

Quantification et Clustering avec des Divergences de Bregman

Aurélie FISCHER

LSTA, Université Pierre et Marie Curie - Paris VI

6 mai 2009

Plan de l'exposé

Plan de l'exposé

- Introduction : quantification et clustering

Plan de l'exposé

- Introduction : quantification et clustering
- Divergences de Bregman

Plan de l'exposé

- Introduction : quantification et clustering
- Divergences de Bregman
- Résultats

Plan de l'exposé

- Introduction : quantification et clustering
- Divergences de Bregman
- Résultats
- Simulations, “comparaisons” entre divergences

Introduction

Quantification et Clustering

Quantification

Introduction

Quantification et Clustering

Quantification

- Compression avec perte

Introduction

Quantification et Clustering

Quantification

- Compression avec perte
- Représenter une variable aléatoire X de loi μ à valeurs dans E par $q(X)$ avec $q : E \rightarrow \mathbf{c} = \{c_1, \dots, c_k\}$ un **quantificateur** (à k points).

Introduction

Quantification et Clustering

Quantification

- Compression avec perte
- Représenter une variable aléatoire X de loi μ à valeurs dans E par $q(X)$ avec $q : E \rightarrow \mathbf{c} = \{c_1, \dots, c_k\}$ un **quantificateur** (à k points).
- Un quantificateur est déterminé par sa **table de codage \mathbf{c}** et la **partition de E** qu'il induit.

Introduction

Quantification et Clustering

Clustering

Introduction

Quantification et Clustering

Clustering

- Données X_1, \dots, X_n supposées indépendantes et de loi μ .

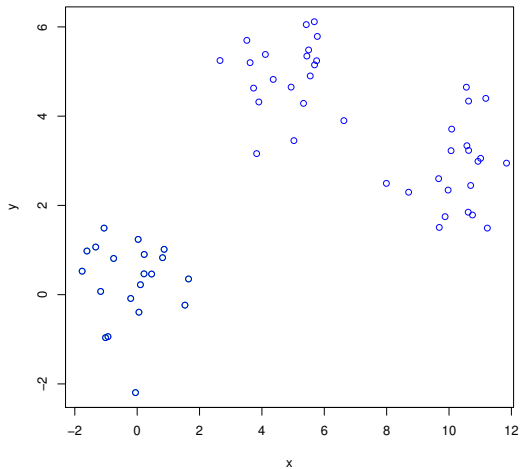
Introduction

Quantification et Clustering

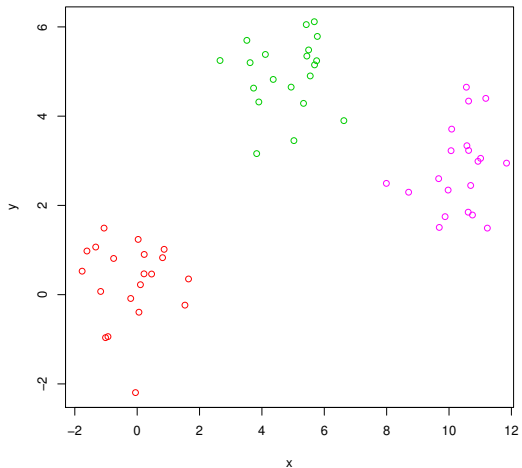
Clustering

- Données X_1, \dots, X_n supposées indépendantes et de loi μ .
- Etant donné un nombre k , former k groupes de données aussi **homogènes** et **séparés** que possible.

Introduction



Introduction



Introduction

La distorsion

L'erreur commise en remplaçant X par $q(X)$ est mesurée par la **distorsion**

Introduction

La distorsion

L'erreur commise en remplaçant X par $q(X)$ est mesurée par la **distorsion**

$$W(\mu, q) = \mathbb{E}d(X, q(X)) = \int_E d(x, q(x))d\mu(x)$$

Introduction

La distorsion

L'erreur commise en remplaçant X par $q(X)$ est mesurée par la **distorsion**

$$W(\mu, q) = \mathbb{E}d(X, q(X)) = \int_E d(x, q(x))d\mu(x)$$

où la **mesure de distorsion** d est une fonction telle que $d(x, y) \geq 0$ pour tous $x, y \in E$ et $d(x, y) = 0$ si et seulement si $x = y$.

Introduction

La distorsion empirique

Soit μ_n la **mesure empirique** correspondant à μ : pour tout A borélien de E ,

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \in A\}}.$$

Introduction

La distorsion empirique

Soit μ_n la **mesure empirique** correspondant à μ : pour tout A borélien de E ,

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \in A\}}.$$

Introduisons la **distorsion empirique** :

$$W(\mu_n, q) = \frac{1}{n} \sum_{i=1}^n d(X_i, q(X_i)).$$

Divergences de Bregman

Une mesure de distorsion

Souvent, on utilise la **distance euclidienne** au carré.

Divergences de Bregman

Une mesure de distorsion

Souvent, on utilise la **distance euclidienne** au carré.

Mais :

- Pas adaptée à tous les types de données

Divergences de Bregman

Une mesure de distorsion

Souvent, on utilise la **distance euclidienne** au carré.

Mais :

- Pas adaptée à tous les types de données
- Données de **dimension infinie** ?
→ dans de nombreux domaines, les données à analyser sont des courbes.

