Elisabeth Bourgeois
MS BD - promo 2022

31 mars 2022

Evaluation en forme de révision pour le module "Fondement mathématique pour l'IA" - page 1/18

Exercice 1 : moments d'ordre 1 et 2

Question a.

* Je propose $\boxed{\hat{\mu}_k = \frac{1}{k} \sum_{i=1}^{k} X_i}$ en supposant $E(X_i) = E(X)$ $\forall i \in [1, k]$

* Justification :
  - Notons $\mu = E(X)$ inconnue et $\sigma^2 = V(X)$.
  - Remarquons que $\sigma^2 = V(X) = V(X_i)$ $\forall i \in [1, k]$ du fait que la variance d'une somme de variables indépendantes est égale à la somme des variances.
  - Posons $\hat{\mu}_k = \frac{1}{k} \sum_{i=1}^{k} X_i$ la moyenne empirique des $\{X_i\}_{i \in [1,k]}$

Montrons que $\hat{\mu}_k$ est un estimateur convergent de μ :

→ D'une part $\boxed{E(\hat{\mu}_k)} = \frac{1}{k} \sum_{i=1}^{k} E(X_k) = \frac{1}{k} \times k \times E(X) = \boxed{E(X)}$ (1)

⚠ (en supposant $E(X_i) = E(X)$ $\forall i \in [1, k]$)

→ D'autre part, en appliquant l'inégalité de Bienaymé-Tchebycheff :

$$\text{Prob}\left(|\hat{\mu}_k - E(\hat{\mu}_k)| \geq \varepsilon\right) \leq \frac{\text{Var}(\hat{\mu}_k)}{\varepsilon^2} \quad \forall \varepsilon > 0$$

$$\text{Or } \text{var}(\hat{\mu}_k) = \frac{\text{Var}\left(\sum_{i=1}^{k} X_i\right)}{k^2} = \frac{k\sigma^2}{k^2} = \frac{\sigma^2}{k} \xrightarrow[k \to \infty]{} 0$$

$\text{Prob}(|\hat{\mu}_k - E(\hat{\mu}_k)| \geq \varepsilon)$ est donc majorée par une suite qui tend vers 0, donc $\text{Prob}(|\hat{\mu}_k - E(\hat{\mu}_k)| \geq \varepsilon) \to 0$

c.à.d $\boxed{\hat{\mu}_k \xrightarrow[\text{en probabilité}]{} E(\hat{\mu}_k) = \mu}$ d'après (1)

Conclusion : pour k grand, $var(\hat{\mu}_k)$ est petite
et $E(\hat{\mu}_k)$ se concentre autour de
l'espérance de X.

Question b :

Estimons μ avec $\hat{\mu}_k$. Notons $\boxed{\widehat{Var}_k^*(X) = \frac{1}{k}\sum_{i=1}^{k}(X_i - \hat{\mu}_k)^2}$

Soit $i \in [\![1,k]\!]$,

$$E[(X_i - \hat{\mu}_k)^2] = E[(X_i - \mu) - (\hat{\mu}_k - \mu)]^2 = V(X_i) + V(\hat{\mu}_k) - 2E[(X_i-\mu)(\hat{\mu}_k - \mu)]$$

$$= \sigma^2 + \frac{\sigma^2}{k} - 2E\left[(X_i - \mu) \times \frac{1}{k}\sum_{j=1}^{k}(X_j - \mu)\right]$$

$$= \sigma^2 + \frac{\sigma^2}{k} - \frac{2}{k}E[(X_i - \mu)^2] = \sigma^2\left(1 + \frac{1}{k} - \frac{2}{k}\right) = \frac{k-1}{k}\sigma^2$$

D'où $E(\widehat{Var}_k^*(X)) = E\left(\frac{1}{k}\sum_{j=1}^{k}(X_j - \hat{\mu}_k)^2\right) = \frac{1}{k}\sum_{j=1}^{k}\underbrace{E[(X_j - \hat{\mu}_k)^2]}_{=\frac{k-1}{k}\sigma^2}$

ie. $\boxed{E(\widehat{Var}_k^*(X)) = \frac{k-1}{k}\sigma^2}$ et donc $\boxed{\widehat{Var}_k^*(X) \text{ biaisé !}}$

Alors que $\widehat{Var}_k(X) = \frac{1}{k-1}\sum_{j=1}^{k}(X_j - \hat{\mu}_k)^2 = \frac{k}{k-1}\widehat{Var}_k^*(X)$

et donc $E(\widehat{Var}_k(X)) = E\left(\frac{k}{k-1}\widehat{Var}_k^2(X)\right) = \frac{k}{k-1}\left(\frac{k-1}{k^2}\right)\sigma^2$

ie $\boxed{E(\widehat{Var}_k(X)) = \sigma^2}$ $\rightarrow$ SANS BIAIS !

