Reconnaissance des formes

Analyse Factorielle Discriminante

- Chapitre 3 -

Analyse factorielle discriminante

Objectif

Contexte : Chaque individu $\boldsymbol{x}_i$ du tableau $\boldsymbol{X}$ est considéré comme un point d'un espace vectoriel $\mathcal{E}$ de dimension p.
L'ensemble des n individus constitue un nuage de points dans $\mathcal{E}$.

On suppose que
— n_1 individus appartiennent à la classe ω_1
— n_2 individus appartiennent à la classe ω_2
tel que $n = n_1 + n_2$.

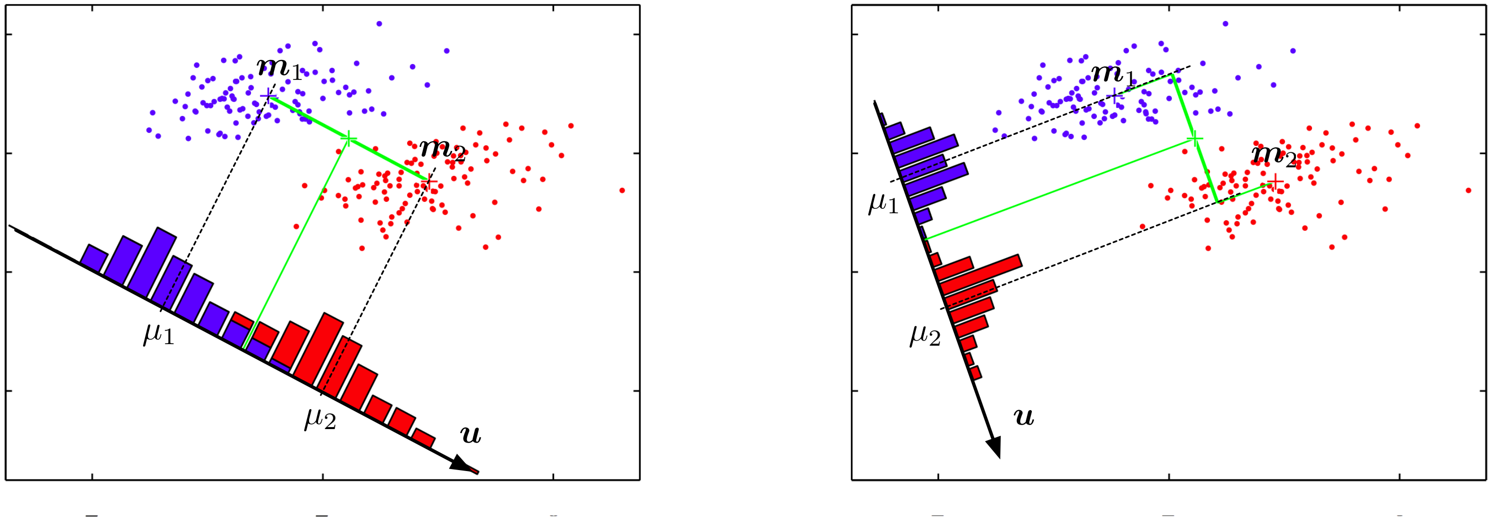
Objectif : On cherche à projeter les données sur un axe $\Delta(\boldsymbol{u})$ de vecteur directeur $\boldsymbol{u}$ maximisant la *séparabilité* de ω_1 et ω_2.
La règle de décision s'écrit

$$d(\boldsymbol{x}) = \begin{cases} \omega_2 & \text{si } \lambda(\boldsymbol{x}) = \boldsymbol{u}^\top \boldsymbol{x} > \lambda_0 \\ \omega_1 & \text{sinon} \end{cases}$$

Analyse factorielle discriminante

Objectif

Objectif : maximiser la séparabilité des données



Analyse factorielle discriminante

Mesure de séparabilité

Séparabilité : Afin de trouver un axe discriminant $\Delta(\boldsymbol{u})$, il convient de définir une mesure de séparabilité.

On note $\boldsymbol{m}_1$ et $\boldsymbol{m}_2$ les moyennes de chaque classe, soit

$$\boldsymbol{m}_1 = \frac{1}{n_1} \sum_{\boldsymbol{x} \in \omega_1} \boldsymbol{x} \qquad \boldsymbol{m}_2 = \frac{1}{n_2} \sum_{\boldsymbol{x} \in \omega_2} \boldsymbol{x}$$

Solution (mauvaise) par l'exemple : On pourrait, par exemple, considérer l'écart quadratique entre les moyennes projetées de chaque classe

$$J(\boldsymbol{u}) = (\boldsymbol{u}^\top [\boldsymbol{m}_1 - \boldsymbol{m}_2])^2 = (\mu_1 - \mu_2)^2$$

Les figures précédentes montrent qu'il convient de prendre en compte la dispersion des classes.

