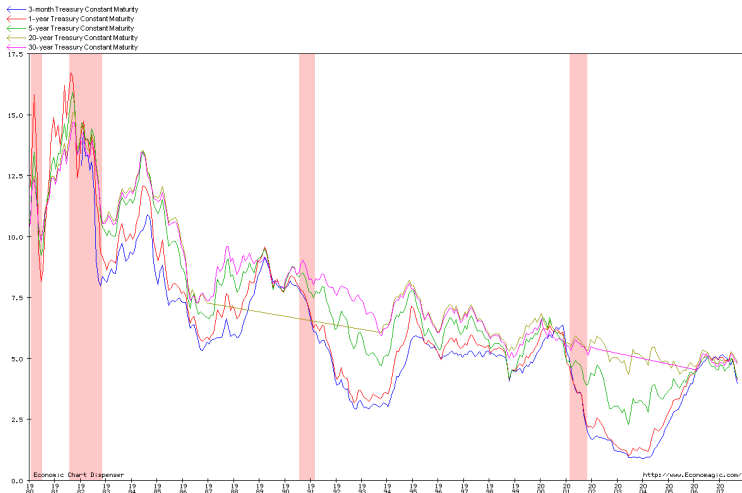


Réduction de la dimension

Exemple 1 - Finance

Analyse des taux d'intérêt



Exemple 1 - Finance (2)

- **Variables** = 18 maturités
= 1M, 3M, 6M, 9M, 1Y, 2Y, ..., 30Y
- **Observations** = Historique mensuel sur 8 ans
= 96 valeurs

Exemple 2 - Web 2.0

Last-FM - webradio de type collaboratif

 the social music revolution

MusiqueUtilisateursÉcouterÉvénementsWidgetsTélécharger

[Uploader de la musique et des vidéos](#)

Vous n'êtes pas connecté(e) [Connexion](#) | [Aide](#) | [Français](#)

Bienvenue sur votre radio

... créez votre propre station avec la musique que vous aimez

Lecture





Shakira - Whenever, Wherever-1:42

☒ Acheter





Autre station

Pop Up

Partage

last.fm

Si vous aimez Shakira, vous devriez également aimer :
Juanes, Paulina Rubio, Jennifer Lopez, Nelly Furtado, Alejandro Sanz et Christina Aguilera (voir plus...)

Exemple 2 - Web 2.0 (2)

- 28302 artistes et leurs "tags"
- **Variables** = 735 tags
= trance, techno, ambient, alternative,
rap metal, rock, ...
- **Observations** = 2840 utilisateurs

Exemple 3 - Reconnaissance de visages



Exemple 3 - Reconnaissance de visages (2)

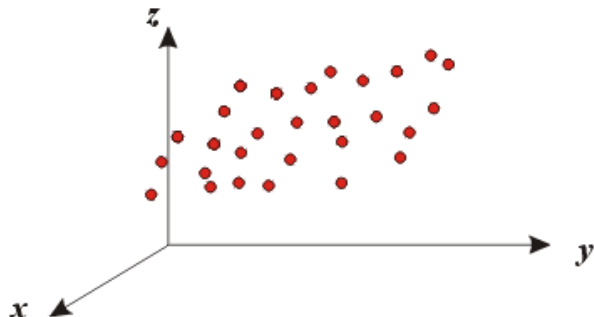
- **Variables** = 256×256 pixels
- **Observations** = 64 images

Traits communs

- Données **multivariées**
- Besoin d'**interprétation**
- **Variabilité** expliquée par des combinaisons de variables

- Dimension = nombre de **variables** = p
- Taille de l'échantillon = nombre d'**observations** = n
- Tableau $n \times p$ de variables **quantitatives**

Représentation graphique



\Rightarrow Nuage de n points dans \mathbb{R}^p

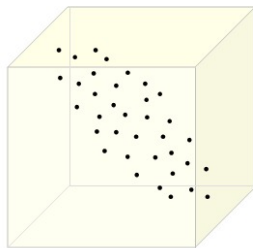
Objectifs

- Réduction de la dimension
- Visualisation du nuage en 2D ou 3D
- Explication de la variabilité

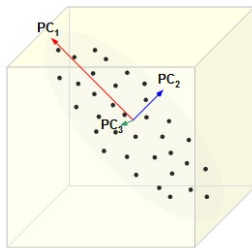
Analyse en Composantes Principales (ACP)