Divergences de Bregman

Une classe de mesures de distorsion

Nous nous intéressons ici aux **divergences de Bregman**, une classe de mesures de distorsion introduite par Bregman en 1967.

Divergences de Bregman

Une classe de mesures de distorsion

Nous nous intéressons ici aux **divergences de Bregman**, une classe de mesures de distorsion introduite par Bregman en 1967.

Pas nécessairement des distances :

Divergences de Bregman

Une classe de mesures de distorsion

Nous nous intéressons ici aux **divergences de Bregman**, une classe de mesures de distorsion introduite par Bregman en 1967.

Pas nécessairement des distances :

- non symétriques

Divergences de Bregman

Une classe de mesures de distorsion

Nous nous intéressons ici aux **divergences de Bregman**, une classe de mesures de distorsion introduite par Bregman en 1967.

Pas nécessairement des distances :

- non symétriques
- ne vérifiant pas l'inégalité triangulaire

Divergences de Bregman

Définition en dimension finie

Définition (Divergence de Bregman fini-dimensionnelle)

Soient $\mathcal{C} \subset \mathbb{R}^d$ un convexe et $\phi : \mathcal{C} \rightarrow \mathbb{R}$ une fonction strictement convexe, différentiable sur $ir(\mathcal{C})$, l'intérieur relatif de \mathcal{C} . La **divergence de Bregman** associée

$$d_\phi : \mathcal{C} \times ir(\mathcal{C}) \rightarrow [0, \infty[$$

est définie par

Divergences de Bregman

Définition en dimension finie

Définition (Divergence de Bregman fini-dimensionnelle)

Soient $\mathcal{C} \subset \mathbb{R}^d$ un convexe et $\phi : \mathcal{C} \rightarrow \mathbb{R}$ une fonction strictement convexe, différentiable sur $ir(\mathcal{C})$, l'intérieur relatif de \mathcal{C} . La **divergence de Bregman** associée

$$d_\phi : \mathcal{C} \times ir(\mathcal{C}) \rightarrow [0, \infty[$$

est définie par

$$d_\phi(x, y) = \phi(x) - \phi(y) - \langle x - y, \nabla\phi(y) \rangle.$$

où $\nabla\phi(y)$ est le gradient de ϕ en y .

Divergences de Bregman

Exemples en dimension finie (1)

Exemples

Divergences de Bregman

Exemples en dimension finie (1)

Exemples

- Distance euclidienne au carré (dim. 1) :

$$\phi(x) = x^2 \rightarrow d_\phi(x, y) = (x - y)^2.$$

Divergences de Bregman

Exemples en dimension finie (1)

Exemples

- **Distance euclidienne au carré (dim. 1) :**

$$\phi(x) = x^2 \rightarrow d_\phi(x, y) = (x - y)^2.$$

- **Perte exponentielle :**

$$\phi(x) = e^x \rightarrow d_\phi(x, y) = e^x - e^y - (x - y)e^y.$$

Divergences de Bregman

Exemples en dimension finie (1)

Exemples

- **Distance euclidienne au carré (dim. 1) :**

$$\phi(x) = x^2 \rightarrow d_\phi(x, y) = (x - y)^2.$$

- **Perte exponentielle :**

$$\phi(x) = e^x \rightarrow d_\phi(x, y) = e^x - e^y - (x - y)e^y.$$

- **Perte logistique :** $\phi(x) = x \ln x + (1 - x) \ln(1 - x)$

$$\rightarrow d_\phi(x, y) = x \ln \frac{x}{y} + (1 - x) \ln \left(\frac{1 - x}{1 - y} \right).$$

Divergences de Bregman

Exemples en dimension finie (1)

Exemples

- **Distance euclidienne au carré (dim. 1) :**

$$\phi(x) = x^2 \rightarrow d_\phi(x, y) = (x - y)^2.$$

- **Perte exponentielle :**

$$\phi(x) = e^x \rightarrow d_\phi(x, y) = e^x - e^y - (x - y)e^y.$$

- **Perte logistique :** $\phi(x) = x \ln x + (1 - x) \ln(1 - x)$

$$\rightarrow d_\phi(x, y) = x \ln \frac{x}{y} + (1 - x) \ln \left(\frac{1 - x}{1 - y} \right).$$

- **Distance euclidienne au carré :**

$$\phi(x) = \|x\|^2 \rightarrow d_\phi(x, y) = \|x - y\|^2.$$

Divergences de Bregman

Exemples en dimension finie (2)

Exemples

- **Distance de Mahalanobis** : pour A définie positive,
 $\phi(x) = {}^t x A x \rightarrow d_\phi(x, y) = {}^t (x - y) A (x - y)$.

