le cnam

Sujet UE RCP208
Reconnaissance des formes et méthodes neuronales

Année universitaire 2020–2021

Examen 1ère session : 28 janvier 2021

Responsable : Michel CRUCIANU

Durée : 2h00

Aucune communication n'est autorisée durant l'examen à l'exception des échanges avec les serveurs du Cnam et avec les enseignants de cette unité d'enseignement.

Sujet de 5 pages, celle-ci comprise.

Vérifiez que vous disposez bien de la totalité des pages du sujet en début d'épreuve et signalez tout problème de reprographie le cas échéant.

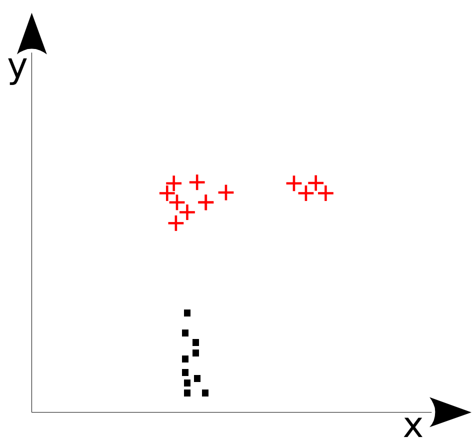
1. Expliquez avec vos propres mots l'utilité des cosinus carrés dans l'interprétation des résultats d'une ACP. **(2 points)**

   Correction : [voir le cours sur les aides à l'interprétation pour l'ACP]

2. L'ACM s'intéresse à l'analyse simultanée des relations entre les modalités de plusieurs variables nominales. Les modalités d'une même variable sont mutuellement exclusives (une observation ne peut posséder qu'une modalité pour chaque variable). Est-ce possible que les projections sur les 2 premiers axes factoriels ACM de deux modalités d'une même variable soient très proches entre elles ? Justifier brièvement. **(2 points)**

   Correction : Bien qu'exclusives l'une de l'autre, deux modalités différentes d'une même variable *peuvent* avoir des projections très proches sur les premiers axes factoriels de l'ACM lorsque les observations qui possèdent l'une ou l'autre de ces modalités pour cette variable sont *très similaires entre elles sur les autres variables* (c'est à dire possèdent en général les mêmes modalités pour les autres variables).

3. Considérons les données bidimensionnelles ($\in \mathbb{R}^2$) représentées dans la figure suivante et appartenant à deux classes différentes (représentées par des '+' et respectivement des '.') :

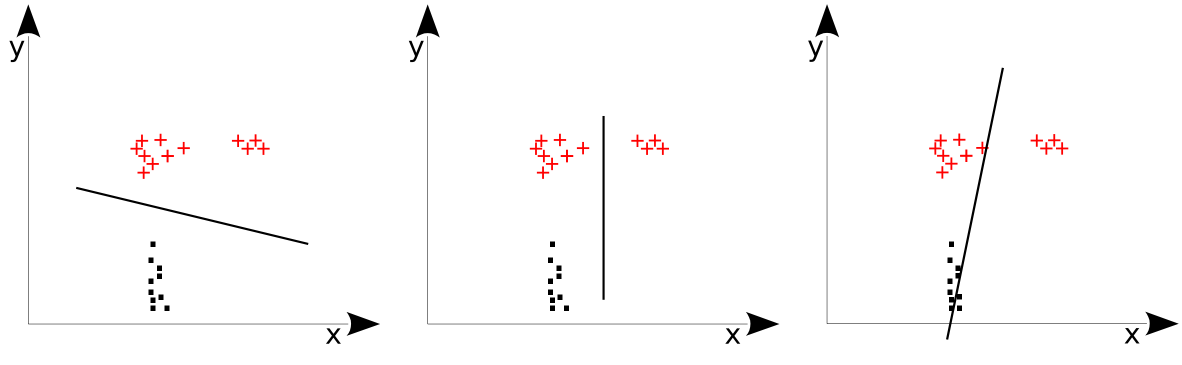


   (a) Parmi les trois directions illustrées dans la figure suivante, laquelle correspond approximativement au premier axe principal d'une ACP centrée ? (**1 point**)

   Correction : Le premier axe principal est la direction de variance maximale des projections par rapport au centre de gravité de l'ensemble du nuage, ce qui dans l'exemple correspond plutôt à la direction 3.