Question c :

Si μ connue, je propose

$$\boxed{\widehat{Var}_k^{\#}(X) = \frac{1}{k}\sum_{i=1}^{k}(X_i - \mu)^2}$$

(cf justification page 3)

Evaluation, page 3 / 18
(Bourgeois)

31 Mars 2022

Justification : $E\left(\widehat{Var}_k^{\#}(X)\right) = \frac{1}{k^2} \sum_{i=1}^{k} var\left((X_k - \mu)^2\right) = \frac{1}{k} \sum_{i=1}^{k} var(X_k) = \sigma^2$

Par ailleurs, les variables $(X_k - \mu)^2$ étant indépendantes

$$var\left(\widehat{Var}_k^{\#}(X)\right) = \frac{1}{k^2} \sum_{i=1}^{k} var\left((X_k - \mu)^2\right)$$

$$= \frac{1}{k}\left(E\left((X-\mu)^4\right) - E\left((X-\mu^2)^2\right)\right)$$

$$= \frac{1}{k}\left(E\left((X-\mu)^4\right) - \sigma^4\right) \xrightarrow[k \to +\infty]{} 0$$

Donc $\widehat{Var}_k^{\#}(X)$ est un estimateur convergent

Le théorème central-limite nous permet de conclure, pour k assez grand (loi asymptotique de $\widehat{Var}_k^{\#}(X)$)

$$\widehat{Var}_k^{\#}(X) \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}\left(\sigma^2, \frac{E\left((X-\mu)^4\right) - \sigma^4}{n}\right)$$

Exercice 2 : corrélation et coefficient de corrélation

Question a :

Non, on ne peut rien affirmer !

Une telle quantité entre deux variables aléatoires ne mesure que les écarts conjoints par rapport à leurs espé-rances respectives.

Pour l'interpréter, il convient de la normaliser sur le produit de l'écart-type de chacune des 2 v.a.

Néanmoins si $E(\bar{X}\bar{Y}) > 0$, on pourra affirmer que les 2 var évoluent dans le même "sens"

Question b :

Notons $cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right)$, $V(X) = \sigma_X$, $V(Y) = \sigma_Y$

Montrons que $\rho \in [-1;1]$:

Soit f une fonction définie sur $\mathbb{R}$ telle que $\alpha \mapsto f(\alpha) = V(-\alpha X + Y)$

$$\forall \alpha \in \mathbb{R},\ f(\alpha) = V(-\alpha X) + 2\,cov(-\alpha X, Y) + V(Y)$$
$$= \alpha^2 V(X) - 2\alpha\, cov(X,Y) + V(Y)$$

f est une fonction polynomiale de degré 2 en α

Elle est toujours positive par propriété des variances

Son discriminant Δ est donc négatif ou nul :

$$\Delta = 4(cov(X,Y))^2 - 4V(X)V(Y) \leq 0$$

donc $(cov(X,Y))^2 \leq V(X)\,V(Y)$,

et en prenant les racines carrées, on obtient :

$$|cov(X,Y)| \leq \sqrt{V(X)}\,\sqrt{V(Y)}$$

c.à.d

$$\boxed{|\rho| \leq 1}$$ (ou encore $\rho \in [-1;1]$)

(inégalité de Cauchy-Schwarz)

Question c.

Non.

Justification :

Soient X et Y deux variables aléatoires discrètes* indépendantes, définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$

univers — événements — mesure de probabilité

Admettons que X et Y admettent un moment d'ordre 2, alors :

→ D'une part, $cov(X,Y) = E((X-E(X))(Y-E(Y)))$

$$= E(XY - E(X).Y - E(Y).X + E(X).E(Y))$$

(par linéarité de l'espérance)

$$= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y)$$

(1) $\boxed{cov(X,Y) = E(XY) - E(X).E(Y)}$ (formule de HUYGENS)

(* : le passage discret → continue se fait aisément en substituant les $\sum_\Omega$ par $\int_\Omega$)

31 Mars 2022

D'autre part $\underline{E(XY)} = \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xy\mathbb{P}([X=x] \cap [Y=y])$

$E(XY) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy\mathbb{P}(X=x)\mathbb{P}(Y=y)$ par indépendance de X et Y

$\boxed{E(XY)} = \left(\sum_{x \in X(\Omega)} x\mathbb{P}(X=x)\right)\left(\sum_{y \in Y(\Omega)} y.\mathbb{P}(Y=y)\right) \boxed{= E(X).E(Y)}$ (2)

$\underline{\underline{cl}}$: (1) et (2) donne $\boxed{cov(X,Y) = 0}$

L'indépendance impose donc la décorrélation.
La réponse de l'énoncé est donc $\underline{\text{NON}}$.

