^{k} \|x_i - \mu_j\|_{\ell_2}^2 \mathbb{1}_{x \in C_j'}$$

(3) $$\frac{\partial}{\partial \mu} \sum_{i=1}^{n} p_i \|x_i - \mu\|_{\ell_2}^2 = 2\left(\sum_{i=1}^{n} p_i \mu - \sum_{i=1}^{n} p_i x_i\right) \overset{!}{=} 0$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i}$$

(notez que la fct est convexe infinie à l'infini et $p_i > 0 \ \forall i$)

(4) k-means: hard assignment, no notion of variance, implicity defines convex clusters through Voronoi partitioning
soft k-means: every point is assigned to every cluster with a certain probability, based on the assumption of Gaussian mixtures. (Weighted average)
When the covariance matrices for soft k-means are fixed to $\sigma^2 \mathbb{1}$ and $\sigma^2 \to 0$, soft k-means becomes k-means.

EXERCICE 2

(1) $\forall x \in \mathbb{R}^d$, $\mathbb{P}(Y=1 \mid X=x) = \mathbb{P}(\mathrm{sgn}(f^*(X)+\varepsilon) = 1 \mid X=x) =$

$= \mathbb{P}(f^*(X)+\varepsilon \geq 0 \mid X=x) = \mathbb{P}(-\varepsilon \leq f^*(X) \mid X=x) =$

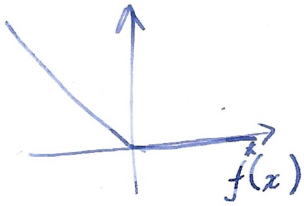
$= \mathbb{P}(\varepsilon \geq -f^*(X) \mid X=x) = 1 - \mathbb{P}(\varepsilon < -f^*(X) \mid X=x) =$

$= 1 - (1-e^{f^*(x)}) \mathbb{1}_{f^*(x) \leq 0} = \mathbb{1}_{f^*(x) \geq 0} + e^{f^*(x)} \mathbb{1}_{f^*(x) \leq 0} =$

$= p_x$ (on a bien $p_x \in]0,1]$ car si $f^*(x) \leq 0$, $e^{f^*(x)} \leq 1$)

(2) $-\ln(\mathbb{P}(Y=1 \mid X=x)) = -\ln(\mathbb{1}_{f^*(x) \geq 0} + e^{f^*(x)} \mathbb{1}_{f^*(x) \leq 0}) =$

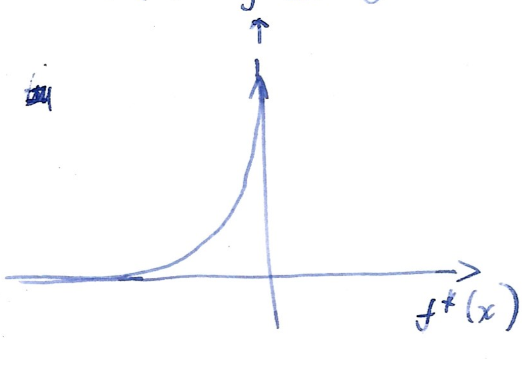
$= \begin{cases} 0 \text{ si } f^*(x) \geq 0 \\ -f^*(x) \text{ si } f^*(x) \leq 0 \end{cases} = -f^*(x) \mathbb{1}_{f^*(x) \leq 0} = \ell_1(f^*(x))$



elle est convexe

(3) $-\ln(\mathbb{P}(Y=-1 \mid X=x)) = -\ln(1 - \mathbb{P}(Y=1 \mid X=x)) =$

$= -\ln(1 - \mathbb{1}_{f^*(x) > 0} - e^{f^*(x)} \mathbb{1}_{f^*(x) \leq 0}) = \begin{cases} -\ln(0) \text{ si } f^*(x) > 0 \\ -\ln(1-e^{f^*(x)}) \text{ si } f^*(x) \leq 0 \end{cases}$



elle est aussi convexe

4. on cherche à maximiser $l_{(X,Y)_1^n}(\beta) = \sum_{i=1}^n \ln(g_\beta(X_i, Y_i))$

où $g_\beta(X, -1) = f_\beta(X)(1-p_X)$, $g_\beta(X, 1) = f_\beta(X) p_X$

le pb est donc $\max_{\beta \in \mathbb{R}^d} \sum_{i=1}^n \mathbb{1}_{Y_i=1} [\ln(\beta^T X_i) + \ln(p_{X_i})]$