Divergences de Bregman

Exemples en dimension finie (2)

Exemples

- **Distance de Mahalanobis** : pour A définie positive,
 $\phi(x) = {}^t x A x \rightarrow d_\phi(x, y) = {}^t (x - y) A (x - y)$.
- **Distance de Kullback-Leibler** : pour $x = (x_1, \dots, x_d)$ avec

$$\sum_{j=1}^d x_j = 1, \phi(x) = \sum_{j=1}^d x_j \ln x_j \rightarrow d_\phi(x, y) = \sum_{i=1}^n x_i \ln \frac{x_i}{y_i}.$$

Divergences de Bregman

Exemples en dimension finie (2)

Exemples

- **Distance de Mahalanobis** : pour A définie positive,
 $\phi(x) = {}^t x A x \rightarrow d_\phi(x, y) = {}^t (x - y) A (x - y)$.
- **Distance de Kullback-Leibler** : pour $x = (x_1, \dots, x_d)$ avec
 $\sum_{j=1}^d x_j = 1$, $\phi(x) = \sum_{j=1}^d x_j \ln x_j \rightarrow d_\phi(x, y) = \sum_{i=1}^n x_i \ln \frac{x_i}{y_i}$.
- **Distance de Kullback-Leibler généralisée** :
 $\phi(x) = \sum_{j=1}^d x_j \ln x_j \rightarrow d_\phi(x, y) = \sum_{i=1}^n x_i \ln \frac{x_i}{y_i} - (x_i - y_i)$.

Divergences de Bregman

Définition en dimension infinie

Définition (Divergence de Bregman fonctionnelle)

Soit E un **espace de Banach séparable et réflexif**. Soient $\mathcal{C} \subset E$ un convexe et $\phi : \mathcal{C} \rightarrow \mathbb{R}$ strictement convexe, de classe C^2 sur $ir(\mathcal{C})$. La **divergence de Bregman** associée est définie par

Divergences de Bregman

Définition en dimension infinie

Définition (Divergence de Bregman fonctionnelle)

Soit E un **espace de Banach séparable et réflexif**. Soient $\mathcal{C} \subset E$ un convexe et $\phi : \mathcal{C} \rightarrow \mathbb{R}$ strictement convexe, de classe C^2 sur $ir(\mathcal{C})$. La **divergence de Bregman** associée est définie par

$$d_\phi(x, y) = \phi(x) - \phi(y) - D_y\phi(x - y)$$

avec $D_y\phi$ la différentielle de ϕ en y .

Divergences de Bregman

Exemples en dimension infinie (1)

Exemples

Divergences de Bregman

Exemples en dimension infinie (1)

Exemples

- Norme L^2 au carré :

$$\phi(x) = \int x^2(t) dt \rightarrow d_\phi(x, y) = \int (x(t) - y(t))^2 dt.$$

Divergences de Bregman

Exemples en dimension infinie (1)

Exemples

- Norme L^2 au carré :

$$\phi(x) = \int x^2(t) dt \rightarrow d_\phi(x, y) = \int (x(t) - y(t))^2 dt.$$

- Biais quadratique :

$$\phi(x) = \left(\int x(t) dt \right)^2 \rightarrow d_\phi(x, y) = \left(\int (x(t) - y(t)) dt \right)^2.$$

Divergences de Bregman

Exemples en dimension infinie (1)

Exemples

- Norme L^2 au carré :

$$\phi(x) = \int x^2(t) dt \rightarrow d_\phi(x, y) = \int (x(t) - y(t))^2 dt.$$

- Biais quadratique :

$$\phi(x) = \left(\int x(t) dt \right)^2 \rightarrow d_\phi(x, y) = \left(\int (x(t) - y(t)) dt \right)^2.$$

- Distance de Kullback-Leibler :

$$\phi(x) = \int x(t) \ln x(t) dt \rightarrow d_\phi(x, y) = \int x(t) \ln \frac{x(t)}{y(t)} dt.$$

Divergences de Bregman

Exemples en dimension infinie (2)

Exemples

- Distance de Kullback-Leibler généralisée :

$$\phi(x) = \int x(t) \ln x(t) dt$$

$$\rightarrow d_{\phi}(x, y) = \int x(t) \ln \frac{x(t)}{y(t)} - (x(t) - y(t)) dt.$$

Divergences de Bregman

Exemples en dimension infinie (2)

Exemples

- **Distance de Kullback-Leibler généralisée** :

$$\phi(x) = \int x(t) \ln x(t) dt$$

$$\rightarrow d_\phi(x, y) = \int x(t) \ln \frac{x(t)}{y(t)} - (x(t) - y(t)) dt.$$

- **Distance de Itakura-Saito** : $\phi(x) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(x(\theta)) d\theta$

$$\rightarrow d_\phi(x, y) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\ln \frac{x(\theta)}{y(\theta)} - \frac{x(\theta)}{y(\theta)} + 1 \right) d\theta.$$

Minimum de la distorsion

Meilleure partition

- A même table de codage, le meilleur quantificateur est le

Minimum de la distorsion

Meilleure partition

- A même table de codage, le meilleur quantificateur est le **quantificateur des plus proches voisins**

Minimum de la distorsion

Meilleure partition

- A même table de codage, le meilleur quantificateur est le **quantificateur des plus proches voisins** associé à la **partition de Voronoi**.