Question d.

* NON, la corrélation implique la dépendance.
* Justification : l'indépendance impliquant la décorrélation, en contraposant cette proposition, ceci reste vrai $((A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A))$:
  la $\underbrace{\text{non décorrélation}}_{(= \text{corrélation !})}$ implique la $\underbrace{\text{non indépendance}}_{(= \text{dépendance !})}$

Question e :

On a montré à la question c) que l'indépendance impliquait la décorrélation. Montrons que dans le cas suivant, la décorrélation impliquait l'indépendance.
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.
Soit (X,Y) un couple gaussien centré (ie un vecteur gaussien standard de dim 2 où X et Y sont 2 v.a gaussiennes $\mathcal{N}(0,1)$ définies sur $(\Omega, \mathcal{F}, \mathbb{P})$)

ie $E((X,Y)) = 0$
$cov((X,Y)) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Bourgeois

Soit $\varphi_X : t \mapsto \varphi_X(t) = E(e^{itX})$ la fonction caractéristique de X
$\mathbb{R} \to \mathbb{C}$

Soit $\varphi_Y : t \mapsto \varphi_Y(t) = E(e^{itY})$ la fonction caractéristique de Y
$\mathbb{R} \to \mathbb{C}$

et $\varphi_{(X,Y)} : (t_1, t_2) \mapsto \varphi_{(X,Y)}(t_1, t_2) = E\left(e^{i(t_1 X + t_2 Y)}\right)$
$\mathbb{R}^2 \to \mathbb{C}$

la fonction caractéristique du couple (X, Y).

L'indépendance de X et Y est établie si

$$\boxed{\varphi_{(X,Y)}(t_1, t_2) = \varphi_X(t_1) \cdot \varphi_Y(t_2)}$$

ceci se montre par le théorème de Fubini :

$$\int_{X \times Y} = \int_X \left(\int_Y \right) = \int_Y \left(\int_X \right)$$

et par "si X est une v.a avec f sa densité, $\varphi : \mathbb{R} \to \mathbb{R}$ C^0 $\int_{-\infty}^{+\infty} |\varphi(x)| f(x)\,dx$ converge alors $E(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x) f(x)\,dx$"

$\Gamma_{(X,Y)}$ = la matrice de covariance de (X, Y) est diagonale : $\begin{bmatrix} \sigma^2_X & 0 \\ 0 & \sigma^2_Y \end{bmatrix}$

(avec $\sigma_X = var(X)$, $\sigma_Y = var(Y)$)

car $cov(X, Y) = cov(Y, X) = 0$

On peut écrire

$$\varphi_{(X,Y)}(t_1, t_2) = \exp\left(-\frac{1}{2} \|(t_1, t_2)\|^2\right)$$

$$= \exp\left(-\frac{1}{2}(t_1^2 + t_2^2)\right) = \exp\left(-\frac{1}{2} t_1^2\right) \times \exp\left(-\frac{1}{2} t_2^2\right)$$

$$\boxed{\varphi_{(X,Y)}(t_1, t_2) = \varphi_X(t_1) \times \varphi_Y(t_2)}$$

d'où X et Y indépendants.

Evaluation page 7 / 18
(Bourgeois)

31 Mars 2022

Partie / Exercice 3 : Corrélation et dépendance

Soit Y et f_Y sa densité telle que $t \mapsto f_Y(t) = \begin{cases} \frac{1}{5} & \text{si } t < -\frac{4}{5} \\ \frac{1}{5} - \frac{1}{2^4} & \text{si } t \in [-\frac{4}{5}, \frac{1}{5}] \\ \text{non défini sinon.} \end{cases}$

Soit X et f_X sa densité telle que

$$t \mapsto f_X(t) = \begin{cases} 0 & \text{si } x < -1 \\ \frac{1}{2} & \text{si } -1 \leq x \leq 1 \\ 0 & \text{si } x > 1 \end{cases}$$