Analyse factorielle discriminante

Discriminant de Fisher

Critère utilisé : Le critère de Fisher suggère de maximiser l'écart inter-classe, normalisé par une mesure de la dispersion intra-classe.

Le discriminant linéaire de Fisher est défini par $\lambda(\boldsymbol{x}) = \boldsymbol{u}^\top \boldsymbol{x}$, où $\boldsymbol{u}$ maximise le critère dit de Fisher donné par

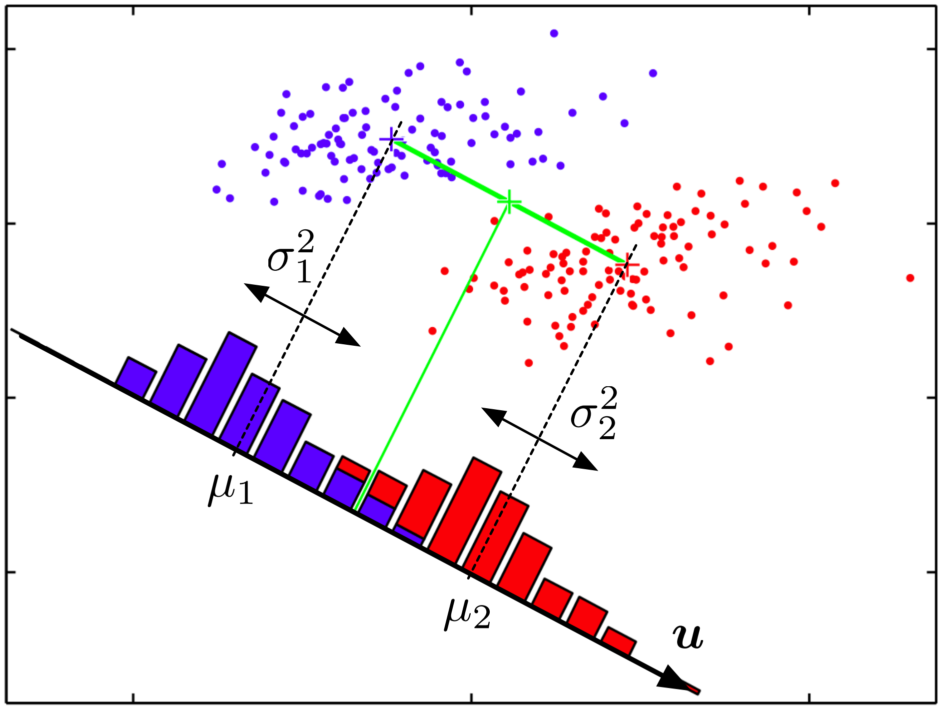
$$J(\boldsymbol{u}) = \frac{(\mu_1 - \mu_2)^2}{p_1\,\sigma_1^2 + p_2\,\sigma_2^2}$$

avec

$$\begin{aligned}
\mu_i &= \boldsymbol{u}^\top \boldsymbol{m}_i \\
\sigma_i^2 &= \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} (\boldsymbol{u}^\top [\boldsymbol{x}_i - \boldsymbol{m}_i])^2 \\
p_i &= \frac{n_i}{n}
\end{aligned}$$

Analyse factorielle discriminante

Discriminant de Fisher



Analyse factorielle discriminante

Discriminant de Fisher

- On définit les quantités suivantes dans l'espace des individus

$$\boldsymbol{\Sigma}_i = \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} (\boldsymbol{x} - \boldsymbol{m}_i)\,(\boldsymbol{x} - \boldsymbol{m}_i)^\top$$

$$\boldsymbol{\Sigma}_w = p_1 \boldsymbol{\Sigma}_1 + p_2 \boldsymbol{\Sigma}_2$$

$$\boxed{\text{La matrice } \boldsymbol{\Sigma}_w \text{ est appelée matrice de dispersion intra-classe}}$$

- On peut exprimer $\boldsymbol{\Sigma}_w$ en fonction de σ_1^2 et σ_2^2. En effet

$$\begin{aligned}\sigma_i^2 &= \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} \boldsymbol{u}^\top (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^\top \boldsymbol{u} \\ &= \boldsymbol{u}^\top \boldsymbol{\Sigma}_i \, \boldsymbol{u}\end{aligned}$$

En conséquence, on a

$$\boxed{p_1 \sigma_1^2 + p_2 \sigma_2^2 = \boldsymbol{u}^\top \boldsymbol{\Sigma}_w \, \boldsymbol{u}}$$

Analyse factorielle discriminante

Discriminant de Fisher

▷ On définit les quantités suivantes dans l'espace des individus

$$\boldsymbol{\Sigma}_b = (\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^\top$$

$\boxed{\text{La matrice } \boldsymbol{\Sigma}_b \text{ est appelée matrice de dispersion inter-classe}}$

Remarque : la matrice $\boldsymbol{\Sigma}_b$ est de rang 1.