   (b) Parmi ces trois directions, laquelle correspond approximativement au premier axe discriminant ? (**1 point**)

   Correction : L'axe discriminant est celui qui maximise le rapport entre l'inertie interclasses et l'inertie totale des projections des deux classes, ce qui dans l'exemple correspond plutôt à la direction 2.

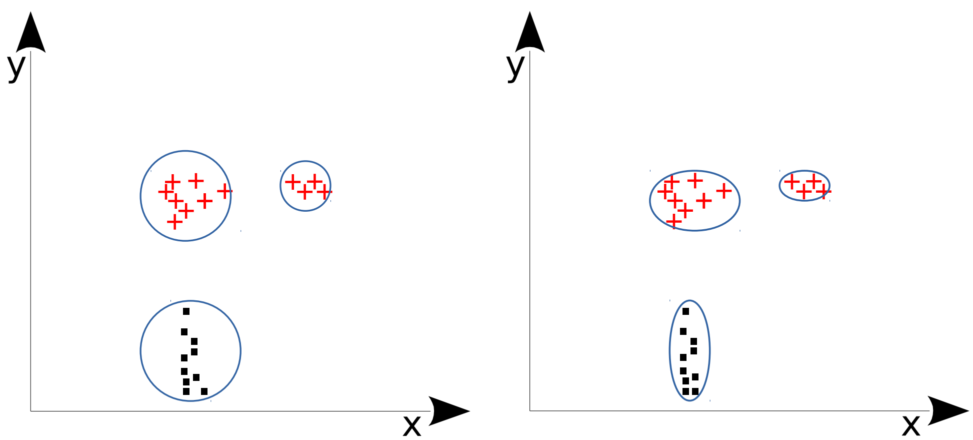


Direction 1 Direction 2 Direction 3

(c) Combien peut-on avoir d'axes discriminants dans cette situation et pourquoi ? (**2 points**)

Correction : Il y a un axe discriminant car il y a seulement deux classes (donc $\text{rang}(\mathbf{E}) \leq 1$) et leurs centres de gravité ne sont pas confondus (donc $\text{rang}(\mathbf{E}) > 0$, l'axe discriminant est bien défini).

(d) Les résultats de deux estimations de mélanges de 3 lois normales bidimensionnelles sont illustrés dans la figure suivante. La forme de chaque enveloppe correspond à la forme de la loi normale bidimensionnelle correspondante. Quel résultat est obtenu en imposant que les matrices de variances-covariances soient « sphériques » (covariances nulles, variances identiques entre les deux variables pour chaque loi normale) ? (**1 point**)



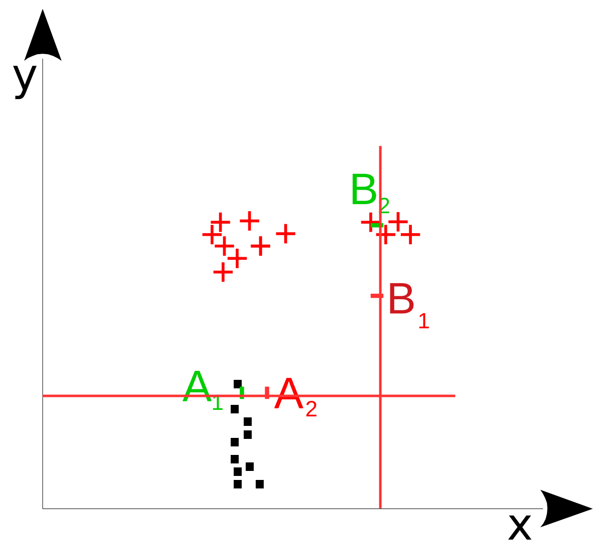
Résultat 1 Résultat 2

Correction : Le résultat 1 correspond à des matrices de variances-covariances « sphériques ».