* $P\left(\left(Y \in]-1 ; -\frac{4}{5}]\right) \cap \left(X \in]-1;0]\right)\right) = 0$

En effet $-1 < X \leq 0$

$\iff 0 \leq X^4 < 1$

$\iff -1 < -X^4 \leq 0$

$\iff -\frac{4}{5} < Y \leq \frac{1}{5}$

donc $\{Y \in]-1 ; -\frac{4}{5}]\} \cap \{X \in]-1;0]\} = \emptyset$

et donc P (" ") = 0

(car P est une mesure de probabilité)

* Or $P(\{Y \in]-1 ; -\frac{4}{5}]\}) = \int_{-1}^{-4/5} f_Y(t)dt = \left(\frac{1}{5} - \frac{1}{2^4}\right)\left(-\frac{4}{5}+1\right)$

$= \frac{1}{25} - \frac{1}{80} \neq 0$

* et $P(\{X \in]-1;0]\}) = \int_{-1}^{0} f_X(t)dt = \frac{1}{2} \neq 0$

L'égalité "$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$" n'est donc pas vérifiée $\forall (A,B) \in \mathcal{P}(\mathbb{R})^2$

X et Y ne sont donc pas indépendantes

Evaluation page 8/18
(Bourgeois)

Question b

$$\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$$

Or $E(XY) = E\left(-\frac{X}{5} + X^6\right) = -\frac{1}{5}E(X) + E(X^5)$

et $E(Y) = E\left(-\frac{1}{5} + X^4\right) = E(X^4)$

Donc $\mathrm{cov}(X,Y) = -\frac{1}{5}E(X) + E(X^5) - E(X)E(X^4)$

Or $X \leadsto \mathcal{U}([-1;1])$
donc $E(X) = 0$

Donc $\mathrm{cov}(X,Y) = E(X^5) = \int_{\mathbb{R}} x^5 f_X(x)\,dx$

$$\boxed{\mathrm{cov}(X,Y)} = \int_{-1}^{1} x^5 \times \frac{1}{2}\,dx = \frac{1}{2}\times\frac{1}{6}\left[x^6\right]_{-1}^{1} = \frac{1}{6} \neq 0$$

Conclusion : X et Y ne sont pas décorrélées.

Exercice 4 : vecteurs aléatoires

Question a :

Soit $u \in \mathbb{R}^m$, $u^T \Gamma u = \sum_{i,j} u_i \,\mathrm{cov}(X_i, X_j)$

($i \in [\![1,m]\!]$, $j \in [\![1,m]\!]$ et i peut être égal à j)

$$u^T\Gamma u = \sum_{i,j} \mathrm{cov}(u_i X_i, u_j X_j) = \mathrm{var}(u_1 X_1) + \ldots + \mathrm{var}(u_m X_m) + \sum_{i\neq j} \mathrm{cov}(X_i, X_j)$$

$i \in [\![1,m]\!]$
$j \in [\![1,m]\!]$
$i \neq j$

$$u^T\Gamma u = \sum_{i=1}^{m} \mathrm{var}(u_i X_i) + 2\sum_{1\leq i<j\leq m} \mathrm{cov}(X_i, X_j)$$

car $\mathrm{cov}(X_i, X_j) = \mathrm{cov}(X_j, X_i)$ $\forall i,j \in [\![1,m]\!]$

$$u^T\Gamma u = \mathrm{Var}(u_1 X_1 + \ldots + u_m X_m) \geq 0$$

$(= \mathrm{Var}(u^T X))$

Evaluation page 9/18
(Bourgeois)

31 Mars 2022

Question b :

$$\Gamma = var(X) = E\left[(X-E(X))\,{}^t(X-E(X))\right] = (cov(X_i, X_j))_{1\leq i,j\leq m}$$

$$= \begin{pmatrix} Var(X_1) & cov(X_1,X_2) & \cdots & cov(X_1,X_m) \\ cov(X_2,X_1) & Var(X_2) & \cdots & cov(X_2,X_m) \\ \vdots & & \ddots & \vdots \\ Cov(X_m,X_1) & Cov(X_m,X_2) & \cdots & VaR(X_m) \end{pmatrix}$$

est la matrice de covariance de X

Comme $cov(X,Y) = cov(Y,X)$ ($\forall X, Y$ v.a.r de carré intégrable)

(ou comme ${}^t(A^tA) = A^tA$)

Alors, la matrice Γ est symétrique.