▷ On peut exprimer $\boldsymbol{\Sigma}_b$ en fonction de μ_1 et μ_2. En effet

$$\begin{aligned}(\mu_1 - \mu_2)^2 &= [\boldsymbol{u}^\top(\boldsymbol{m}_1 - \boldsymbol{m}_2)]^2 \\ &= \boldsymbol{u}^\top \boldsymbol{\Sigma}_b\, \boldsymbol{u}\end{aligned}$$

Analyse factorielle discriminante

Discriminant de Fisher

Définition : Le critère de Fisher s'exprime ainsi :

$$J(\boldsymbol{u}) = \frac{\boldsymbol{u}^\top \boldsymbol{\Sigma}_b\, \boldsymbol{u}}{\boldsymbol{u}^\top \boldsymbol{\Sigma}_w\, \boldsymbol{u}}$$

Maximisation : On calcule le gradient de $J(\boldsymbol{u})$, qu'on annule ensuite

$$\begin{aligned}
\nabla J(\boldsymbol{u}) &= (\boldsymbol{u}^\top \boldsymbol{\Sigma}_w\, \boldsymbol{u})\, \nabla\{\boldsymbol{u}^\top \boldsymbol{\Sigma}_b\, \boldsymbol{u}\} - (\boldsymbol{u}^\top \boldsymbol{\Sigma}_b\, \boldsymbol{u})\, \nabla\{\boldsymbol{u}^\top \boldsymbol{\Sigma}_w\, \boldsymbol{u}\} \\
&= 2(\boldsymbol{u}^\top \boldsymbol{\Sigma}_w\, \boldsymbol{u})\boldsymbol{\Sigma}_b\, \boldsymbol{u} - 2(\boldsymbol{u}^\top \boldsymbol{\Sigma}_b\, \boldsymbol{u})\boldsymbol{\Sigma}_w\, \boldsymbol{u} \\
\nabla J(\boldsymbol{u}) = 0 \qquad &\Rightarrow \qquad (\boldsymbol{u}^\top \boldsymbol{\Sigma}_w\, \boldsymbol{u})\boldsymbol{\Sigma}_b\, \boldsymbol{u} = (\boldsymbol{u}^\top \boldsymbol{\Sigma}_b\, \boldsymbol{u})\boldsymbol{\Sigma}_w\, \boldsymbol{u}
\end{aligned}$$

En divisant chaque membre par $(\boldsymbol{u}^\top \boldsymbol{\Sigma}_w\, \boldsymbol{u})$, on aboutit à

$$\boxed{\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}_b\, \boldsymbol{u} = J(\boldsymbol{u})\, \boldsymbol{u}}$$

Analyse factorielle discriminante

Discriminant de Fisher

Problème : Pour maximiser $J(\boldsymbol{u})$, Il convient de résoudre

$$\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}_b\,\boldsymbol{u} = J(\boldsymbol{u})\,\boldsymbol{u}$$

▷ On sait que $\boldsymbol{\Sigma}_b = (\boldsymbol{m}_1 - \boldsymbol{m}_2)\,(\boldsymbol{m}_1 - \boldsymbol{m}_2)^\top$. En remplaçant ci-dessus, on obtient

$$J(\boldsymbol{u})\,\boldsymbol{u} = \boldsymbol{\Sigma}_w^{-1}\,(\boldsymbol{m}_1 - \boldsymbol{m}_2)\,\underbrace{(\boldsymbol{m}_1 - \boldsymbol{m}_2)^\top\,\boldsymbol{u}}_{\text{constante}}$$

▷ A une constante près, ceci implique que :

$$\boxed{\boldsymbol{u} = \boldsymbol{\Sigma}_w^{-1}\,(\boldsymbol{m}_1 - \boldsymbol{m}_2)}$$

Le discriminant $\lambda(\boldsymbol{x}) = \boldsymbol{u}^\top\boldsymbol{x}$ correspondant est dit *discriminant de Fisher* (1936).