Philosophie de l'ACP

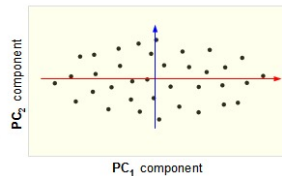
→ Projeter le nuage selon la "bonne" direction



a



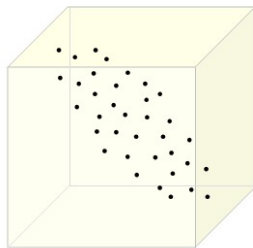
b



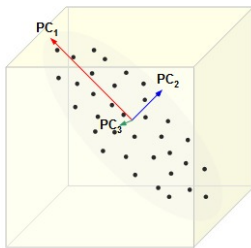
c

Philosophie de l'ACP

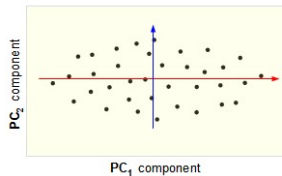
→ Projeter le nuage selon la "bonne" direction



a



b



c

Idée : maximiser la **dispersion**

Cadre statistique : Tableau de données

- Observations : $X_i \in \mathbb{R}^p$, $1 \leq i \leq n$
- Variable j : X_{1j}, \dots, X_{nj}
- Matrice $n \times p$ de données $X = (X_1, \dots, X_n)^T$

$$X = (X_{ij})_{i,j} = \begin{pmatrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}$$

Matrice de covariance empirique

- Barycentre

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \in \mathbb{R}^p$$

- Matrice de covariance empirique ($p \times p$)

$$S = (s_{kj})_{k,j} = \frac{1}{n} \sum_{i=1}^n X_i X_i^T - \bar{X} \bar{X}^T$$

Meilleure direction

- Direction de projection $a \in \mathbb{R}^p$
- Echantillon (1D) $= (a^T X_1, \dots, a^T X_n)$
- Maximiser la variance empirique en a :

$$s_a^2 = a^T S a$$

- Solution :

vecteur propre $g_{(1)}$ de la plus grande valeur propre l_1

Diagonalisation de S symétrique réelle

- Valeurs propres : $l_1 \geq \dots \geq l_p$
- Vecteurs propres orthonormés $g_{(1)}, \dots, g_{(p)}$
- Réduction de la matrice $S = GLG^T$ où
 - ▶ $L = \text{diag}(l_1, \dots, l_p)$ matrice diagonale $p \times p$
 - ▶ G matrice orthogonale $p \times p$

$$G = (g_{(1)}, \dots, g_{(p)}) = (g_{kj})_{k,j}$$

Composantes principales (CP)

- Composantes principales : pour tout vecteur $z \in \mathbb{R}^p$

$$y_j(z) = g_{(j)}^T(z - \bar{X}) , \quad 1 \leq j \leq p$$

- La matrice $n \times p$

$$Y = (y_j(X_i))_{1 \leq i \leq n, 1 \leq j \leq p}$$

remplace la matrice X des données initiales.

Corrélation empirique "Variable vs. CP"

- Corrélations empiriques entre la variable k et la CP y_j :

$$\tilde{r}_{kj} = g_{kj} \sqrt{\frac{l_j}{s_{kk}}} \quad (\text{définition})$$

- Propriété :

$$\sum_{j=1}^p \tilde{r}_{kj}^2 = 1$$

Variance empirique de la k -ème variable

- Part de la variance empirique de la k -ème variable expliquée par les 2 premières CP (y_1, y_2) :

$$\tilde{r}_{k1}^2 + \tilde{r}_{k2}^2$$

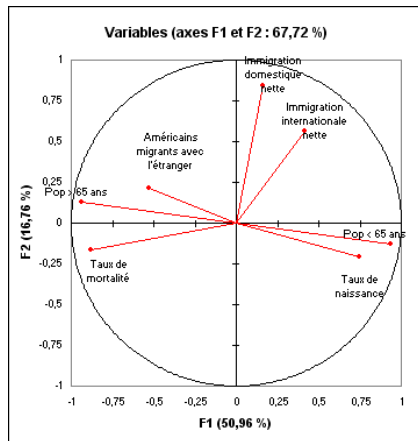
- On a :

$$I_1 + I_2 = \sum_{k=1}^p s_{kk}(\tilde{r}_{k1}^2 + \tilde{r}_{k2}^2)$$

- Visualisation 2D : **Disque des corrélations**

Disque des corrélations

- Point $(\tilde{r}_{k1}, \tilde{r}_{k2})$ correspond la variable k



Variance empirique des données

- Part de la variance empirique du nuage de points expliquée par la CP y_j :