$$+ \mathbb{1}_{Y_i=-1} [\ln(\beta^T X_i) + \ln(1-p_{X_i})]$$

Ceci est équivalent à

$$\boxed{\min_{\beta \in \mathbb{R}^d} - \left\{ \sum_{i=1}^n \mathbb{1}_{Y_i=1} [\ln(\beta^T X_i) + \ln(p_{X_i})] + \mathbb{1}_{Y_i=-1} [\ln(\beta^T X_i) + \ln(1-p_{X_i})] \right\}}$$

$$\text{s.t.} \begin{cases} -\ln(p_{X_i}) = -\beta^T X_i \mathbb{1}_{\beta^T X_i < 0} & \forall i \in [n] \\ -\ln(1-p_{X_i}) = -\ln(1-e^{\beta^T X_i}) \mathbb{1}_{\beta^T X_i < 0} + \chi_{\beta^T X_i \geq 0} & \forall i \in [n] \end{cases}$$

5. D'après les question 2 et 3, les fonctions contraintes sont convexes, et $-\ln$ (fonction linéaire) est convexe. Si $\beta^T X_i \geq 0$, $\chi_{\beta^T X_i} = \infty$ donc le minimum n'existe pas. Si $\beta^T X_i < 0$ $\forall i \in [n]$, on a

$$\nabla_\beta l_{(X,Y)_1^n}(\beta) = \sum_{i=1}^n \mathbb{1}_{Y_i=1} \left(X_i \left(1 - (\beta^T X_i)^{-1}\right)\right) + \mathbb{1}_{Y_i=-1} \left(X_i \left(\frac{e^{\beta^T X_i}}{1-e^{\beta^T X_i}} - (\beta^T X_i)^{-1} \right)\right)$$

EXERCICE 3

PARTIE A

1. $\mathbb{E}[e^{i\omega^T u}] = \int_{\mathbb{R}^d} \frac{d\omega}{(2\pi)^{d/2}\sqrt{2\gamma}} \cancel{\exp\{-\frac{\omega^2}{}\}}$

$$\exp\left\{-\frac{\omega^2}{2\gamma}\right\} \exp\{i\omega^T u\} =$$

$$= \underbrace{\int_{\mathbb{R}^d} \frac{d\omega}{(2\pi)^{d/2}\sqrt{2\gamma}} \exp\left\{-\frac{1}{2\gamma}(\omega - \gamma i u)^2\right\}}_{\substack{= 1 \text{ car } \omega \sim \mathcal{N}(0, 2\gamma \mathbb{1}_d) \text{ et translation ne} \\ \text{change pas l'intégrale sur tout } \mathbb{R}^d}} \exp\{-\gamma u^2\} =$$

$= e^{-\gamma \|u\|_{\ell_2}^2}$ autrement dit $e^{\gamma \|u\|_{\ell_2}^2} \mathbb{E}[e^{i\omega^T u}] = 1$

Notez que $\mathbb{E}[e^{i\omega^T u}] = \mathbb{E}[\cos(\omega^T u) + i\sin(\omega^T u)] = e^{-\gamma \|u\|_{\ell_2}^2}$
donc, puisque $e^{-\gamma \|u\|_{\ell_2}^2} \in \mathbb{R}$, $\mathbb{E}[\cos(\omega^T u)] = e^{-\gamma \|u\|_{\ell_2}^2}$

2.
$$2\mathbb{E}[\cos(\omega^T x + \varphi)\cos(\omega^T y + \varphi)] = 2\int_{\mathbb{R}^d} d\omega \int_0^{\pi} d\varphi \frac{1}{(2\pi)^{d/2}\sqrt{2\gamma}} e^{-\frac{\omega^2}{2\gamma}} \cdot \frac{[\cos(\omega^T(x+y) + 2\varphi) + \cos(\omega^T(x-y))]}{2} =$$

$$= \int_{\mathbb{R}^d} d\omega \frac{1}{(2\pi)^{d/2}\sqrt{2\gamma}} e^{-\frac{\omega^2}{2\gamma}} \cos(\omega^T(x-y)) = \mathbb{E}[\omega^T(x-y)]$$

car cos est 2π-périodique et on intègre 2φ de 0 à π

4. Posons $\frac{1}{N} S_N = \frac{2}{N} \sum_{i=1}^{N} \cos(\omega_i^T x + \varphi_i) \cos(\omega_i^T y + \varphi_i)$

Alors, puisque $\omega_i \sim \omega$ iid, $\varphi_i \sim \varphi$ iid,

$$\frac{1}{N} \mathbb{E}[S_N] = 2\,\mathbb{E}[\cos(\omega^T x + \varphi)\cos(\omega^T y + \varphi)] =$$

$$\underset{\text{question 2}}{=} \mathbb{E}[\cos(\omega^T (x-y))] \underset{\text{question 1}}{=} e^{-\gamma \|x-y\|_2^2}$$