Minimum de la distorsion

Meilleure partition

- A même table de codage, le meilleur quantificateur est le quantificateur des plus proches voisins associé à la partition de Voronoi.
- Donc, trouver la meilleure table de codage $\mathbf{c}^* = \{c_1, \dots, c_k\}$,

Minimum de la distorsion

Meilleure partition

- A même table de codage, le meilleur quantificateur est le **quantificateur des plus proches voisins** associé à la **partition de Voronoi**.
- Donc, trouver la **meilleure table de codage** $\mathbf{c}^* = \{c_1, \dots, c_k\}$, en minimisant

$$W(\mu, \mathbf{c}) = \mathbb{E} \min_{j=1, \dots, k} d_\phi(X, c_j),$$

Minimum de la distorsion

Meilleure partition

- A même table de codage, le meilleur quantificateur est le **quantificateur des plus proches voisins** associé à la **partition de Voronoi**.
- Donc, trouver la **meilleure table de codage** $\mathbf{c}^* = \{c_1, \dots, c_k\}$, en minimisant

$$W(\mu, \mathbf{c}) = \mathbb{E} \min_{j=1, \dots, k} d_\phi(X, c_j),$$

$$W(\mu_n, \mathbf{c}) = \frac{1}{n} \sum_{i=1}^n \min_{j=1, \dots, k} d_\phi(X_i, c_j).$$

Minimum de la distorsion

Conditions assurant l'existence d'un minimum

→ Argument de **compacité**

Minimum de la distorsion

Conditions assurant l'existence d'un minimum

→ Argument de **compacité**

- \mathcal{C} est de **dimension finie**, $y \mapsto d_\phi(x, y)$ est semi-continue inférieurement et $d_\phi(x, y) \leq \lim_{\|z\| \rightarrow +\infty} d_\phi(x, z)$.

Minimum de la distorsion

Conditions assurant l'existence d'un minimum

→ Argument de **compacité**

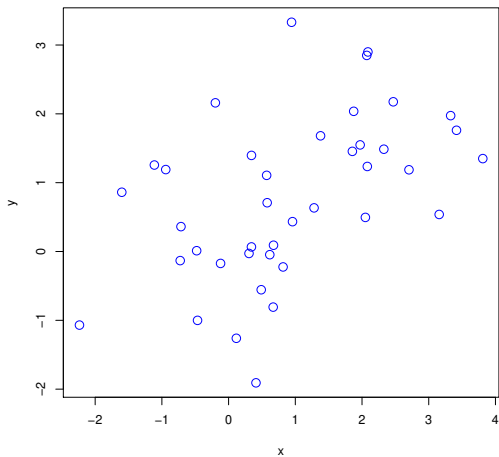
- \mathcal{C} est de **dimension finie**, $y \mapsto d_\phi(x, y)$ est semi-continue inférieurement et $d_\phi(x, y) \leq \lim_{\|z\| \rightarrow +\infty} d_\phi(x, z)$.
- Il existe $R > 0$ tel que $\mathbb{P}\{\|X\| \leq R\} = 1$, et $y \mapsto d_\phi(x, y)$ est semi-continue inférieurement pour la topologie faible $\sigma(E, E')$.

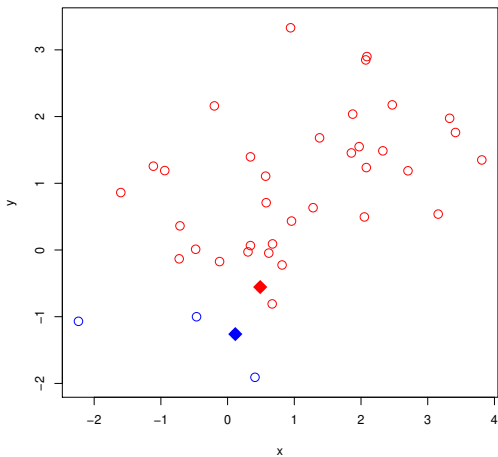
Minimum de la distorsion

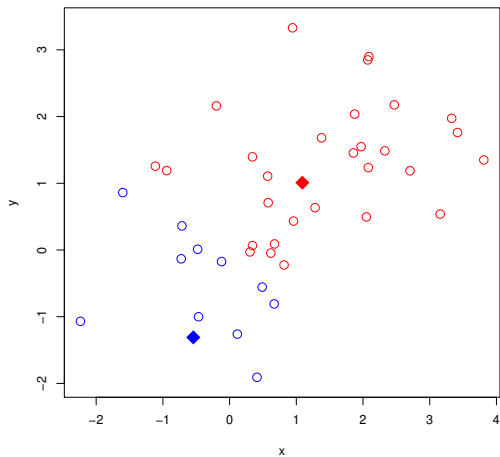
Conditions assurant l'existence d'un minimum

