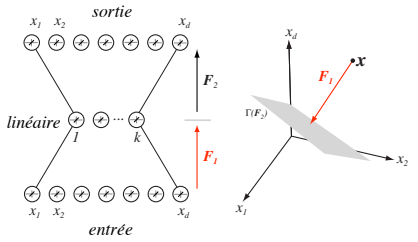
Apprentissage non-supervisé

- Typologie de la réduction de dimension
  - méthode de base: ACP
  - “groupement (clustering) des dimensions”
  - extensions:
    - ACP non-linéaire (NLPCA)
    - échelonnement multidimensionnel (multidimensional scaling – MDS)
    - cartes auto-organisatrices (self-organizing maps – SOM)
    - local linear embedding (LLE)
    - ISOMAP
    - courbes principales (principal curves)

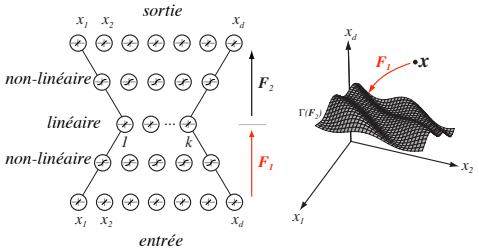
Apprentissage non-supervisé

- ACP non-linéaire – auto-encodage
  - modèle de réseau de ACP



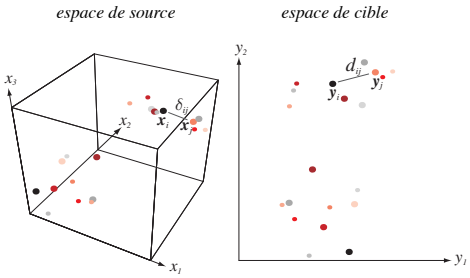
Apprentissage non-supervisé

- ACP non-linéaire – auto-encodage
  - extension non-linéaire



Apprentissage non-supervisé

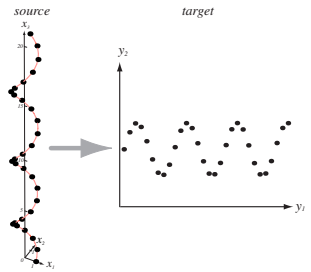
- Échelonnement multidimensionnel (MDS)
  - représentation de dimension réduite qui préserve les distances



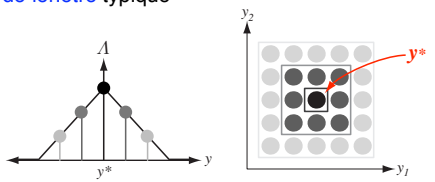
- Échelonnement multidimensionnel (MDS)
  - fonctions d'erreur
    - $J_{ee} = \frac{\sum_{i<j}(d_{ij} - \delta_{ij})^2}{\sum_{i<j}\delta_{ij}^2}$
    - $J_{ff} = \sum_{i<j}\left(\frac{d_{ij} - \delta_{ij}}{\delta_{ij}}\right)^2$
    - $J_{ef} = \frac{1}{\sum_{i<j}\delta_{ij}}\sum_{i<j}\frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}}$

- Échelonnement multidimensionnel (MDS)
  - minimisation
    - descente de gradient standard
  - initialisation
    - les d' coordonnées avec les variances plus grandes
    - ACP avec d' composantes

- Échelonnement multidimensionnel (MDS)



- Cartes auto-organisatrices (SOM)
  - $\mathbf{x}_i$ appartient à V_ℓ avec un poids $\mathbf{W}_{i,\ell}$
  - $\mathbf{W}_{i,\ell}$ ne dépend que de la distance entre $\mathbf{v}_\ell$ et $\mathbf{v}_{(\mathbf{x}_i)}$
  - fonction de fenêtre typique