Analyse factorielle discriminante

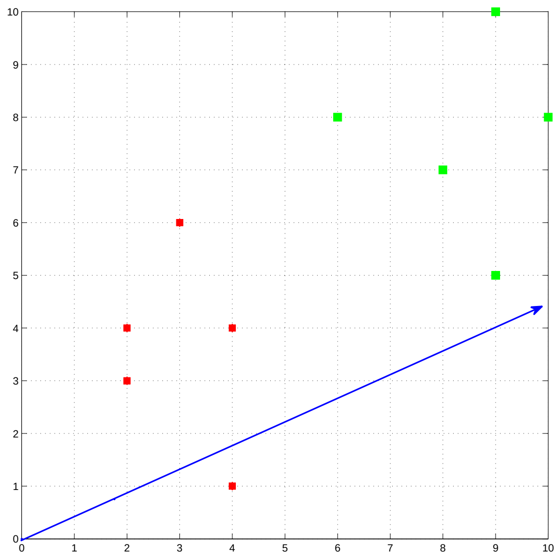
Exemple

Données : On considère les deux classes d'individus dans $\mathbb{R}^2$ suivantes :

$$\omega_1 = \{(4,1);(2,4);(2,3);(3,6);(4,4)\}$$

$$\omega_2 = \{(9,10);(6,8);(9,5);(8,7);(10,8)\}$$

$$\Rightarrow \qquad \boldsymbol{u} = (1.76, 0.75)^\top$$



Critères du second ordre

Taxonomie

Les critères *du second ordre*, tels que le critère de Fisher se contentent d'une quantité d'information modeste, limitée aux 2 premiers moments de $\lambda(\boldsymbol{x})$:

$$\begin{aligned} \mu_i &= \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} \lambda(\boldsymbol{x}) = \boldsymbol{u}^\top \boldsymbol{m}_i - \lambda_0 \\ \sigma_i^2 &= \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} (\lambda(\boldsymbol{x}) - \mu_i)^2 = \boldsymbol{u}^\top \boldsymbol{\Sigma}_i \, \boldsymbol{u} \end{aligned}$$

Exemple : le rapport signal-sur-bruit généralisé

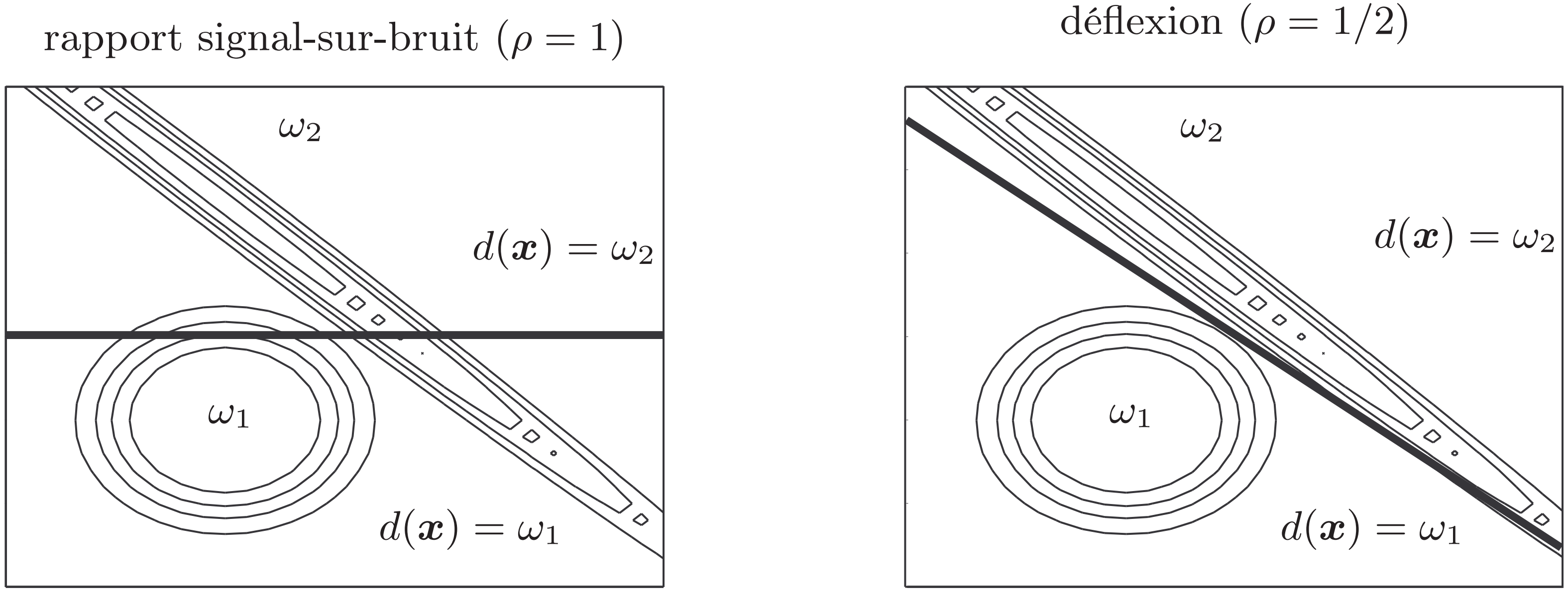
$$J_{\text{rsbg}}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{(\mu_1 - \mu_2)^2}{\rho \, \sigma_1^2 + (1-\rho) \, \sigma_2^2}.$$