Par Hoeffding, il s'ensuit que

$$\mathbb{P}(|S_N - \mathbb{E}[S_N]| \geq N\varepsilon) \leq 2e^{-\frac{\varepsilon^2 N^2}{2Na^2}} = 2e^{-\frac{\varepsilon^2 N}{8}}$$

~~tat~~ avec $a = 2$ car $\cos(\omega_i^T x + \varphi_i)\cos(\omega_i^T y + \varphi_i) \in [-1, 1]$ $\forall i \in [n]$

5. Posons $\frac{1}{N}|S_N - \mathbb{E}[S_N]| = R$. On veut montrer que

$$\mathbb{P}\left(R \leq \sqrt{\frac{16}{N} \ln\left(\frac{n+1}{\delta}\right)}\right) \geq 1-\delta \quad \text{où } \delta \in]0,1[$$

autrement dit, $\mathbb{P}\left(R \geq \sqrt{\frac{16}{N} \ln\left(\frac{n+1}{\delta}\right)}\right) \leq \delta$

en posant $\delta = 2e^{-\frac{\varepsilon^2 N}{8}} \in]0,1[$ (puisque $\frac{\varepsilon^2 N}{8} > 1$)

on obtient $\varepsilon = \sqrt{\frac{8}{N} \ln\left(\frac{2}{\delta}\right)} \underset{n>1}{\leq} \sqrt{\frac{16}{N} \ln\left(\frac{n+1}{\sqrt{\delta}}\right)}$

et en remarquant que R est positive, on obtient le résultat.

PARIE B

1. D'une part on a $tr(Var(\omega)) = tr(\mathbb{E}[\|\omega\|_{\ell_2}^2] - \cancel{\mathbb{E}[\omega]^2}^{\,0})$

$\underset{\uparrow}{\geq} tr(\mathbb{E}[\omega_N^2]) = tr(\sigma^2) = \sigma^2$ (trace d'un scalaire)

trace est une fct linéaire et toutes les composantes sont ≥ 0

D'autre part on a $tr(Var(\omega)) = tr(2\gamma \mathbb{1}_d) = 2\gamma d$

② $|f_N(u)| - |f_N(U_i)| \leq |f_N(u) - f_N(U_i)| \underset{\substack{\uparrow \\ f_N \text{ est } L_N\text{-lipsch}}}{\leq} \overset{\leq \varepsilon/2r}{L_N} \underbrace{\|u - U_i\|_{\ell_2}}_{\leq r}$

$$\leq \frac{\varepsilon}{2r} \cdot r = \frac{\varepsilon}{2}$$

donc, puisque $|f_N(U_i)| \leq \varepsilon/2$, $|f_N(u)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$

③ La question précédente nous renseigne que

$$[|f_N(u)| \leq \varepsilon] \supset [|f_N(U_i)| \leq \varepsilon/2 \text{ et } L_N \leq \tfrac{\varepsilon}{2r}]$$

donc $\mathbb{P}(|f_N(u)| \leq \varepsilon) \geq \mathbb{P}(|f_N(U_i)| \leq \frac{\varepsilon}{2}, L_N \leq \frac{\varepsilon}{2r})$

$\Leftrightarrow \mathbb{P}(|f_N(u)| > \varepsilon) \leq \mathbb{P}(|f_N(U_i)| > \frac{\varepsilon}{2} \text{ \underline{ou} } L_N > \frac{\varepsilon}{2r})$

④ D'ailleurs, $\mathbb{P}(\sup_{u \in \mathcal{U}} |f_N(u)| \leq \varepsilon) = \mathbb{P}(|f_N(u)| \leq \varepsilon \quad \forall u \in \mathcal{U})$.

On a, $\forall u \in \mathcal{U}\ \exists i \in [\![1, T]\!]$ tq $\|u - \mathcal{U}_i\|_{\ell_2} \leq r$

En rassemblant,

$$\mathbb{P}(\sup_{u \in \mathcal{U}} |f_N(u)| > \varepsilon) \leq \sum_{i=1}^{T} \mathbb{P}(|f_N(U_i)| > \tfrac{\varepsilon}{2}) + \mathbb{P}(L_N > \varepsilon/2r)$$