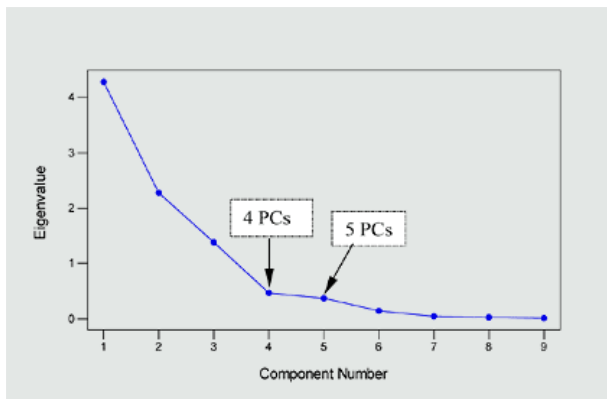
$$v_j = \frac{l_j}{\text{Tr}(S)}$$

où $\text{Tr}(S) = \sum_{j=1}^p l_j.$

- Visualisation : **scree-graph**

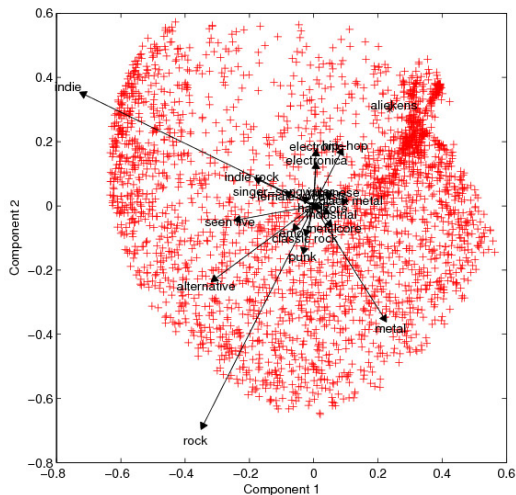
Scree-graph

- Axes = indice j de la CP et part de variance v_j



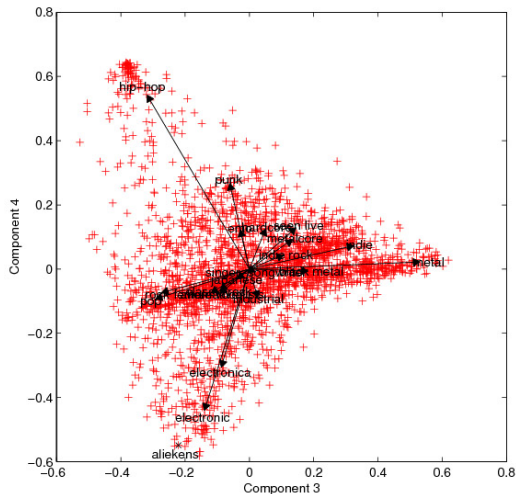
Résultats de l'ACP - Last-FM (1)

Projection du nuage de points sur (CP1, CP2)



Résultats de l'ACP - Last-FM (2)

Projection du nuage de points sur (CP3, CP4)



Résultats de l'ACP - Visages (1)

Données



Résultats de l'ACP - Visages (2)

"Visages propres"



Résultats de l'ACP - Visages (3)

Reconstruction partielle (sous-colonne de la matrice Y)



Résultats de l'ACP - Visages (4)

Projection d'autres images



Quelques remarques

- ACP = outil **linéaire**
- **Orthogonalité** des composantes principales

- **En pratique :**

Réduction de la matrice $R = (r_{kj})_{k,j}$ des **corrélations**

$$r_{kj} = \frac{s_{kj}}{\sqrt{s_{kk}s_{jj}}}$$

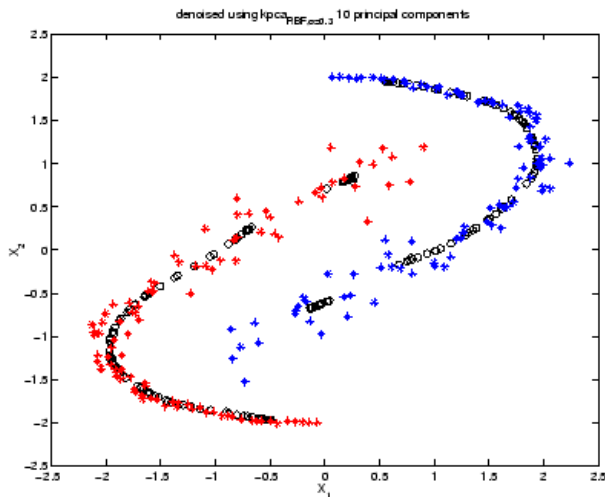
- **Obstacle numérique :**

Réduction de S en **très grande dimension**

Quand est-ce que ça marche ?