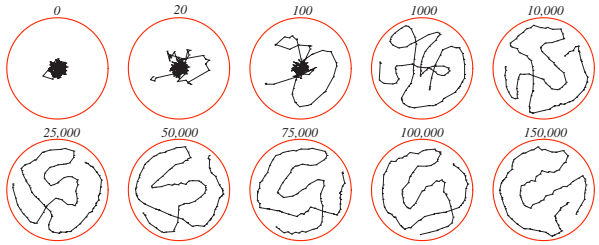


- Cartes auto-organisatrices (SOM)

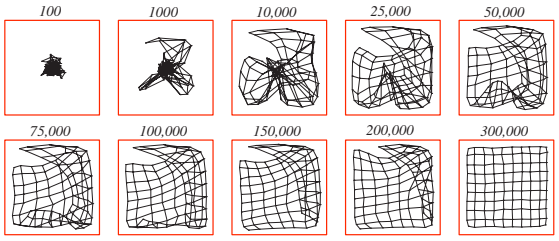
SOM(X_n)

1 $C^{(0)} \leftarrow \left\{\mathbf{v}_1^{(0)}, \ldots, \mathbf{v}_k^{(0)}\right\}$
2 $j \leftarrow 0$
3 **faire**
4 recalculer $\mathbf{W}^{(j)}$
5 **pour** $\ell \leftarrow 1$ **à** k **faire**
6 $\mathbf{v}_\ell^{(j+1)} \leftarrow \frac{1}{n}\sum_{i=1}^{n} \mathbf{W}_{i,\ell}^{(j)} \mathbf{x}_i$
7 $j \leftarrow j+1$
8 **jusqu'à** *changement* $>$ *seuil*

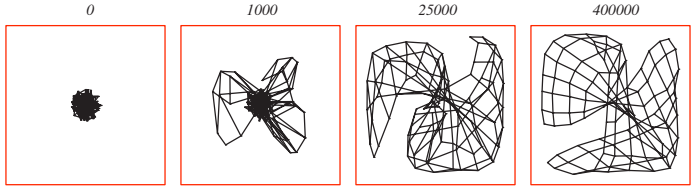
- Cartes auto-organisatrices (SOM)
  - 2 dimensions ⟶ 1 dimension



- Cartes auto-organisatrices (SOM)
  - 2 dimensions ⟶ 2 dimensions

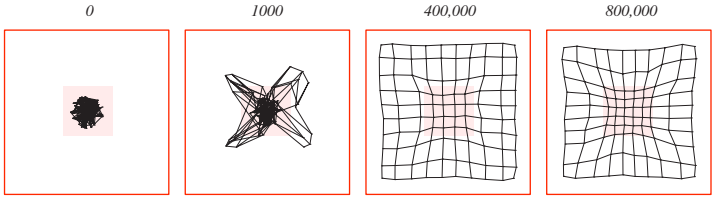


- Cartes auto-organisatrices (SOM)
  - problème: minimum local



Apprentissage non-supervisé

- Cartes auto-organisatrices (SOM)
  - estimation de densité



Apprentissage non-supervisé

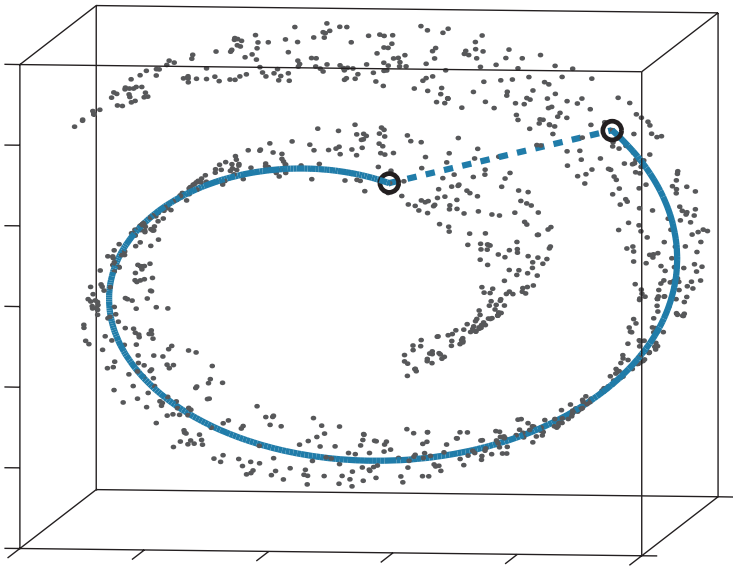
- Cartes auto-organis. (SOM) – théorie de communication
- Codage de source – quantification vectorielle:
  - fonction d'erreur: $J_s = \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{v}_{(\mathbf{x}_i)}\|^2$