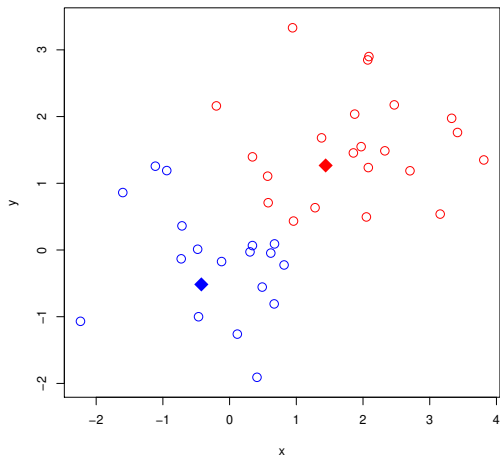
→ Argument de **compacité**

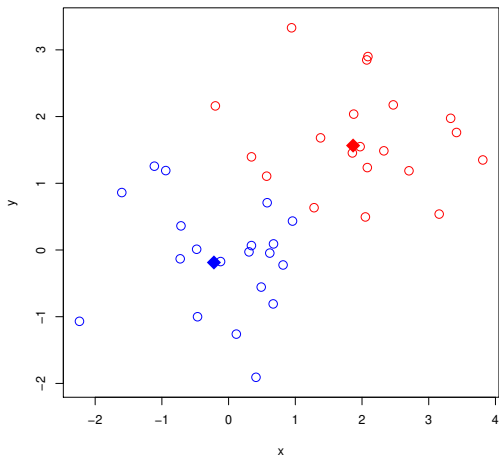
- \mathcal{C} est de **dimension finie**, $y \mapsto d_\phi(x, y)$ est semi-continue inférieurement et $d_\phi(x, y) \leq \lim_{\|z\| \rightarrow +\infty} d_\phi(x, z)$.
- Il existe $R > 0$ tel que $\mathbb{P}\{\|X\| \leq R\} = 1$, et $y \mapsto d_\phi(x, y)$ est semi-continue inférieurement pour la topologie faible $\sigma(E, E')$.
- Carré de la **norme** d'un espace réflexif.