Avec Γ diagonale, $cov(X_i, X_j) = 0 \quad \forall i \neq j,\ i,j \in [\![1,m]\!]$

ce qui signifie que les composantes de X sont décorrélées

(les m variables aléatoires composant le vecteur X)

Et si X est gaussien, alors on pourra en déduire l'indépendance (mais rien ne nous l'indique ici).

Question c :

Pour définir une loi gaussienne, il faut définir entièrement ses moments d'ordre 1 et 2.

-> pour définir μ_X, $\boxed{m}$ paramètres sont nécessaires
($les\ (E(X_i))_{i\in[\![1,m]\!]}$)

-> pour définir Γ, sont nécessaires autant de paramètres qu'il y a de $cov(X_i, X_j)$ $i,j \in [\![1,m]\!]$ $i \neq j$ divisé par 2 (en effet $cov(X_i, X_j) = cov(X_j, X_i)$)

soit $1+2+\ldots+(m-1)$ covariances à définir

i.e. $\boxed{\dfrac{m(m-1)}{2}}$

Evaluation page 10/18
(Bourgeois)

→ pour définir Γ, suite, $\underline{\underline{n}}$ variances sont également à définir

__Bilan__ $\boxed{2n + \frac{n(n-1)}{2}}$ paramètres sont nécessaires

__Question d__ : __malédiction de la dimensionnalité.__

Ici le nombre de caractéristiques est supérieur au nombre d'observations. Les caractéristiques agissent comme du bruit dans les données, ce qui diminue la précision du modèle. Dans un espace de grande dimension, les points ont tendance à se distribuer plus faiblement, les observations deviennent plus difficiles à regrouper (difficulté de peuplement dans l'espace). Une dimension trop grande fait que chaque observation apparait équidistantes, aucun cluster significatif ne peut se former (du moins, pas sur une approche de distance euclidienne). Avec une forte dimension, il y a également un plus haut risque de colinéarités (dépendance/redondance), ce qui génère une instabilité du modèle.

(On peut alors envisager une réduction de dimension, ou utiliser les régressions pénalisées lasso pour un apprentissage au plus "juste".

__Question e__

Soit X un vecteur aléatoire de dimension 4 dont les composantes sont les v.a X_1, X_2, X_3, X_4.

La matrice de covariance empirique s'écrit

$$\Gamma_e = \begin{pmatrix} Ve(X_1) & cov(X_1, X_2) & & cov(X_1, X_4) \\ cov(X_2, X_1) & \ddots & & \\ cov(X_3, X_1) & & \ddots & \\ cov(X_4, X_1) & & & Ve(X_4) \end{pmatrix}$$

Evaluation page 11/18
(Bourgeois)

31-03-22

* Calculons la moyenne arithmétique des $(X_i)_{i\in[\![1,4]\!]}$ notés $(\overline{X_i})_{i\in[\![1,4]\!]}$
avec $\overline{X_i} = \frac{1}{N}\sum_{j=1}^{N} X'_{ij}$ où N = nombre de réalisations
X'_{ij} = l'observation de X_i, à la jème observation.

$$\overline{X_1} = \frac{1}{5}(1+1+2+3+4) = \frac{11}{5}$$

$$\overline{X_2} = \frac{1}{5}(1+2+3+4+5) = 3$$

$$\overline{X_3} = \frac{1}{5}(1+3+4+5+6) = \frac{19}{5}$$

$$\overline{X_4} = \frac{1}{5}(1+4+5+6+7) = \frac{23}{5}$$

* Notons $Ve(X)$ la variance empirique de X
$Ve(X_i)$ " de la composante X_i.

$$Ve(X_i) = \frac{1}{N}\sum_{j=1}^{N}(X'_{ij}-\overline{X}_i)^2$$

où X'_{ij} est la valeur observée de X_i à la jème réalisation
N = nb de réalisations.

car variance EMPIRIQUE (cf partie/ex 1)

On a $Ve(X_1) = \frac{1}{5}\left(\left(1-\frac{11}{5}\right)^2+\left(1-\frac{11}{5}\right)^2+\left(2-\frac{11}{5}\right)^2+\left(3-\frac{11}{5}\right)^2+\left(4-\frac{11}{5}\right)^2\right)$

$Ve(X_1) = 1,36$ De même $Ve(X_2) = 2$ $Ve(X_3) = 2,96$
$Ve(X_4) = 4,24$

* Aussi, $\forall (p,q) \in [\![1,4]\!]^2$, $p \neq q$,

$$cov(X_p, X_q) = \frac{1}{5}\sum_{j=1}^{5}(X'_{pj} * X'_{pq} - \overline{X}_p\overline{X}_q)$$