Apprentissage non-supervisé

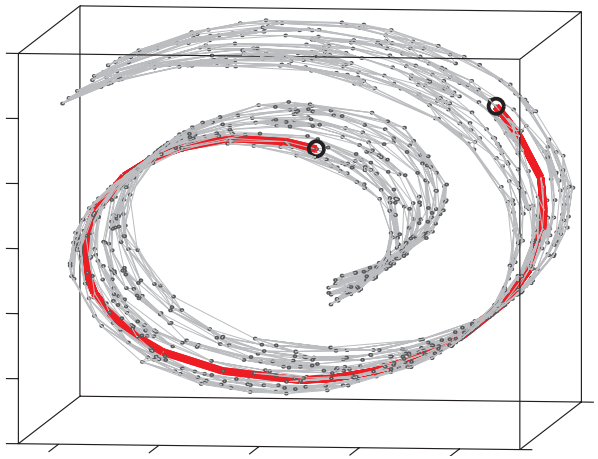
- Codage de canal – correction d'erreur:
  - probabilité d'erreur d'un bit: p
  - distance de Hamming entre des mots de code: $d_{i,j} = d_H\big(c(\mathbf{v}_i), c(\mathbf{v}_j)\big)$
  - probabilité d'erreur de code: $p_{i,j} = p^{d_{i,j}}(1-p)^{d-d_{i,j}}$
  - fonction d'erreur: $J_c = \sum_{i=1}^{n}\sum_{j=1}^{c} \|\mathbf{v}_{(\mathbf{x}_i)} - \mathbf{v}_j\|^2 p_{x_i,j}$
- Codage conjoint de canal-source
  - fonction d'erreur: $J_{s+c} = \sum_{i=1}^{n}\sum_{j=1}^{c} \|\mathbf{x}_i - \mathbf{v}_j\|^2 p_{x_i,j}$

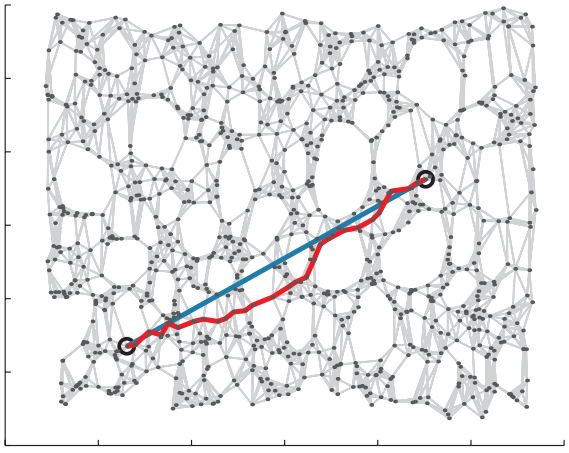
Apprentissage non-supervisé

- Problème générale: surfaces compliquées → minima local
- Solution 1: ISOMAP
  - distance geodésique: chemins plus courts dans le graphe de similarité
  - MDS standard sur les distances geodésiques