(e) Quel résultat est obtenu en imposant que les matrices de variances-covariances soient diagonales (covariances nulles) ? (**1 point**)

Correction : Les *deux* résultats correspondent à des matrices de variances-covariances diagonales. Le résultat 2 est plus probable car il correspond à une vraisemblance plus élevée (les lois sont mieux « ajustées » aux observations).

(f) Une observation incomplète A, pour laquelle seule l'ordonnée (valeur de la variable y) est connue, et une observation incomplète B, pour laquelle seule l'abscisse (valeur de la variable x) est connue, sont représentées dans la figure par un trait horizontal et respectivement un trait vertical. Sur la figure suivante nous avons illustré les résultats d'imputations par les moyennes et par les 3 plus proches voisins pour ces deux données incomplètes. Quelle notation sur la figure correspond à quelle imputation ? (**2 points**)



Correction : A_1 imputation par les 3 plus proches voisins (ppv) de A, A_2 imputation par la moyenne (sur x) de A, B_1 imputation par la moyenne (sur y) de B, B_2 imputation par les 3 ppv de B.

4. Dans une carte topologique (que nous notons $\mathcal{C}$), comment peut-on définir le voisinage d'ordre d d'un neurone c ($c \in \mathcal{C}$) ? (**1,5 point**)

Correction : Le voisinage d'ordre d d'un neurone c peut-être défini comme étant le sous-ensemble des neurones r ($r \in \mathcal{C}$) dont la distance avec le neurone c (c'est-à-dire $\delta(c, r)$) est inférieure ou égale à d.
La distance $\delta(c, r)$ est définie comme étant la longueur du plus court chemin entre c et r sur le graphe de $\mathcal{C}$.

5. On s'intéresse au paramètre T (la température) utilisé dans l'apprentissage des cartes topologiques. On décide de faire un apprentissage avec une température constante, c'est-à-dire un apprentissage avec une seule phase et Tmax=Tmin=T (radius_ini=radius_fin dans somtoolbox). Décrivez le comportement de la carte selon la valeur de T. Que se passe-t-il si la température est trop grande ou trop faible ? (**2,5 points**)

   Correction : Quand la valeur de T est fixée alors plus la valeur de T est grande plus les référents se regroupent d'une manière très dense au centre de gravité du nuage de points. Pour une valeur élevée de T, la carte s'organise mais ne parvient pas à se développer sur l'ensemble des données. Pour une petite valeur de T les relations de voisinages interviennent moins. Donc une petite valeur de T permet le déploiement de la carte sur les données, les référents se répartissent plus finement sur les données, mais la carte obtenue n'est pas bien ordonnée.

6. On veut résoudre un problème de prédiction à l'aide d'un perceptron multicouches. Pour ce problème, on essaie de prédire une valeur en utilisant les 7 valeurs précédentes. Le réseau de neurones utilisé possède une couche cachée de 5 neurones. Quel est le nombre de poids (biais ou seuil + poids) de ce réseau ? Justifiez votre réponse (**1,5 point**)

   Correction : - Nombre de poids de la couche d'entrée vers la couche cachée : 7 x 5
   - Nombre de poids de la couche cachée vers la couche de sortie : 5 x 1
   - Nombre de biais de la couche cachée : 5
   - Nombre de biais de la couche de sortie : 1
   Total : (7 x 5) + (5 x 1) + 5 + 1 = 46

7. Expliquez pourquoi le critère d'arrêt prématuré (early stopping : stopper l'apprentissage avant convergence) est important pour favoriser la généralisation dans l'entraînement d'un perceptron multicouche. (**2,5 points**)

   Correction : Le critère d'arrêt prématuré permet d'interrompre l'apprentissage lorsque l'on observe une dégradation des performances sur un ensemble de validation (estimation de la généralisation). Le early stopping est particulièrement utile avec un perceptron multicouche, qui a tendance à surapprendre significativement s'il est entraîné trop longuement, en minimisant jusqu'au bout la fonction erreur définie sur l'ensemble d'apprentissage. Voir la fiche de TP 6.