Il se décline suivant le critère de Fisher ($\rho = p_1$), la déflexion ($\rho = 1/2$) et le rapport signal-sur-bruit ($\rho = 1$).

Critères du second ordre

Choix optimum du critère

Le choix d'un critère du second ordre constitue un problème crucial car le paramètre ρ influe largement sur les performances :



Critères du second ordre

Choix optimum du critère

Soit $\lambda(\boldsymbol{x})$ la statistique linéaire définie par $\lambda(\boldsymbol{x}) = \boldsymbol{u}^\top \boldsymbol{x} - \lambda_0$. Soit $J(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$ un critère du second ordre, ne dépendant donc que des variables suivantes :

$$\begin{aligned}
\mu_i &= \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} \lambda(\boldsymbol{x}) = \boldsymbol{u}^\top \boldsymbol{m}_i - \lambda_0 \\
\sigma_i^2 &= \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} (\lambda(\boldsymbol{x}) - \mu_i)^2 = \boldsymbol{u}^\top \boldsymbol{\Sigma}_i \, \boldsymbol{u}
\end{aligned}$$

avec $\boldsymbol{m}_i = \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} \boldsymbol{x}$ et $\boldsymbol{\Sigma}_i = \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} (\boldsymbol{x} - \boldsymbol{m}_i)\,(\boldsymbol{x} - \boldsymbol{m}_i)^\top$.

Objectif :

Rechercher $\boldsymbol{u}$ et λ_0 maximisant $J(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ sans expression du critère

Critères du second ordre

Choix optimum du critère

Les dérivés partielles de J par rapport à $\boldsymbol{u}$ et λ_0 doivent être nulles, c'est-à-dire

$$\left\{\begin{array}{l} \frac{\partial J}{\partial \boldsymbol{u}} = \frac{\partial J}{\partial \sigma_1^2} \cdot \frac{\partial \sigma_1^2}{\partial \boldsymbol{u}} + \frac{\partial J}{\partial \sigma_2^2} \cdot \frac{\partial \sigma_2^2}{\partial \boldsymbol{u}} + \frac{\partial J}{\partial \mu_1} \cdot \frac{\partial \mu_1}{\partial \boldsymbol{u}} + \frac{\partial J}{\partial \mu_2} \cdot \frac{\partial \mu_2}{\partial \boldsymbol{u}} = 0 \\ \frac{\partial J}{\partial \lambda_0} = \frac{\partial J}{\partial \sigma_1^2} \cdot \frac{\partial \sigma_1^2}{\partial \lambda_0} + \frac{\partial J}{\partial \sigma_2^2} \cdot \frac{\partial \sigma_2^2}{\partial \lambda_0} + \frac{\partial J}{\partial \mu_1} \cdot \frac{\partial \mu_1}{\partial \lambda_0} + \frac{\partial J}{\partial \mu_2} \cdot \frac{\partial \mu_2}{\partial \lambda_0} = 0, \end{array}\right.$$

où les dérivées partielles de μ_i et σ_i^2 par rapport à $\boldsymbol{u}$ et λ_0 sont données par

$$\frac{\partial \sigma_i^2}{\partial \boldsymbol{u}} = 2\,\boldsymbol{\Sigma}_i\,\boldsymbol{u}, \quad \frac{\partial \mu_i}{\partial \boldsymbol{u}} = \boldsymbol{m}_i, \quad \frac{\partial \sigma_i^2}{\partial \lambda_0} = 0, \quad \frac{\partial \mu_i}{\partial \lambda_0} = -1.$$