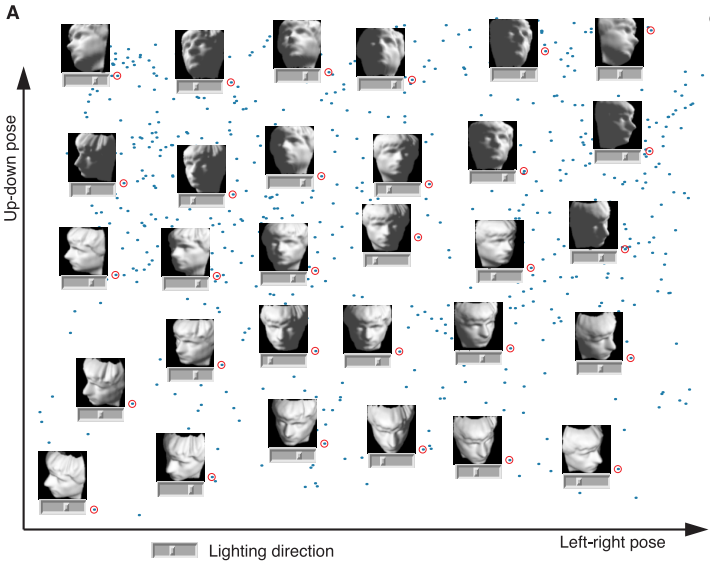


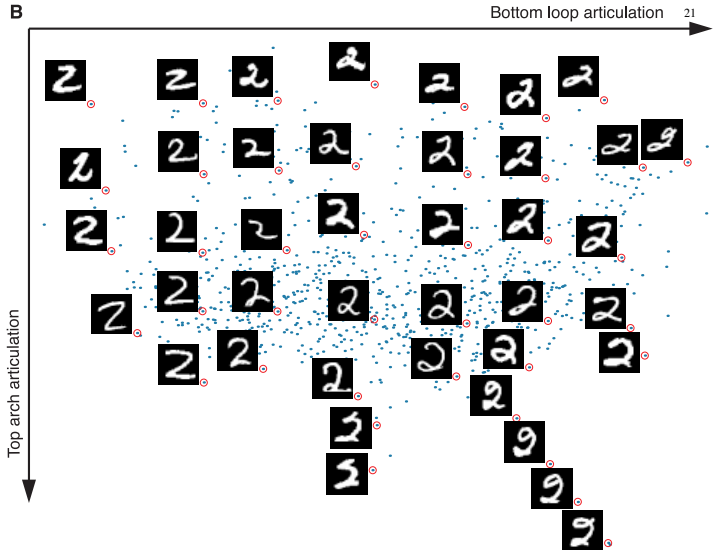
C





A





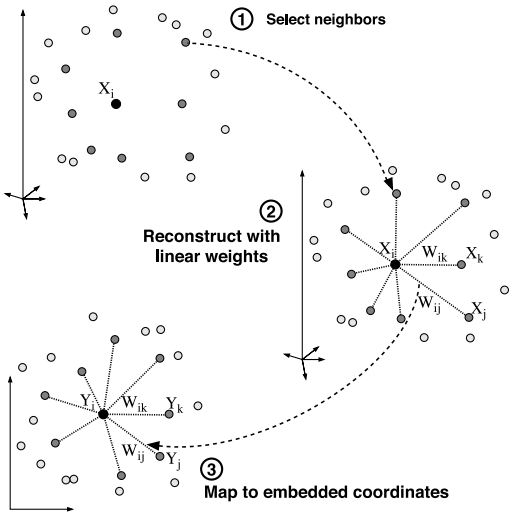
Apprentissage non-supervisé

- Solution 2: Local linear embedding (LLE)
  - Étape 1: trouver l'ensemble des voisins $V_{\mathbf{x}_i}$
  - Étape 2: approximer les points avec une combinaison linéaire de leurs plus proches voisins:

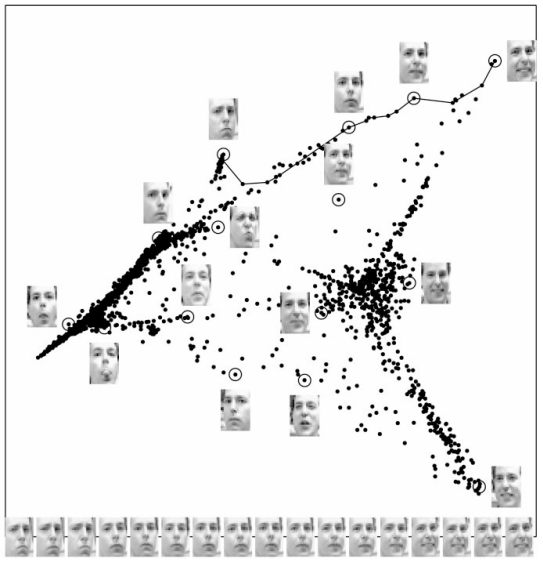
$$\min_{\mathbf{W}} \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in V_{\mathbf{x}_i}} w_{i,j} \mathbf{x}_j \right\|^2$$

  - Étape 3: reconstruire les points dans l'espace de projection en utilisant les mêmes poids:

$$\min_{\mathbf{Y}} \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{\mathbf{x}_j \in V_{\mathbf{x}_i}} w_{i,j} \mathbf{y}_j \right\|^2$$



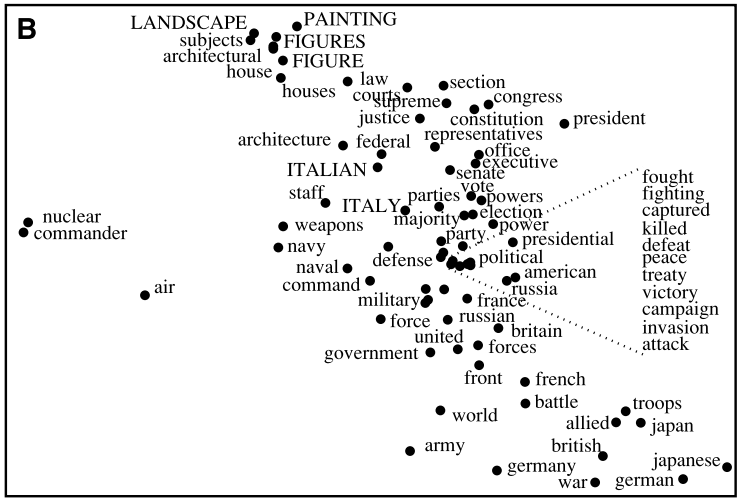
The con
these recons
symmetry: f
are invaria
translations
bors. By sy
struction w
metric prop
opposed to
ticular fram
invariance t
forced by t
rows of the
Suppose
nonlinear m
$\ll D$. To a
exists a li
translation,
maps the l
each neighb
nates on the
struction we
ric propertie
exactly suc
expect their
try in the o



valid for
particular
struct the
should al
fold coor
LLE c
mapping
step of th
observatio
vector $\vec{Y}_i$
nates on th
d-dimensi
embeddin

Φ

This cost
based on
but here
mizing th
cost in E
vectors $\vec{Y}$
the proble
by solvin
lem (9),
tors prov
coordinat
Imple
straightfo
points we
est neighb
tance or



Apprentissage non-supervisé

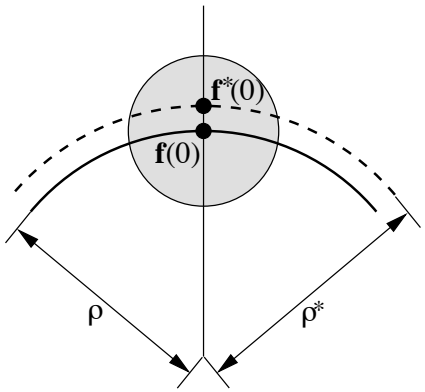
- désavantage d'ISOMAP:
  - temps d'exécution: $O(n^3)$
- projeter des nouveaux points
  - construire la fonction de projection explicitement
  - problème d'interpolation
  - problème d'apprentissage supervisé (régression multidimensionnelle)