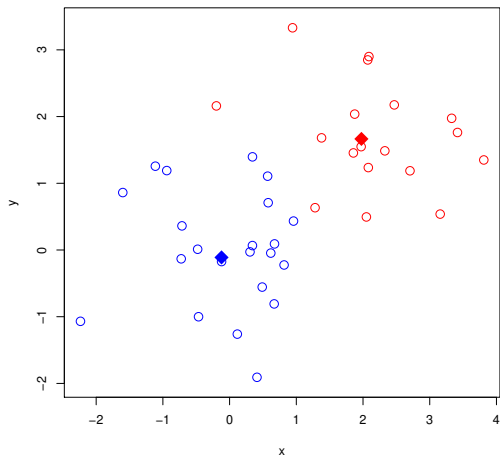


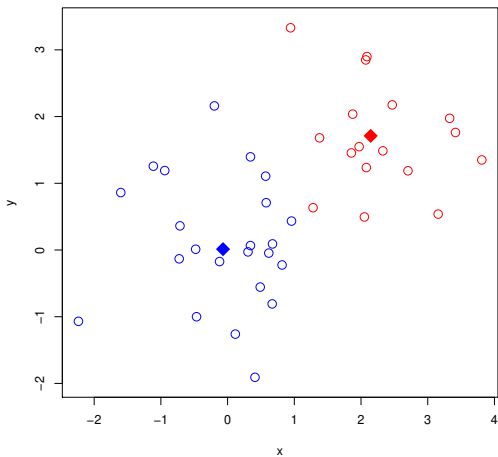


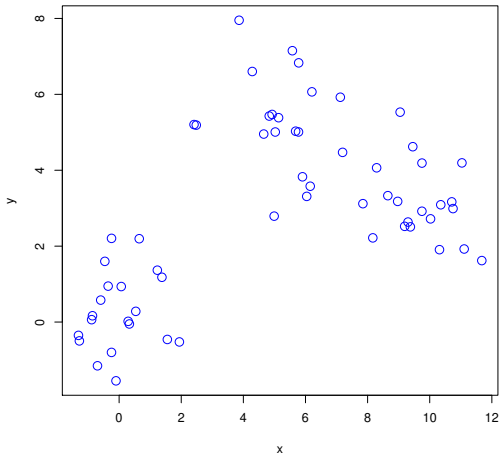


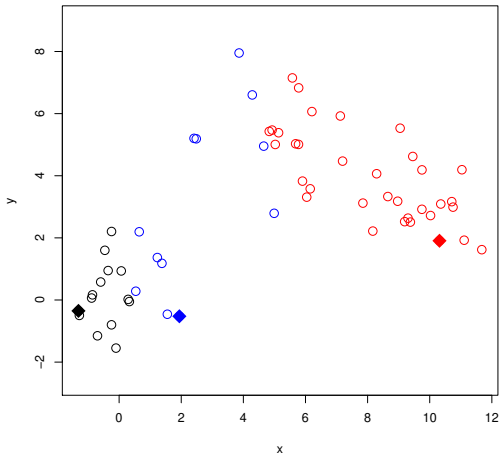


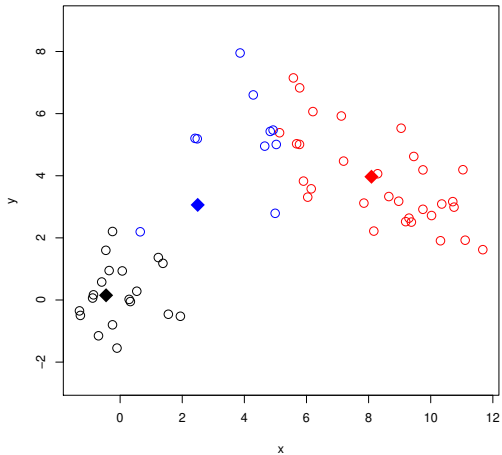


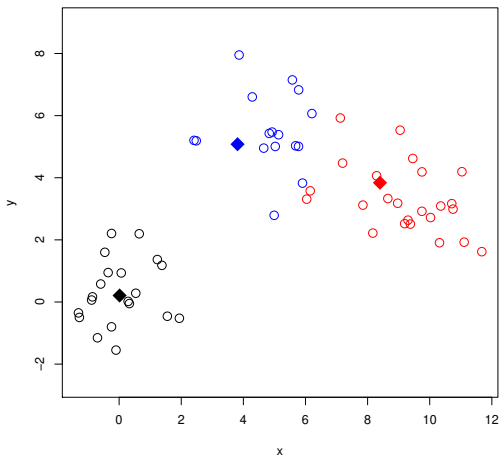


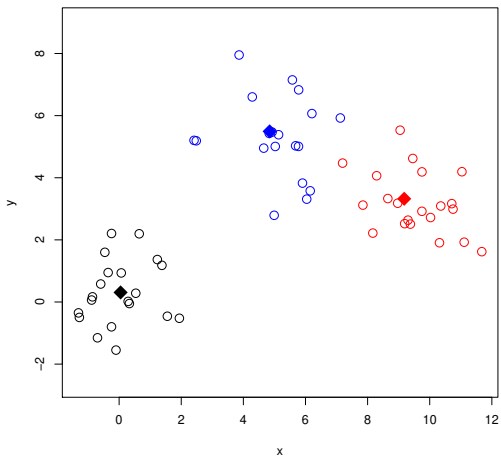


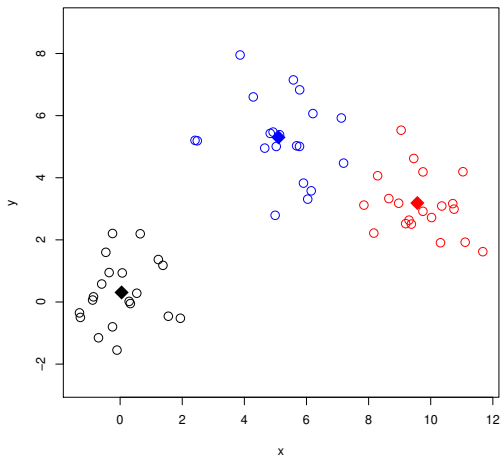


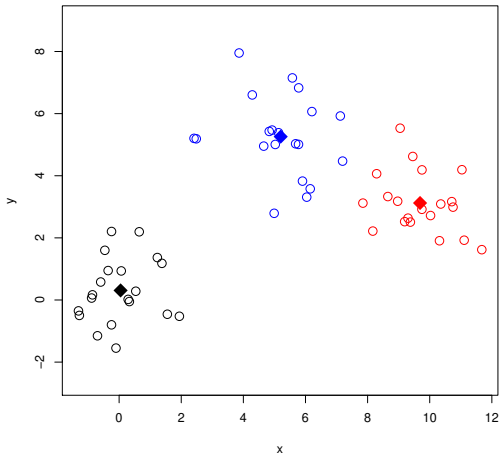












Convergence de la distorsion

Sous de bonnes conditions, on a

$$\lim_{n \rightarrow \infty} W(\mu, \mathbf{c}_n^*) = \inf_{\mathbf{c} \in \mathcal{C}^k} W(\mu, \mathbf{c}) \quad p.s.$$

où \mathbf{c}_n^* est un minimiseur de la distorsion empirique.

Conditions assurant la convergence

- \mathcal{C} est de **dimension finie**, $d_\phi(\cdot, \cdot)$ est continue, $d_\phi(x, y) \leq \lim_{\|z\| \rightarrow +\infty} d_\phi(x, z)$, et $y \mapsto d_\phi(x, y)$ est convexe.
- Il existe $R > 0$ tel que $\mathbb{P}\{\|X\| \leq R\} = 1$ et $M = M(\phi, R) \geq 0$ tel que $\|D_c \phi\| \leq M$ pour tout $c \in B_R$, et $y \mapsto d_\phi(x, y)$ est semi-continue inférieurement pour la topologie faible $\sigma(E, E')$.
- Carré d'une **norme** hilbertienne.

Convergence de la distorsion

Vitesse de convergence

S'il existe $R > 0$ tel que $\mathbb{P}\{\|X\| \leq R\} = 1$, et si $\|D_c\phi\|$ et $|\phi(c) - D_c\phi(c)|$ sont uniformément bornés sur B_R par $M_1 = M_1(\phi, R) \geq 0$ et $M_2 = M_2(\phi, R) \geq 0$ respectivement, alors, on a

$$\mathbb{E}W(\mathbf{c}_n^*, \mu) - \inf_{\mathbf{c} \in \mathcal{C}^k} W(\mu, \mathbf{c}) \leq \frac{4k}{\sqrt{n}} (M_1 R + M_2).$$

Applications

En pratique, les différentes divergences de Bregman sont **adaptées au clustering** de différents types de données en dimension **finie** ou **infinie**. Donnons quelques illustrations.

Applications

En pratique, les différentes divergences de Bregman sont **adaptées au clustering** de différents types de données en dimension **finie** ou **infinie**. Donnons quelques illustrations.

Remarque (“Comparer” les divergences de Bregman)

Applications

En pratique, les différentes divergences de Bregman sont **adaptées au clustering** de différents types de données en dimension **finie** ou **infinie**. Donnons quelques illustrations.

Remarque (“Comparer” les divergences de Bregman)

- **Comparer les distorsions** obtenues avec différentes divergences de Bregman **n'a pas de sens**.

Applications

En pratique, les différentes divergences de Bregman sont **adaptées au clustering** de différents types de données en dimension **finie** ou **infinie**. Donnons quelques illustrations.