Ces résultats permettent de réécrire le système précédent ainsi

$$\left\{\begin{array}{l} 2\left[\frac{\partial J}{\partial \sigma_1^2}\boldsymbol{\Sigma}_1 + \frac{\partial J}{\partial \sigma_2^2}\boldsymbol{\Sigma}_2\right]\boldsymbol{u} = -\left[\frac{\partial J}{\partial \mu_1}\,\boldsymbol{m}_1 + \frac{\partial J}{\partial \mu_2}\,\boldsymbol{m}_2\right] \\ \frac{\partial J}{\partial \mu_1} + \frac{\partial J}{\partial \mu_2} = 0. \end{array}\right.$$

Critères du second ordre

Choix optimum du critère

En introduisant la deuxième équation dans la première, et en notant que $\boldsymbol{u}$ peut être défini à un coefficient multiplicatif près, on aboutit au système linéaire

$$[\rho\boldsymbol{\Sigma}_1 + (1-\rho)\boldsymbol{\Sigma}_2]\,\boldsymbol{u} = (\boldsymbol{m}_2 - \boldsymbol{m}_1),$$

avec

$$\rho = \frac{\frac{\partial J}{\partial \sigma_1^2}}{\frac{\partial J}{\partial \sigma_1^2} + \frac{\partial J}{\partial \sigma_2^2}}.$$

Remarque : On a $0 \leq \rho \leq 1$ si $J(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ varie dans le même sens par rapport à σ_1^2 et σ_2^2 (dérivées de même signe). Cette condition est raisonnable.

Conclusion. Ce résultat est particulièrement intéressant puisqu'il établit que l'expression de $J(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ n'intervient dans $\boldsymbol{u}$ que par l'intermédiaire de ρ.