On a alors $cov(X_1,X_2) = \frac{1}{5}\left((1\times1 - \frac{11}{5}\times3) + (1\times2 - \frac{11}{5}\times3) + (2\times3 - \frac{11}{5}\times3) + (3\times4 - \frac{11}{5}\times3) + (4\times5 - \frac{11}{5}\times3)\right) = 1,6$

On calcule de même $cov(X_1,X_3)$, $cov(X_1,X_4)$
$cov(X_2,X_3)$, $cov(X_2,X_4)$
$cov(X_3,X_4)$

Evaluation page 12/18
(Bourgeois)

D'où $\Gamma_e = \begin{pmatrix} 1,36 & 1,6 & 1,84 & 2,08 \\ 1,6 & 2 & 2,4 & 2,8 \\ 1,84 & 2,4 & 2,96 & 3,52 \\ 2,08 & 2,8 & 3,52 & 4,24 \end{pmatrix}$

$\det(\Gamma_e) = 0$ donc $\boxed{\Gamma_e \text{ NON INVERSIBLE.}}$

(détails des calculs disponibles sur demande)

<u>Question f</u>

Soit $B = \begin{pmatrix} 31 & 41 & 51 & 61 \\ 41 & 55 & 69 & 83 \\ 51 & 69 & 87 & 105 \\ 61 & 83 & 105 & 127 \end{pmatrix}$

Une matrice de covariance est semi définie positive (cf question a). <u>Or B n'est pas défini positive</u> (et donc B n'est pas une matrice de covariance empirique).

Pour démontrer ceci, <u>nous allons utiliser la propriété suivante</u> : $\boxed{(\forall A \in S_n(\mathbb{R}),\ A \in S_n^+(\mathbb{R})) \Leftrightarrow (Sp(A) \subset \mathbb{R}^+)}$ (1)

Il ne s'agit donc que de montrer que $\boxed{Sp(B) \not\subset \mathbb{R}^+}$

On a $Sp(B) = \{\lambda \in \mathbb{R} / \det(B - \lambda I_4) = 0\}$

or $\det(B - \lambda I_4) = \begin{vmatrix} 31-\lambda & 41 & 51 & 61 \\ 41 & 55-\lambda & 69 & 83 \\ 51 & 69 & 87-\lambda & 105 \\ 61 & 83 & 105 & 127-\lambda \end{vmatrix}$

(détails de calculs disponibles sur demande)

$\det(B - \lambda I_4) = \lambda^4 - 300\lambda^3 + 480\lambda^2 = \lambda^2(\lambda^2 - 300\lambda + 480)$

$= \lambda^2\left((\lambda + (2\sqrt{5505} - 150))(\lambda - 12\sqrt{5505} + 150)\right) = 0$

d'où $Sp(B) = \{0 ; \underbrace{-2\sqrt{5505} + 150}_{<0} ; 2 + \sqrt{5505} + 150\} \not\subset \mathbb{R}^+$

d'où $B \notin S_n^+(\mathbb{R})$ d'où B n'est pas une matrice de covariance empirique

Evaluation, page 13/18
(Bourgeois)

31-03-22

(1) : démo de $\left((\forall A \in S_n(\mathbb{R}))\left(A \in S_n^+(\mathbb{R})\right) \Leftrightarrow (Sp(A) \subset \mathbb{R}^+)\right)$:

$\Leftarrow$: Soient $n \in \mathbb{N}^*$, $A \in S_n(\mathbb{R})$, $Sp(A) \subset \mathbb{R}$

Supposons A positive, Soit $\lambda \in Sp(A)$, $X \in \mathcal{M}_n(\mathbb{R}) \setminus \{0\}$ un vect propre associé.

Puisque $X \neq 0$, on a ${}^tXX = \|X\|_2^2 > 0$ et $\lambda = \frac{{}^tX(\lambda X)}{{}^tXX} = \frac{{}^tXAX}{{}^tXX} \geq 0$

et donc $Sp(A) \subset \mathbb{R}^+$

$\Rightarrow$: Supposons $Sp(A) \subset \mathbb{R}^+$, d'après le th spectral,

$\exists P \in O_n(\mathbb{R}) / \exists D = \operatorname{diag}(\lambda_i)_{1 \leq i \leq n} \in \mathcal{D}_n(\mathbb{R}) / A = PDP^{-1} = PD{}^tP$