- Nuages de points **ellipsoïdaux**
- Modèle implicite = modèle **gaussien**
- Information portée par les **statistiques d'ordre 2**
- Absence de **valeurs aberrantes**

Echec de l'ACP



⇒ **Extension** : ACP non-linéaire (à noyau)

Noyaux positifs

Définition

Soit \mathcal{X} l'espace où vivent les observations.

Noyau positif

Une fonction $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ est un noyau positif si et seulement si

- ❶ k est symétrique: $k(x, x') = k(x', x)$, $\forall x, x' \in \mathcal{X}$
- ❷ k est positive:

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) \geq 0, \quad \forall c_i \in \mathbb{R}, \quad \forall x_i \in \mathcal{X}, \quad \forall n \geq 1$$

Un théorème d'analyse

Théorème de Mercer

Pour tout noyau positif k sur \mathcal{X} il existe un espace de Hilbert \mathcal{H} et une application Φ tels que:

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle, \quad \forall x, x' \in \mathcal{X}$$

où $\langle \cdot, \cdot \rangle$ représente le produit scalaire sur \mathcal{H} .

- Le théorème de Mercer est non constructif: il ne fournit ni \mathcal{H} , ni Φ

Commentaires

- Le théorème de Mercer est non constructif: il ne fournit ni \mathcal{H} , ni Φ
- En pratique:

- Le théorème de Mercer est non constructif: il ne fournit ni \mathcal{H} , ni Φ
- En pratique:
 - ▶ \mathcal{H} est un espace de grande dimension

- Le théorème de Mercer est non constructif: il ne fournit ni \mathcal{H} , ni Φ
- En pratique:
 - ▶ \mathcal{H} est un espace de grande dimension
 - ▶ Φ est une application non-linéaire

- Le théorème de Mercer est non constructif: il ne fournit ni \mathcal{H} , ni Φ
- En pratique:
 - ▶ \mathcal{H} est un espace de grande dimension
 - ▶ Φ est une application non-linéaire
- \mathcal{H} est un espace de représentation des données, connu sous le nom de "feature space" ou espace de caractéristiques

- Le théorème de Mercer est non constructif: il ne fournit ni \mathcal{H} , ni Φ
- En pratique:
 - ▶ \mathcal{H} est un espace de grande dimension
 - ▶ Φ est une application non-linéaire
- \mathcal{H} est un espace de représentation des données, connu sous le nom de "feature space" ou espace de caractéristiques
- L'astuce du noyau consiste à faire l'impasse sur \mathcal{H} et Φ si on sait qu'ils existent!

Distances images

- Norme euclidienne sur \mathbb{R}^m : $\forall u \in \mathbb{R}^m$, $\|u\| = \sqrt{\langle u, u \rangle}$ où \langle , \rangle produit scalaire sur \mathbb{R}^m

Distances images

- Norme euclidienne sur \mathbb{R}^m : $\forall u \in \mathbb{R}^m$, $\|u\| = \sqrt{\langle u, u \rangle}$ où \langle , \rangle produit scalaire sur \mathbb{R}^m
- Distance euclidienne:
$$d(u, v) = \|u - v\| = \sqrt{\langle u, u \rangle + \langle v, v \rangle - 2 \langle u, v \rangle}$$

Distances images

- Norme euclidienne sur \mathbb{R}^m : $\forall u \in \mathbb{R}^m$, $\|u\| = \sqrt{\langle u, u \rangle}$ où \langle , \rangle produit scalaire sur \mathbb{R}^m
- Distance euclidienne:
$$d(u, v) = \|u - v\| = \sqrt{\langle u, u \rangle + \langle v, v \rangle - 2 \langle u, v \rangle}$$
- Transformation non-linéaire : $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ avec $m > d$