Critères du second ordre

Choix optimum du critère

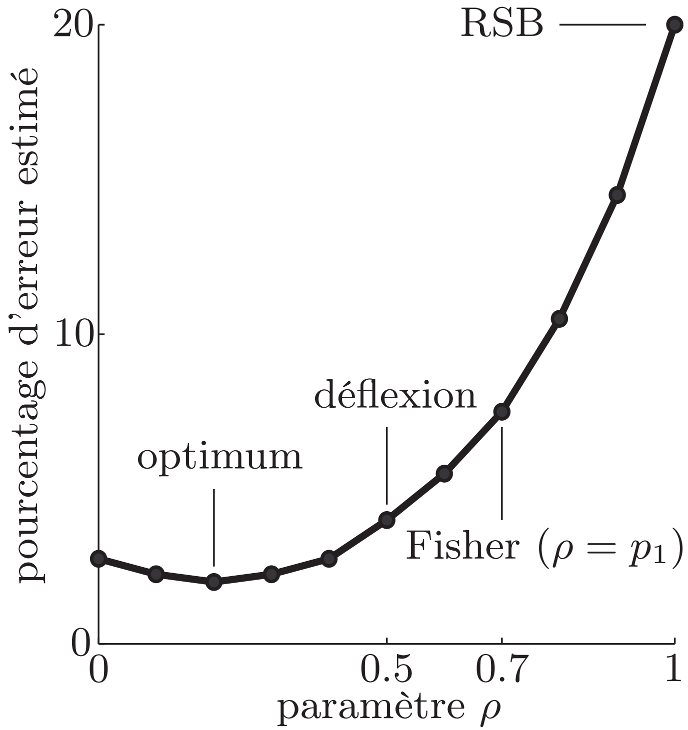
Algorithme de la méthode

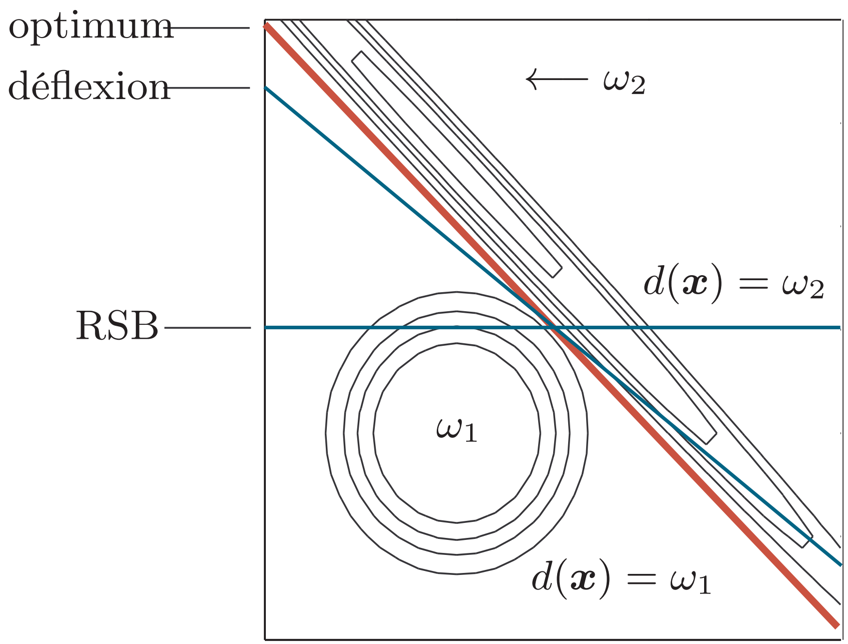
1. Initialiser ρ à 0
2. Tant que $\rho \leq 1$, répéter
   - ▷ résoudre $[\rho\boldsymbol{\Sigma}_1 + (1-\rho)\boldsymbol{\Sigma}_2]\,\boldsymbol{u} = (\boldsymbol{m}_2 - \boldsymbol{m}_1)$ pour obtenir $\boldsymbol{u}_\rho$
   - ▷ déterminer le seuil $\lambda_{\rho,0}$ minimisant par exemple $P_e(d_\rho)$
   - ▷ mise à jour de ρ: $\rho \leftarrow \rho + \Delta\rho$, avec $\Delta\rho$ préalablement choisi
3. Sélectionner le meilleur détecteur d_ρ obtenu

Optimisation des critères du second ordre

Choix optimum du critère

Le classifieur obtenu par la méthode précédente est au moins aussi performant que ceux résultant de la maximisation du rapport signal-sur-bruit, etc. En effet, il leur correspond à chacun une valeur particulière de ρ.





Optimisation des critères du second ordre

Choix optimum du critère

Le calcul du paramètre ρ correspondant au rapport signal-sur-bruit conduit au résultat connu suivant :

$$\boldsymbol{u}_{\mathrm{rsb}} = \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1),$$

car $\rho_{\mathrm{rsb}} = 1$. De façon analogue, on établit pour la déflexion et le critère de Fisher :

$$\begin{aligned}
\boldsymbol{u}_{\mathrm{deflex}} &= 2\,(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1) \\
\boldsymbol{u}_{\mathrm{Fisher}} &= (p_1\,\boldsymbol{\Sigma}_1 + p_2\,\boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1),
\end{aligned}$$

car $\rho_{\mathrm{deflex}} = 1/2$ et $\rho_{\mathrm{Fisher}} = p_1$.

Analyse factorielle discriminante

Cas multi-classe

Généralisation du critère de Fisher à C classes :

On cherche $(C-1)$ projections $\lambda_i(\boldsymbol{x}) = \boldsymbol{u}_i^\top \boldsymbol{x}$ à partir de $(C-1)$ vecteurs $\boldsymbol{u}_i$

On range les vecteurs $\boldsymbol{u}_i$ dans la matrice $\boldsymbol{U}$, soit $\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{U}^\top \boldsymbol{x}$.

Notation :

On généralise la dispersion intra-classe comme ceci

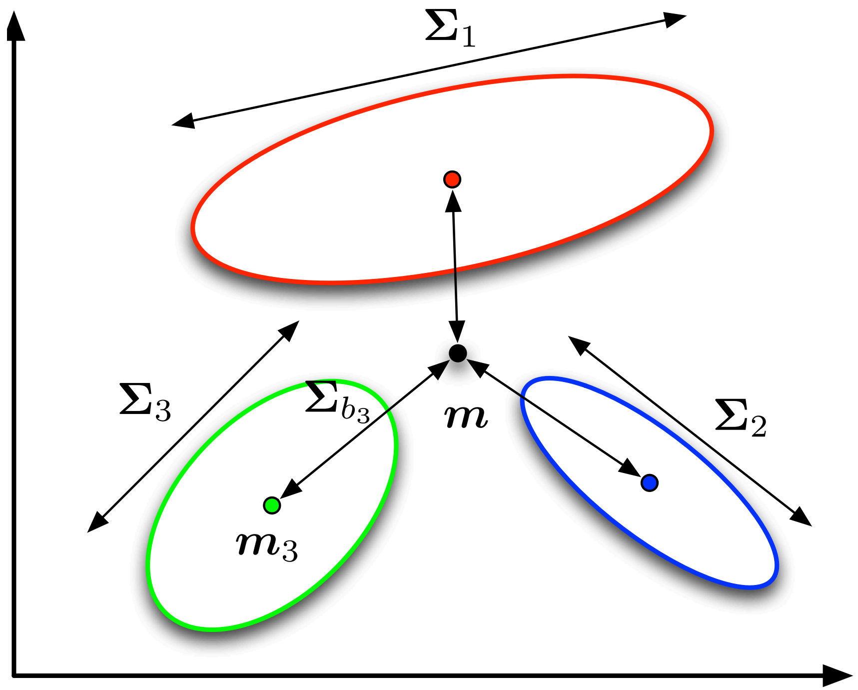
$$\boldsymbol{\Sigma}_w = \sum_{i=1}^{C} \boldsymbol{\Sigma}_i$$