Soit $X \in \mathcal{M}_{n,1}(\mathbb{C})$. En posant $X = (x_i)_{1 \leq i \leq n}$ puis $X' = {}^tPX = (x'_i)_{1 \leq i \leq n}$,

$${}^tXAX = {}^tX(PD{}^tP)X = {}^t({}^tPX)D({}^tPX) = {}^tX'DX$$
$$= \sum_{i=1}^{n} \lambda_i (x'_i)^2 \geq 0 \text{ et } A \text{ positive}$$

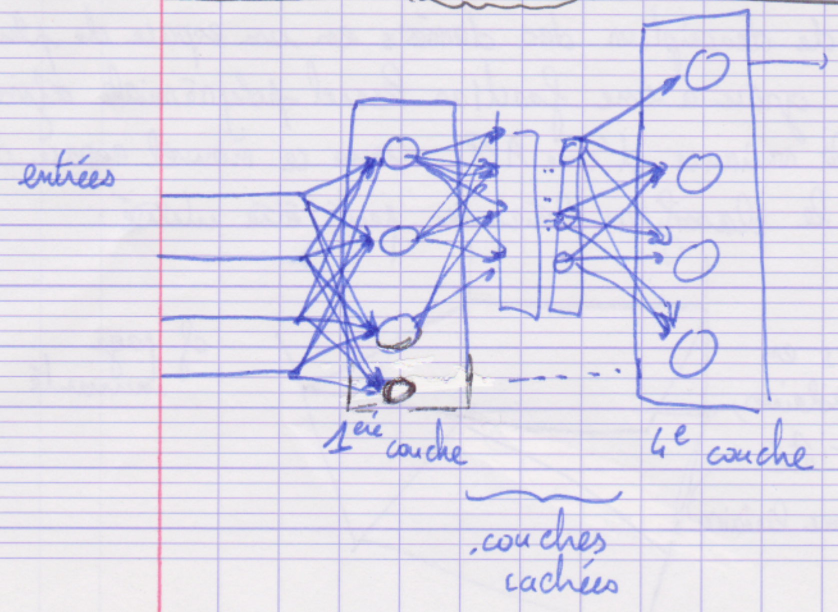
Question g

$$\boxed{\operatorname{Card}(\{\lambda \in Sp(B) / \lambda \neq 0\}) = 2}$$

D'après question f.

Exercice 5 : apprentissage supervisé

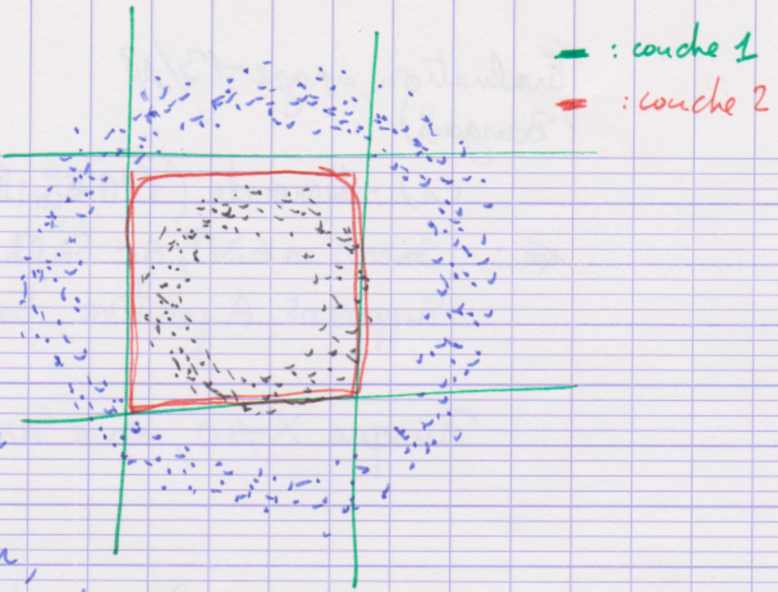
Question a Idée 1 "Réseau de neurones"



"Une combinaison de séparateurs linéaires permet de produire un séparateur global non linéaire."
(Rumelhart, 1986)

Chaque couche est un classifieur linéaire

Evaluation, page 14/18
(Bourgeois)



Dans notre cas, pour obtenir ce qui se rapprocherait du séparateur vert de l'énoncé, 2 couches seraient à envisager, la couche d'entrée serait à 4 entrées