Distances images

- Norme euclidienne sur \mathbb{R}^m : $\forall u \in \mathbb{R}^m, \|u\| = \sqrt{\langle u, u \rangle}$ où \langle , \rangle produit scalaire sur \mathbb{R}^m
- Distance euclidienne:
$$d(u, v) = \|u - v\| = \sqrt{\langle u, u \rangle + \langle v, v \rangle - 2 \langle u, v \rangle}$$
- Transformation non-linéaire : $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ avec $m > d$
- Noyau: $k(x, x') = \langle \Phi(x), \Phi(x') \rangle$

Distances images

- Norme euclidienne sur \mathbb{R}^m : $\forall u \in \mathbb{R}^m$, $\|u\| = \sqrt{\langle u, u \rangle}$ où \langle , \rangle produit scalaire sur \mathbb{R}^m
- Distance euclidienne:
$$d(u, v) = \|u - v\| = \sqrt{\langle u, u \rangle + \langle v, v \rangle - 2\langle u, v \rangle}$$
- Transformation non-linéaire : $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ avec $m > d$
- Noyau: $k(x, x') = \langle \Phi(x), \Phi(x') \rangle$
- Distance image:

$$d_\Phi(x, x') = \|\Phi(x) - \Phi(x')\| = \sqrt{k(x, x) + k(x', x') - 2k(x, x')}$$

\Rightarrow la distance induite par Φ ne fait intervenir que le noyau

Intérêt pour la classification

- **Aucune complication algorithmique** en remplaçant le produit scalaire par une autre mesure de similarité

Intérêt pour la classification

- Aucune complication algorithmique en remplaçant le produit scalaire par une autre mesure de similarité
- Transformer un problème initialement non-linéaire en un problème linéaire en envoyant les données dans un espace plus grand

Intérêt pour la classification

- Aucune complication algorithmique en remplaçant le produit scalaire par une autre mesure de similarité
- Transformer un problème initialement non-linéaire en un problème linéaire en envoyant les données dans un espace plus grand

Intérêt pour la classification

- **Aucune complication algorithmique** en remplaçant le produit scalaire par une autre mesure de similarité
- **Transformer un problème initialement non-linéaire en un problème linéaire** en envoyant les données dans un espace plus grand

Exemple

Soit $f(x, y) = ax^2 + bx + c - y = 0$ une surface de décision polynomiale (parabole dans \mathbb{R}^2).

Intérêt pour la classification

- **Aucune complication algorithmique** en remplaçant le produit scalaire par une autre mesure de similarité
- **Transformer un problème initialement non-linéaire en un problème linéaire** en envoyant les données dans un espace plus grand

Exemple

Soit $f(x, y) = ax^2 + bx + c - y = 0$ une surface de décision polynomiale (parabole dans \mathbb{R}^2).

Rôle clé de la transformation:

$$\begin{aligned}\Phi : \mathbb{R}^2 &\rightarrow \mathbb{R}^4 \\ x &\mapsto (x^2, x, 1, y)^T\end{aligned}$$

Du non-linéaire au linéaire

Exemple (suite)

On peut écrire:

$$g(x^2, x, 1, y) = ax^2 + bx + c - y = 0$$

où $g(u, v, w, y) = au + bv + cw - y$.

L'équation $g(u, v, w, y) = 0$ définit une surface de décision linéaire dans \mathbb{R}^4 .

Du non-linéaire au linéaire

Exemple (suite)

On peut écrire:

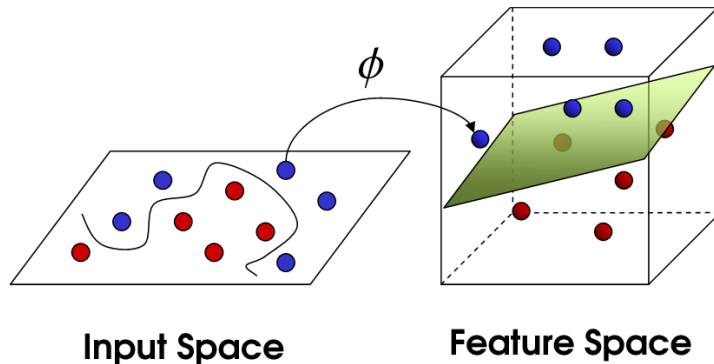
$$g(x^2, x, 1, y) = ax^2 + bx + c - y = 0$$

où $g(u, v, w, y) = au + bv + cw - y$.