Remarque (“Comparer” les divergences de Bregman)

- **Comparer les distorsions** obtenues avec différentes divergences de Bregman **n'a pas de sens**.
- Le critère permettant d'évaluer la performance d'une divergence de Bregman dans les simulations suivantes repose sur la **comparaison de la partition obtenue avec une partition “attendue”**.

Applications

En pratique, les différentes divergences de Bregman sont **adaptées au clustering** de différents types de données en dimension **finie** ou **infinie**. Donnons quelques illustrations.

Remarque (“Comparer” les divergences de Bregman)

- **Comparer les distorsions** obtenues avec différentes divergences de Bregman **n'a pas de sens**.
- Le critère permettant d'évaluer la performance d'une divergence de Bregman dans les simulations suivantes repose sur la **comparaison de la partition obtenue avec une partition “attendue”**.
Pas très satisfaisant en apprentissage non supervisé!

Applications

Bijection Famille exponentielle \leftrightarrow Divergence de Bregman

(Banerjee, Merugu, Dhillon et Gosh)

→ Telle divergence de Bregman convient mieux à des données suivant telle loi.

Applications

Bijection Famille exponentielle \leftrightarrow Divergence de Bregman

(Banerjee, Merugu, Dhillon et Gosh)

→ Telle divergence de Bregman convient mieux à des données suivant telle loi.

Exemples

- Loi gaussienne \leftrightarrow Distance euclidienne au carré

Applications

Bijection Famille exponentielle \leftrightarrow Divergence de Bregman

(Banerjee, Merugu, Dhillon et Gosh)

→ Telle divergence de Bregman convient mieux à des données suivant telle loi.

Exemples

- Loi gaussienne \leftrightarrow Distance euclidienne au carré
- Loi de Poisson \leftrightarrow Distance de Kullback-Leibler généralisée

Applications

Bijection Famille exponentielle \leftrightarrow Divergence de Bregman

(Banerjee, Merugu, Dhillon et Gosh)

→ Telle divergence de Bregman convient mieux à des données suivant telle loi.

Exemples

- Loi gaussienne \leftrightarrow Distance euclidienne au carré
- Loi de Poisson \leftrightarrow Distance de Kullback-Leibler généralisée
- Loi binomiale \leftrightarrow Perte logistique

Simulations gaussienne/Poisson/binomiale

Pour chacune des lois gaussienne, Poisson et binomiale, on simule 90 données réparties en 3 ensembles de 30 données centrées en 10, 20 et 40 respectivement.

Simulations gaussienne/Poisson/binomiale

Pour chacune des lois gaussienne, Poisson et binomiale, on simule 90 données réparties en 3 ensembles de 30 données centrées en 10, 20 et 40 respectivement.

	Euclidien	Kullback-Leibler	Logistique
Gaussienne	0,689	0,672	0,685
Poisson	0,702	0,732	0,728
Binomiale	0,791	0,806	0,813

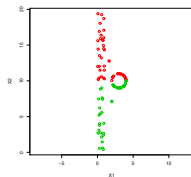
Table: Information mutuelle normalisée (100 essais)

Simulations gaussienne/Poisson/binomiale

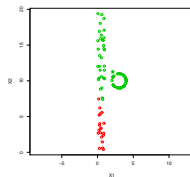
	Euclidien	Kullback-Leibler	Logistique
Gaussienne	52	35	42
Poisson	37	63	56
Binomiale	38	57	62

Table: Nombre de cas (sur 100) où chaque divergence est meilleure (au sens large)

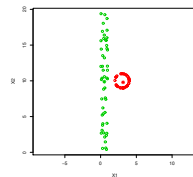
Simulations “bande et cercle”



a)



b)



c)

Figure: a = Euclidien, b = l-divergence, c = Itakura-Saito.

Simulations “bande et cercle”

Euclidien	I-divergence	Logistique	Itakura-Saito
0.302	0.327	0.320	0.986

Table: Information mutuelle normalisée (50 essais)

Simulations “bande et cercle”

Euclidien	I-divergence	Logistique	Itakura-Saito
0.302	0.327	0.320	0.986

Table: Information mutuelle normalisée (50 essais)

La distance de Itakura-Saito, non symétrique et non convexe en la deuxième variable, sépare la bande et le cercle, alors que les autres divergences de Bregman coupent les données différemment.

Simulations sur le simplexe

On simule 45 données distribuées sur le simplexe de dimension 2 suivant une loi de Dirichlet.

Simulations sur le simplexe

On simule 45 données distribuées sur le simplexe de dimension 2 suivant une loi de Dirichlet.

Euclidien	Kullback-Leibler	Logistique	Itakura-Saito
0.674	0.714	0.689	0.673

Table: Information mutuelle normalisée (100 essais)

Simulations sur le simplexe

On simule 45 données distribuées sur le simplexe de dimension 2 suivant une loi de Dirichlet.

Euclidien	Kullback-Leibler	Logistique	Itakura-Saito
0.674	0.714	0.689	0.673

Table: Information mutuelle normalisée (100 essais)

La distance de Kullback-Leibler est ici la plus appropriée.

Simulations sur le simplexe

On simule 45 données distribuées sur le simplexe de dimension 2 suivant une loi de Dirichlet.

Euclidien	Kullback-Leibler	Logistique	Itakura-Saito
0.674	0.714	0.689	0.673

Table: Information mutuelle normalisée (100 essais)

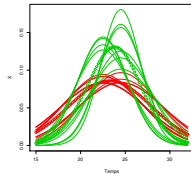
La distance de Kullback-Leibler est ici la plus appropriée.
→ Cohérent avec l'utilisation habituelle de cette divergence en classification de textes.