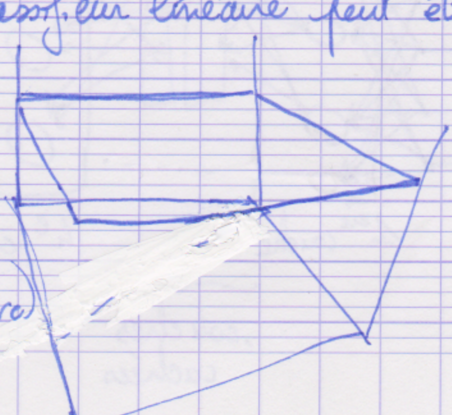
Mais la causalité serait difficile à faire émerger, il y a également un risque de surapprentissage (plus il y a de couches, plus petit est le volume de chaque groupe, plus grand est le risque d'une variance élevée), enfin l'algorithme est coûteux (complexité ++). D'où...

... Idée 2 : "ACP à noyau" (kernel trick)

On change de dimension de l'espace d'entrée par une combinaison linéaire des axes ($\mathbb{R}^2$ à $\mathbb{R}^3$) tel que pour $(x,y) \in \mathbb{R}^2$ on peut exprimer $z \in \mathbb{R}$ / $z = x^2 + y^2$. Les kernels remplacent ce type de combinaison et donnent directement le résultat du produit scalaire $\langle x, y \rangle$.

l'espace de description des données en un espace de plus grande dimension grâce à une fonction kernel polynômiale afin de maximiser la variance d'une ACP. Dans ce nouvel espace de description, le classifieur linéaire peut être utilisé.

(Aussi si on avait exprimé les points en coordonnées polaires, le problème aurait été ramené à un problème linéaire)

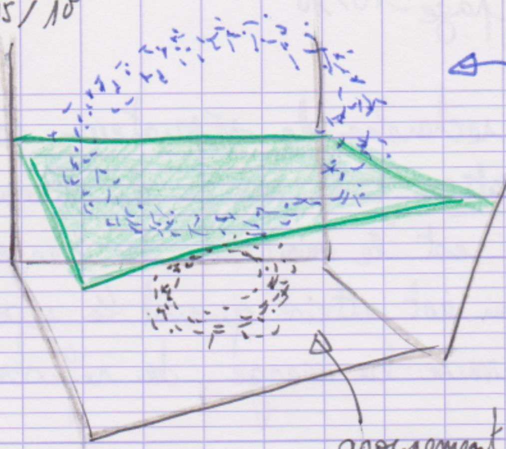


cf page suivante

31 Mars 2022

Evaluation, page 15 / 18

(Bourgeois)



Questions b et c

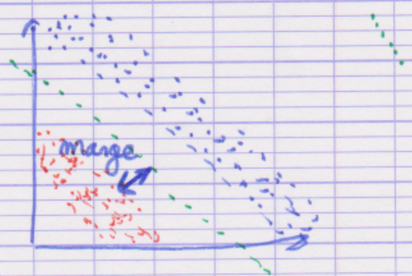
Le problème est un problème de classification.

* On peut envisager l'algorithme des k-means mais les contours des potentiels clusters semblent mal délimités, tant et si bien que le maximum des distance intra-cluster pourrait être supérieur au minimum de la distance inter-cluster.
* L'algorithme des k-médoïdes (PAM) est plus robuste que celui des k-means (moins dépendant des conditions d'initialisation) mais cet algorithme se confronterait tout autant aux écueils du k-means.
* Une régression logistique binaire pourrait être envisagée mais l'algorithme des séparateurs à vastes marges semble plus adapté à l'allure présentée dans l'énoncé.

En effet, l'allure se prête à la recherche d'une marge maximale: la variabilité des marges entre les points (x,y) et les séparateurs candidats est faible, avec une valeur de marge plus grande. Cette recherche de séparateur

(cf page suivante)

marge = distance euclidienne d'1 pt à l'hyperplan séparateur



- Cette recherche de séparateur est un problème d'optimisation (la marge de référence du séparateur candidat est le minimum des marges (x, y) au séparateur, et l'optimum est atteint quand le séparateur choisi est celui qui maximise la marge de référence).

- La convexité de ce problème (d'optimisation) nous permet d'user de la norme $\| \|_0$ (distance euclidienne)

l'hyperplan semble "évident" (peu d'outliers, de bruits qui viennent fausser la recherche de l'optimum
Ici, on obtiendrait une précision dans la prédiction a priori assez satisfaisante.

PARTIE/ EXERCICE 6

cf pages suivantes.
langage choisi ici : Python
Script exécutable en ligne