L'équation $g(u, v, w, y) = 0$ définit une surface de décision linéaire dans \mathbb{R}^4 .

Un problème non-linéaire dans un certain espace peut parfois se formuler comme un problème linéaire dans un espace plus grand.

Du non-linéaire au linéaire



ACP à noyau

ACP classique

On considère un nuage de points x_1, \dots, x_n centrés en l'origine.

ACP classique

On considère un nuage de points x_1, \dots, x_n centrés en l'origine.

Buts de l'ACP

- méthode de visualisation des données
- réduction de la dimension effective des données

ACP classique

On considère un nuage de points x_1, \dots, x_n centrés en l'origine.

Buts de l'ACP

- méthode de visualisation des données
- réduction de la dimension effective des données

L'ACP consiste à identifier les composantes principales de l'échantillon constituées par

- 1 la meilleure direction de projection du nuage de points
i.e. celle de variance maximale

ACP classique

On considère un nuage de points x_1, \dots, x_n centrés en l'origine.

Buts de l'ACP

- méthode de visualisation des données
- réduction de la dimension effective des données

L'ACP consiste à identifier les composantes principales de l'échantillon constituées par

- 1 la meilleure direction de projection du nuage de points
i.e. celle de variance maximale
- 2 puis, la meilleure direction de projection orthogonale à la première

ACP classique

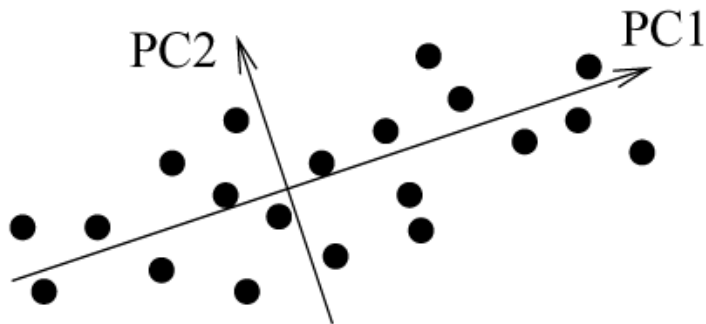
On considère un nuage de points x_1, \dots, x_n centrés en l'origine.

Buts de l'ACP

- méthode de visualisation des données
- réduction de la dimension effective des données

L'ACP consiste à identifier les composantes principales de l'échantillon constituées par

- ❶ la meilleure direction de projection du nuage de points
i.e. celle de variance maximale
- ❷ puis, la meilleure direction de projection orthogonale à la première
- ❸ et, ainsi de suite, jusqu'à la n -ième



ACP (suite)

- Projection orthogonale d'un vecteur x sur la direction $w \in \mathbb{R}^d$:

$$p_w(x) = \frac{\langle x, w \rangle}{\|w\|}$$

ACP (suite)

- Projection orthogonale d'un vecteur x sur la direction $w \in \mathbb{R}^d$:

$$p_w(x) = \frac{\langle x, w \rangle}{\|w\|}$$

- Variance empirique du nuage de points selon la direction w :

$$\mathbb{V}(p_w) = \frac{1}{n} \sum_{i=1}^n \frac{\langle x_i, w \rangle^2}{\|w\|^2}$$

ACP (suite)

- Projection orthogonale d'un vecteur x sur la direction $w \in \mathbb{R}^d$:

$$p_w(x) = \frac{\langle x, w \rangle}{\|w\|}$$

- Variance empirique du nuage de points selon la direction w :

$$\mathbb{V}(p_w) = \frac{1}{n} \sum_{i=1}^n \frac{\langle x_i, w \rangle^2}{\|w\|^2}$$

- Matrice de covariance empirique $\Sigma = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$

- Projection orthogonale d'un vecteur x sur la direction $w \in \mathbb{R}^d$:

$$p_w(x) = \frac{\langle x, w \rangle}{\|w\|}$$

- Variance empirique du nuage de points selon la direction w :

$$\mathbb{V}(p_w) = \frac{1}{n} \sum_{i=1}^n \frac{\langle x_i, w \rangle^2}{\|w\|^2}$$

- Matrice de covariance empirique $\Sigma = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$
- On a donc :

$$\mathbb{V}(p_w) = \frac{w^T \Sigma w}{\|w\|^2}$$

Problème d'optimisation

Première composante principale

$$\arg \max_w \mathbb{V}(p_w) = \frac{w^T \Sigma w}{\|w\|^2}$$

Problème d'optimisation

Première composante principale

$$\arg \max_w \mathbb{V}(p_w) = \frac{w^T \Sigma w}{\|w\|^2}$$

Solution

Les composantes principales sont les vecteurs propres de la Σ rangés selon la décroissance des valeurs propres correspondantes.