et la dispersion inter-classe comme cela

$$\boldsymbol{\Sigma}_b = \frac{1}{n} \sum_{i=1}^{C} n_i (\boldsymbol{m}_i - \boldsymbol{m}) (\boldsymbol{m}_i - \boldsymbol{m})^\top$$

Analyse factorielle discriminante

Cas multi-classe



Analyse factorielle discriminante

Cas multi-classe

Définition : Dans le cas multi-classe, le critère de Fisher s'exprime ainsi

$$J(\boldsymbol{U}) = \frac{|\boldsymbol{U}^\top \boldsymbol{\Sigma}_b \boldsymbol{U}|}{|\boldsymbol{U}^\top \boldsymbol{\Sigma}_w \boldsymbol{U}|}$$

Maximisation : La solution $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{C-1}]$ du problème est obtenue en résolvant le problème aux valeurs propres généralisé suivant

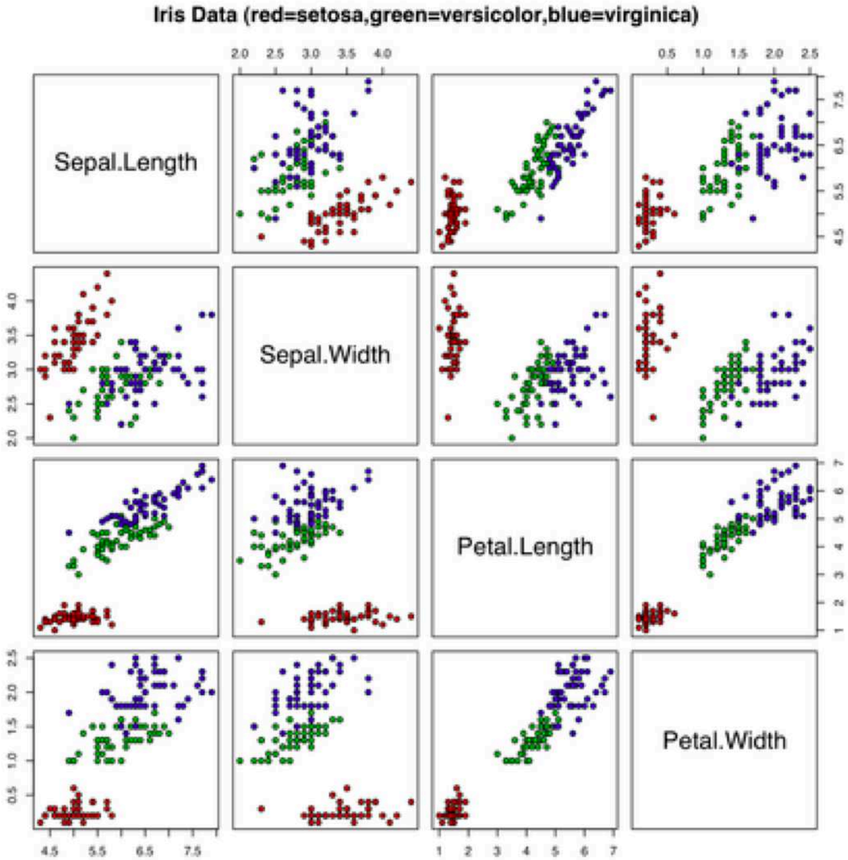
$$\boxed{(\boldsymbol{\Sigma}_b - \lambda_i \boldsymbol{\Sigma}_w)\boldsymbol{u}_i = 0}$$

pour $i = 1, \ldots, C-1$.

Notes : Les directions de projection $\boldsymbol{u}_i$ correspondent aux vecteurs propres associés aux plus grandes valeurs propres de $\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}_b$. Au plus $C-1$ valeurs propres sont non-nulles car $\text{rang}(\boldsymbol{\Sigma}_b) = C-1$.

Analyse factorielle discriminante

Cas multi-classe



Analyse factorielle discriminante

Cas multi-classe