Simulation de courbes gaussiennes (1)

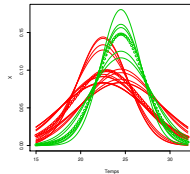
Nous considérons 40 courbes de densité gaussienne dont 20 centrées en 22.5 et 20 autres en 24.5 avec un écart-type choisi uniformément entre 2 et 5.

Simulation de courbes gaussiennes (1)

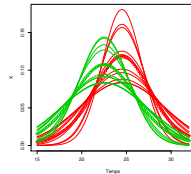
Nous considérons 40 courbes de densité gaussienne dont 20 centrées en 22.5 et 20 autres en 24.5 avec un écart-type choisi uniformément entre 2 et 5.



a)



b)



c)

Figure: a = Biais quadratique, b = Norme L^2 , c = I-divergence

Simulation de courbes gaussiennes (1)

l-divergence	Norme 2	Biais quadratique
0.910	0.799	0.017

Table: Information mutuelle normalisée (30 essais)

Simulation de courbes gaussiennes (1)

I-divergence	Norme 2	Biais quadratique
0.910	0.799	0.017

Table: Information mutuelle normalisée (30 essais)

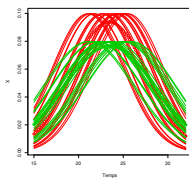
C'est la distance de Kullback-Leibler généralisée qui sépare le mieux les deux groupes.

Simulation de courbes gaussiennes (2)

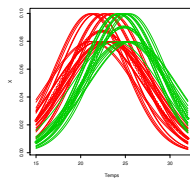
Considérons 20 courbes gaussiennes d'écart-type 4 et 20 d'écart-type 5, avec une moyenne choisie uniformément entre 21 et 26.

Simulation de courbes gaussiennes (2)

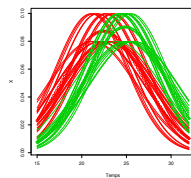
Considérons 20 courbes gaussiennes d'écart-type 4 et 20 d'écart-type 5, avec une moyenne choisie uniformément entre 21 et 26.



a)



b)



c)

Figure: a = Biais quadratique, b = Norme L^2 , c = l-divergence

Simulation de courbes gaussiennes (2)

I-divergence	Norme 2	Biais quadratique
0.017	0.022	1

Table: Information mutuelle normalisée (30 essais)

Simulation de courbes gaussiennes (2)

I-divergence	Norme 2	Biais quadratique
0.017	0.022	1

Table: Information mutuelle normalisée (30 essais)

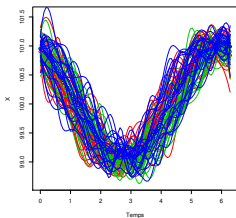
Ici, retrouver les deux groupes signifie classer les courbes selon la variance de la gaussienne (et plus selon la moyenne comme dans ce qui précède). Cette fois, c'est le biais quadratique qui conduit au meilleur résultat.

Simulation de sinusoïdes déphasées

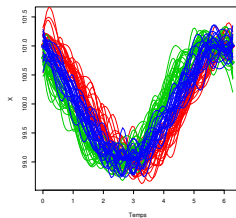
Nous avons créé 3 groupes de sinusoïdes, correspondant à trois phases différentes 0 , $\frac{\pi}{4}$ et $\frac{\pi}{8}$, avec un bruit gaussien.

Simulation de sinusoïdes déphasées

Nous avons créé 3 groupes de sinusoïdes, correspondant à trois phases différentes 0 , $\frac{\pi}{4}$ et $\frac{\pi}{8}$, avec un bruit gaussien.



a)



b)

Figure: a = Biais quadratique, b = Norme L^2

Simulation de sinusoïdes déphasées

Norme 2	Biais quadratique	Itakura-Saito
0.858	0.043	0.853

Table: Information mutuelle normalisée pour 3 groupes de sinusoïdes de phases 0 , $\frac{\pi}{6}$ et $\frac{\pi}{12}$ (30 essais)

Bibliographie

- [A. Banerjee, S. Merugu, I. S. Dhillon et J. Ghosh](#) : Clustering with Bregman Divergences. *Journal of Machine Learning Research*, 6 :1705–1749, 2005.
- [L. M. Bregman](#) : The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7 :200–217, 1967.
- [B. A. Frigiyik, S. Srivastava et M. R. Gupta](#) : Functional Bregman Divergence and Bayesian Estimation of Distributions. *IEEE International Symposium on Information Theory*, 1681–1685, 2008.
- [A. Gersho et R. M. Gray](#) : *Vector quantization and signal compression*. Kluwer Academic Publishers, 1992.
- [S. Graf et H. Luschgy](#) : *Foundations of quantization for probability distributions*. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2000.
- [T. Linder](#) : Learning-Theoretic Methods in Vector Quantization. In L. Györfi, éditeur : *Principles of Nonparametric Learning*. Springer, 2001.
- [A. Strehl et J. Ghosh](#) : Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions, *Journal of Machine Learning Research*, 3 :583–617, 2002.
- [R. Tibshirani, G. Walther et T. Hastie](#) : Estimating the number of clusters in a dataset via the Gap statistic, 2000.