Problème d'optimisation

Première composante principale

$$\arg \max_w \mathbb{V}(p_w) = \frac{w^T \Sigma w}{\|w\|^2}$$

Solution

Les composantes principales sont les vecteurs propres de la Σ rangés selon la décroissance des valeurs propres correspondantes.

Remarque : la matrice Σ est symétrique réelle donc diagonalisable dans une base orthonormée.

ACP (suite)

On cherche un vecteur v et un réel λ tels que:

$$\Sigma v = \lambda v$$

Or, on a :

$$\Sigma v = \frac{1}{n} \sum_{i=1}^n \langle x_i, v \rangle x_i$$

D'où:

$$v = \sum_{i=1}^n \left(\frac{\langle x_i, v \rangle}{n\lambda} \right) x_i = \sum_{i=1}^n \alpha_i x_i$$

ACP (suite)

On cherche un vecteur v et un réel λ tels que:

$$\Sigma v = \lambda v$$

Or, on a :

$$\Sigma v = \frac{1}{n} \sum_{i=1}^n \langle x_i, v \rangle x_i$$

D'où:

$$v = \sum_{i=1}^n \left(\frac{\langle x_i, v \rangle}{n\lambda} \right) x_i = \sum_{i=1}^n \alpha_i x_i$$

On utilise

$$x_j^T \Sigma v = \lambda \langle x_j, v \rangle, \quad \forall j$$

et on y substitue les expressions de Σ et v :

$$\frac{1}{n} \sum_{i=1}^n \alpha_i \left\langle x_j, \sum_{k=1}^n \langle x_k, x_i \rangle x_k \right\rangle = \lambda \sum_{i=1}^n \alpha_i \langle x_j, x_i \rangle$$

ACP (suite)

- On note $K = (\langle x_i, x_j \rangle)_{i,j}$ la matrice de Gram

ACP (suite)

- On note $K = (< x_i, x_j >)_{i,j}$ la matrice de Gram
- On peut écrire alors le système:

$$K^2 \alpha = n \lambda K \alpha$$

ACP (suite)

- On note $K = (< x_i, x_j >)_{i,j}$ la matrice de Gram
- On peut écrire alors le système:

$$K^2 \alpha = n \lambda K \alpha$$

- Pour résoudre en α , on résout donc le problème aux éléments propres

$$K \alpha = n \lambda \alpha$$

Défauts de l'ACP classique

- elle est adaptée surtout au cas de réalisations de gaussiennes multivariées

Défauts de l'ACP classique

- elle est adaptée surtout au cas de réalisations de gaussiennes multivariées
 - ▶ en général, la non-corrélation n'implique pas l'indépendance des directions principales

Défauts de l'ACP classique

- elle est adaptée surtout au cas de réalisations de gaussiennes multivariées
 - ▶ en général, la non-corrélation n'implique pas l'indépendance des directions principales
 - ▶ alternative : Analyse en Composantes **Indépendantes** (plutôt que Principales)

Défauts de l'ACP classique

- elle est adaptée surtout au cas de réalisations de gaussiennes multivariées
 - ▶ en général, la non-corrélation n'implique pas l'indépendance des directions principales
 - ▶ alternative : Analyse en Composantes **Indépendantes** (plutôt que Principales)
- elle est adaptée aux structures linéaires

Défauts de l'ACP classique

- elle est adaptée surtout au cas de réalisations de gaussiennes multivariées
 - ▶ en général, la non-corrélation n'implique pas l'indépendance des directions principales
 - ▶ alternative : Analyse en Composantes **Indépendantes** (plutôt que Principales)
- elle est adaptée aux structures linéaires
 - ▶ les nuages de points ne sont pas tous ellipsoïdaux!!