Apprentissage non-supervisé

- Problème: bruit

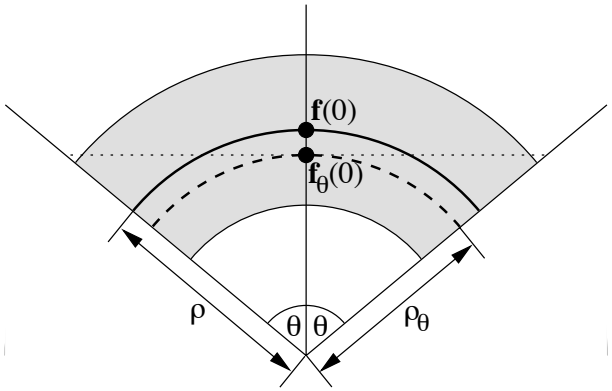
Apprentissage non-supervisé

- Le biais du modèle



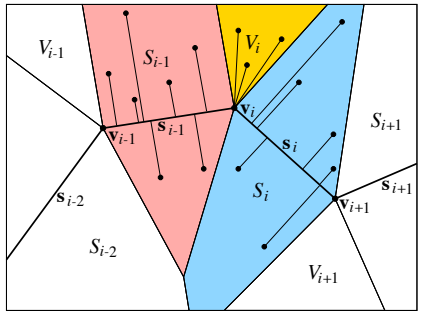
Apprentissage non-supervisé

- Le biais de l'estimation



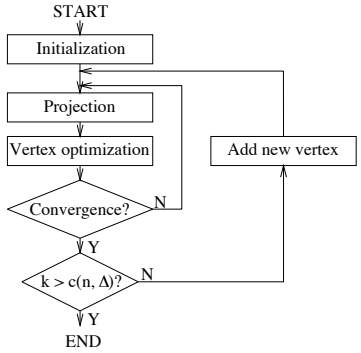
Apprentissage non-supervisé

- Solution: courbes principales polygonales
  - Mesurer la distance de la courbe au lieu des sommets



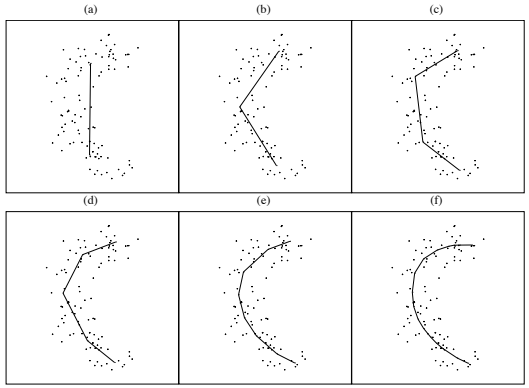
Apprentissage non-supervisé

- Courbes principales polygonales

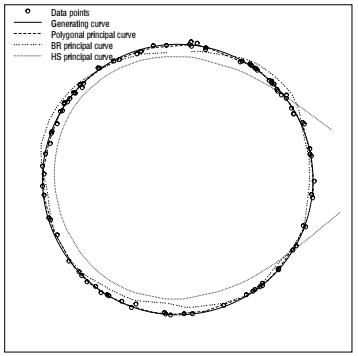


Apprentissage non-supervisé

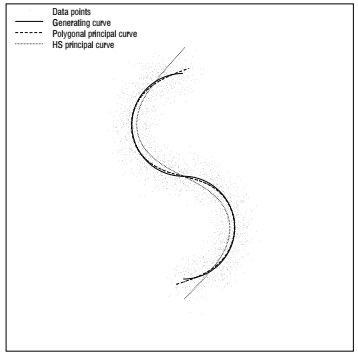
- Courbes principales polygonales



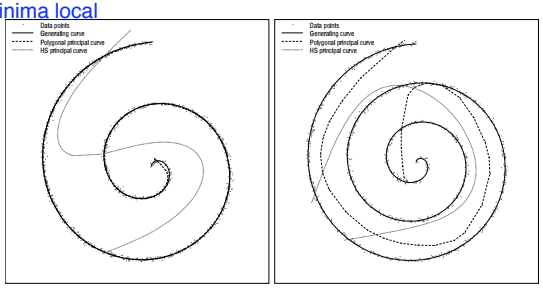
- Courbes principales polygonales
  - bruit réduit



