

Modélisation et Reconnaissance des formes

Epreuve correspondant au cours de Marie-Odile Berger

Une feuille recto-verso format A4 de notes sur le cours est autorisée pendant l'examen. Aucun autre document admis. J'attends de vous des réponses précises et argumentées aux questions posées.

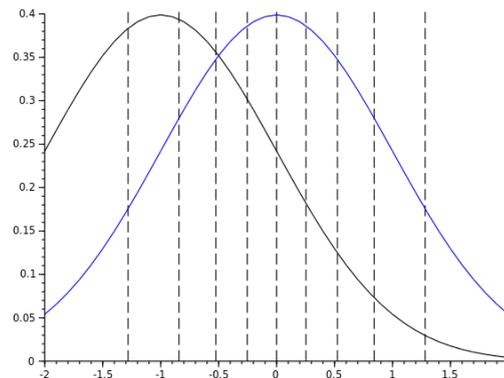
1 Probabilités

soit X une variable aléatoire (v.a.) réelle d'espérance -1 et de variance 2.

1. Calculer l'espérance et la variance de la v.a. $-3X + 2$.
2. En utilisant l'inégalité de Bienaymé Tchebichef, que pouvez vous dire de $p(|X + 1| \leq 2)$?

2 Estimation au Maximum de vraisemblance

1. Soient neuf tirages aléatoires x_1, \dots, x_9 suivant une même loi; les valeurs tirées sont représentées sur le diagrammes ci-dessous par des traits verticaux pointillés. On souhaite estimer au maximum de vraisemblance une loi gaussienne suivant ces échantillons. Dans le graphique ci dessous vous avez deux densités gaussiennes (en noir et en bleu).

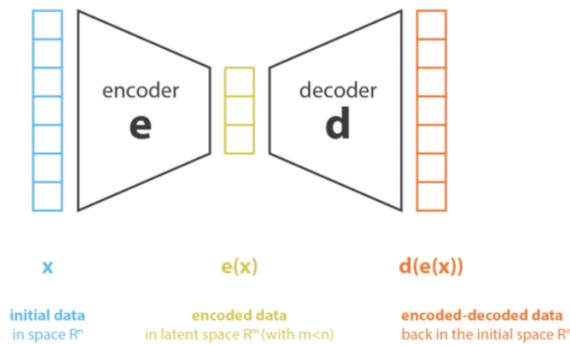


- Comment calcule-t-on la vraisemblance des neuf tirages pour une de ces lois ?
 - Quelle est parmi ces deux densités celle maximisant la vraisemblance des tirages ?
2. La loi exponentielle dépend d'un paramètre α est définie sur R^+ par sa densité $f(x) = \alpha e^{-\alpha x}$. Soit (x_1, \dots, x_n) n tirages de cette loi. Déterminer au maximum de vraisemblance le paramètre α .

3 Réduction de la dimensionalité

On considère un ensemble de formes, chaque forme étant décrite par un vecteur $x \in R^n$. Le processus de réduction de la dimension consiste à représenter une forme avec un vecteur plus compact de R^m avec $m \ll n$ avec une perte minimale d'information. L'idée sous-jacente est d'effectuer les processus de reconnaissance, de classification... avec la représentation réduite pour gagner en efficacité.

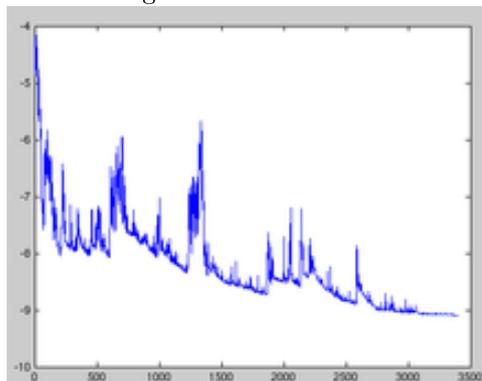
De manière générale (voir figure ci-dessous), cela nécessite de construire un encodeur e (dont le rôle est de construire une représentation $e(x)$ de taille réduite des données x) et d'un décodeur d dont le rôle est de reconstruire la forme à partir de la représentation réduite $e(x)$.



1. Si le système encodeur/décodeur est efficace, comment devraient être x et $d(e(x))$?
2. Nous avons étudié en cours l'analyse en composantes principales, notamment dans le cas de la représentation des visages. Expliquer en quoi l'ACP est un système de type encodeur/décodeur. Préciser quel est la fonction d'encodage et quel est la fonction de décodage.
3. On regarde dans la suite comment implanter l'encodeur et le décodeur par des réseaux de neurones.
 - Quelle serait la fonction de coût minimisée? L'apprentissage est-il supervisé (y-a-t'il besoin de données étiquetées?)
 - Parmi les couches des réseaux convolutionnels que vous connaissez (par exemple couche convolutionnelle, de pooling,...), pouvez vous citer des couches linéaires et des couches non linéaires?
 - Si on ne met que des couches de convolution, aura t'on un résultat fondamentalement différent de celui de l'ACP? Pourquoi? Expliquer quelles couches introduire dans le CNN pour amener un résultat différent?