Défauts de l'ACP classique

- elle est adaptée surtout au cas de réalisations de gaussiennes multivariées
 - ▶ en général, la non-corrélation n'implique pas l'indépendance des directions principales
 - ▶ alternative : Analyse en Composantes **Indépendantes** (plutôt que Principales)
- elle est adaptée aux structures linéaires
 - ▶ les nuages de points ne sont pas tous ellipsoïdaux!!
 - ▶ alternative : **Kernel PCA**

- On applique une transformation Φ qui envoie le nuage de points X dans un espace où la structure est linéaire

- On applique une transformation Φ qui envoie le nuage de points X dans un espace où la structure est linéaire
- La matrice de covariance de $\Phi(X) = (\Phi(x_1), \dots, \Phi(x_n))^T$ est alors

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \Phi(x_i) \Phi(x_i)^T$$

- On applique une transformation Φ qui envoie le nuage de points X dans un espace où la structure est linéaire
- La matrice de covariance de $\Phi(X) = (\Phi(x_1), \dots, \Phi(x_n))^T$ est alors

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \Phi(x_i) \Phi(x_i)^T$$

- **Astuce du noyau** : $K = (k(x_i, x_j))_{i,j} = (\Phi(x_i)^T \Phi(x_j))_{i,j}$

ACP à noyau (suite)

- "Directions" principales de la forme:

$$p_i(x) = \sum_{j=1}^n \alpha_{i,j} k(x_j, x)$$

ACP à noyau (suite)

- "Directions" principales de la forme:

$$p_i(x) = \sum_{j=1}^n \alpha_{i,j} k(x_j, x)$$

- le vecteur $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,n})$ est solution du problème d'optimisation:

$$\max_{\alpha} \frac{\alpha^T K^2 \alpha}{\alpha^T K \alpha}$$

sous les contraintes: $\alpha_i^T K \alpha_j$ pour $j = 1, \dots, i-1$

ACP à noyau (suite)

- "Directions" principales de la forme:

$$p_i(x) = \sum_{j=1}^n \alpha_{i,j} k(x_j, x)$$

- le vecteur $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,n})$ est solution du problème d'optimisation:

$$\max_{\alpha} \frac{\alpha^T K^2 \alpha}{\alpha^T K \alpha}$$

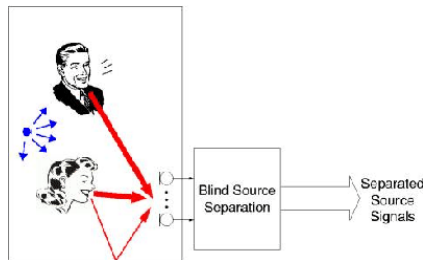
sous les contraintes: $\alpha_i^T K \alpha_j$ pour $j = 1, \dots, i-1$

- on résout le problème aux éléments propres:

$$K\alpha = n\lambda\alpha$$

Analyse en Composantes Indépendantes (ACI)

Problème du "cocktail-party"



- ACP = fondée sur la notion de **corrélation**
- Bonne notion = notion d'**indépendance**

Corrélation vs. Indépendance

- Or : X et Y indépendants $\Rightarrow \text{cov}(X, Y) = 0$
- Réciproque fausse en général, sauf cas gaussien...
- De l'ACP vers l'ACI... (beaucoup plus difficile !)

Formulation du problème

- $S = (S_1, \dots, S_d)^T$ sources indépendantes et non-gaussiennes inconnues
- \mathbf{A} matrice de mélange $d \times d$ inconnue
- $X = (X_1, \dots, X_d)^T$ observations (capteurs), on suppose $\text{Cov}(X) = \mathbf{I}$
- On a le système : $X = \mathbf{A}S$
- On cherche \mathbf{A} orthogonale telle que :

$$S = \mathbf{A}^T X \quad \text{ait des composantes indépendantes}$$

Théorie de l'information

- Entropie d'une v.a. $Z \sim p(z)$:

$$H(Z) = -\mathbb{E}(\log(p(Z)))$$

- Considérons les v.a. T de variance v , alors

$$Z \sim \mathcal{N}(0, 1) \rightarrow \max_T H(T)$$

- Information mutuelle pour $S = (S_1, \dots, S_d)^T$:

$$I(S) = \sum_{i=1}^d H(S_i) - H(S)$$

ACI par méthode entropique

- Propriété de l'entropie : si $S = \mathbf{A}^T X$

$$H(S) = H(X) + \log(|\det(\mathbf{A})|)$$

- On a donc le problème d'optimisation suivant :

$$\rightarrow \min_{\mathbf{A}: \mathbf{A}^T \mathbf{A} = \mathbf{I}} I(\mathbf{A}^T X) = \sum_{i=1}^d H(S_i) - H(X)$$

- Interprétation : écart du comportement gaussien (minimisation de l'entropie des composantes)