- Courbes principales polygonales
  - beaucoup de points

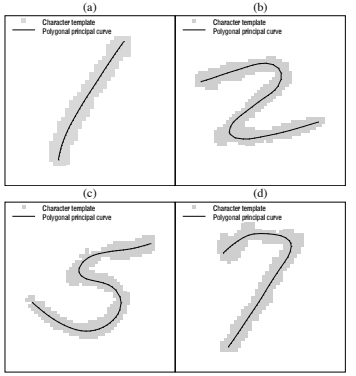


- désavantages des courbes principales:
  - minima local

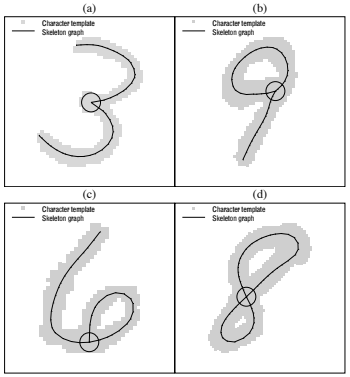


  - extension aux surfaces n'est pas évident
    → la plupart des applications sont dans le traitement d'image

- Skeletisation des caractères

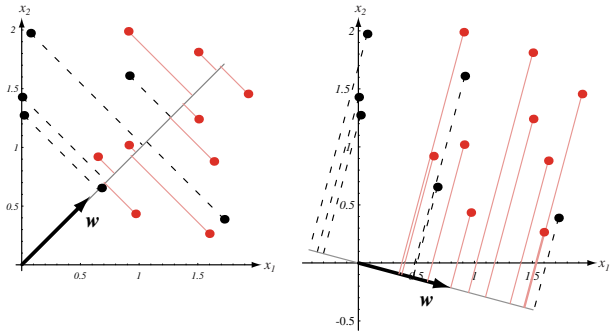


- Skeletisation des caractères



- Apprentissage non-supervisé pour la classification: analyse discriminante
  - but: trouver la meilleure projection qui préserve l'information discriminante
- Discriminante de Fisher
  - $y = \mathbf{w}^t \mathbf{x}$

- Analyse discriminante



- Idée 1: séparer les moyennes projetées
  - $_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}$
  - $\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{w}^t \mathbf{x}$
  - trouver $\mathbf{w}$ qui maximise $|\tilde{m}_1 - \tilde{m}_2| = |\mathbf{w}^t(_1 - _2)|$
- Idée 2: séparer les moyennes projetées normalisées par les variances par classe
  - $\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2$
  - $J(\mathbf{w}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$

Apprentissage non-supervisé

- Maximiser $J(\mathbf{w})$:
  - $\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - _i)(\mathbf{x} - _i)^t$
  - $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$
  - $\tilde{s}_i^2 = \sum_{\mathbf{x} \in D_i} (\mathbf{w}^t \mathbf{x} - \mathbf{w}_i^t)^2 = \sum_{\mathbf{x} \in D_i} \mathbf{w}^t (\mathbf{x} - _i)(\mathbf{x} - _i)^t \mathbf{w} = \mathbf{w}^t S_i \mathbf{w}$
  - $\tilde{s}_1^2 + \tilde{s}_2^2 = \mathbf{w}^t S_W \mathbf{w}$
  - $\mathbf{S}_B = (_1 - _2)(_1 - _2)^t$
  - $(\tilde{m}_1 - \tilde{m}_2)^2 = (\mathbf{w}_1^t - \mathbf{w}_2^t)^2 = \mathbf{w}^t (_1 - _2)(_1 - _2)^t \mathbf{w} = \mathbf{w}^t S_B \mathbf{w}$
  - $J(\mathbf{w}) = \dfrac{\mathbf{w}^t S_B \mathbf{w}}{\mathbf{w}^t S_W \mathbf{w}}$
  - $\mathbf{w}_{max} = S_W^{-1} (_1 - _2)$