4 CNN et fonctions de coût

1. Quel est le lien entre mise en oeuvre des réseaux convolutionnels (CNN) et optimisation de fonction de coût?
2. Quel est la différence entre optimisation par méthode de gradient et optimisation par gradient stochastique (SGD)? A quoi sert la méthode SGD par rapport à une méthode de descente de gradient classique?
3. La figure ci-dessous vous montre la minimisation d'une fonction au cours des itérations en utilisant une méthode SGD. Pourquoi observe-t-on des irrégularités fortes?



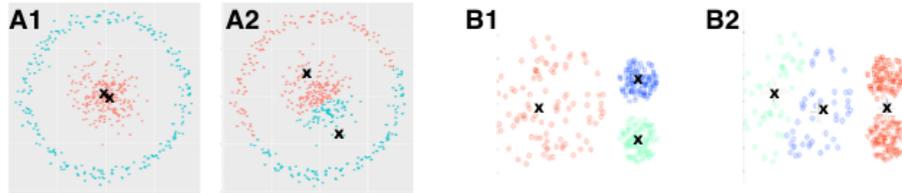
5 K means

On dispose de n observations x_1, \dots, x_n appartenant à R^N . On souhaite former k classes S_1, \dots, S_k à partir de ces données. Dans la méthode des k -means, on cherche un tel partitionnement minimisant :

$$\sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

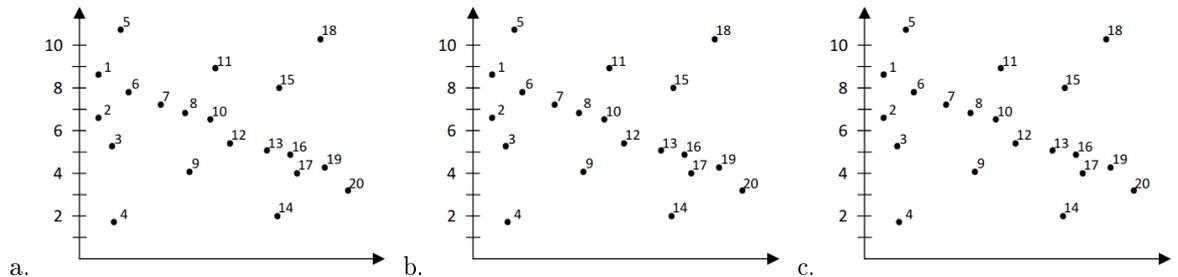
où μ_i est la moyenne des points dans S_i .

1. Décrire un algorithme permettant le calcul des k-means. Cet algorithme est-il optimal? Pourquoi?
2. Etant donnée une nouvelle mesure x dont on veut déterminer la classe d'appartenance, sur quel principe décide-t-on d'affecter x à l'une des k classes?
3. Dans les deux cas A et B de la figure suivante, chaque ensemble de données est classifié avec deux méthodes différentes notées A_1 et A_2 (resp. B_1 et B_2), une des deux méthodes étant un K-means. Une croix indique le centre de chaque classe. Indiquez pour chacun des deux cas A et B le résultat qui est issu des K-means et justifiez votre réponse.



6 Estimation robuste

1. Application de RANSAC : On considère le problème de détecter des droites approchant au mieux les points dessinés dans la figure ci dessous. Je vous demande d'appliquer trois fois l'algorithme RANSAC pour détecter des droites et d'indiquer les inliers (en les entourant d'un cercle) respectivement sur les figures a, b et c. Les deux points sélectionnés aléatoirement pour mettre en oeuvre l'algorithme sont respectivement pour les trois exemples a, b, c les points d'indice : [13,16], [3, 10] et [1,10]. Le paramètre de tolérance d'erreur pour évaluer l'adéquation d'un point au modèle sera fixé à 1.0. Dans chacun des cas, entourer les points considérés comme inlier¹. Quelle sera parmi ces 3 essais celui qui sera retenu par l'algorithme?



2. Nous avons vu en cours deux méthodes robustes : les M-estimateurs et l'algorithme RANSAC. Quelle est selon vous la méthode la plus efficace en temps de calcul? (expliquer pourquoi)
3. Une méthode de type RANSAC peut-elle prendre en comptes des problèmes où le taux de données erronées est de 80%?
4. Le point de rupture d'un estimateur est la proportion d'observations incorrectes qu'un estimateur peut supporter sans donner de valeurs très erronées
 - Expliquer pourquoi le point de rupture de l'estimateur moyenne est 0.
 - Quel est le point de rupture de l'estimateur médiane?
 - On considère l'estimateur suivant d'une grandeur p : $\hat{p}_h = \sum_{i=1}^h r_i(p)$ où les résidus $r_i(p)$ sont ordonnés par valeurs croissantes et h est un entier fixé (pour fixer les idées, $p = (a, b)$ est par exemple une droite à estimer et $r_i(p) = |y_i - ax_i + b|$ est le résidu associé à chaque mesure (x_i, y_i)).
Quel est le point de rupture de cet estimateur?

1. Je ne vous demande pas un dessin très précis...