LDA (Linear Discriminant Analysis)

Théorie, intuition, démonstration et exemple 2D $\to$ 1D

1 Idée / intuition (vs PCA)

PCA (rappel en une phrase)

La PCA est **non supervisée** : elle cherche des directions qui **maximisent la variance** des données, sans tenir compte des classes.

LDA

La LDA est **supervisée** : on a des **étiquettes de classes** et on cherche une projection qui :

- **rapproche** les points d'une même classe (faible dispersion intra-classe),
- **éloigne** les classes entre elles (forte séparation inter-classe).

En pratique, LDA construit un sous-espace de dimension au plus $C-1$ si on a C classes.

2 Cadre et notations (dimension générale)

On observe N points $x_1, \dots, x_N \in \mathbb{R}^d$ avec des labels $y_i \in \{1, \dots, C\}$.

Pour chaque classe c :

- N_c = nombre de points dans la classe c,
- μ_c = moyenne de la classe c,
- μ = moyenne globale.

Définitions :

$$\mu_c = \frac{1}{N_c} \sum_{i:\, y_i = c} x_i, \qquad \mu = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

3 Matrices de dispersion : intra-classe et inter-classe

3.1 Dispersion intra-classe (S_W)

$$S_W = \sum_{c=1}^{C} \sum_{i:\, y_i = c} (x_i - \mu_c)(x_i - \mu_c)^\top \in \mathbb{R}^{d \times d}.$$

Intuition. S_W mesure à quel point les points **se dispersent à l'intérieur** de chaque classe.

3.2 Dispersion inter-classe (S_B)

$$S_B = \sum_{c=1}^{C} N_c (\mu_c - \mu)(\mu_c - \mu)^\top \in \mathbb{R}^{d \times d}.$$

Intuition. S_B mesure à quel point les **moyennes de classes** sont éloignées du centre global.

4 Objectif LDA en 1D (une direction)

On cherche une direction $w \in \mathbb{R}^d$ (non nulle) et on projette :

$$z = w^\top x.$$

Après projection :

- la variance inter-classe projetée vaut $w^\top S_B w$,
- la variance intra-classe projetée vaut $w^\top S_W w$.

4.1 Critère de Fisher

$$\boxed{J(w) = \frac{w^\top S_B w}{w^\top S_W w}} \qquad \text{(on veut maximiser } J(w)\text{).}$$

Intuition. On veut un grand numérateur (classes séparées) et un petit dénominateur (classes compactes).

5 Démonstration : solution via problème aux valeurs propres généralisé

Le critère $J(w)$ est invariant par changement d'échelle ($J(\alpha w) = J(w)$). On peut imposer la contrainte :

$$w^\top S_W w = 1.$$

On maximise donc $w^\top S_B w$ sous contrainte $w^\top S_W w = 1$.
Lagrangien :

$$\mathcal{L}(w, \lambda) = w^\top S_B w - \lambda(w^\top S_W w - 1).$$

Condition stationnaire :

$$\nabla_w \mathcal{L} = 2S_B w - 2\lambda S_W w = 0 \quad \Longrightarrow \quad \boxed{S_B w = \lambda S_W w}$$

C'est un **problème de valeurs propres généralisé**. Si S_W est inversible :

$$S_W^{-1} S_B w = \lambda w.$$

Conclusion (1D). La meilleure direction w est le vecteur propre associé à la plus grande valeur propre de $S_W^{-1} S_B$.

6 LDA en dimension k (projection supervisée)

On cherche $W \in \mathbb{R}^{d\times k}$ (colonnes = directions), et on projette $z = W^\top x$.
Objectif (forme trace) :

$$\boxed{\max_{W \neq 0} \operatorname{Tr}\Big((W^\top S_W W)^{-1}(W^\top S_B W)\Big)}$$

La solution consiste à prendre les k vecteurs propres généralisés associés aux plus grandes valeurs propres :

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1, \dots, k.$$

6.1 Dimension maximale

$$\boxed{k \leq C - 1}$$

car $\operatorname{rang}(S_B) \leq C - 1$.

7 Cas particulier : deux classes ($C = 2$)

Quand $C = 2$, on a :

$$S_B = N_1(\mu_1 - \mu)(\mu_1 - \mu)^\top + N_2(\mu_2 - \mu)(\mu_2 - \mu)^\top$$

et il existe une forme simple (à un facteur près) :

$$S_B \propto (\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top.$$

La direction optimale (Fisher) devient :

$$\boxed{w \propto S_W^{-1}(\mu_1 - \mu_2).}$$

Intuition. On prend la différence des moyennes de classes, puis on la corrige par la covariance intra-classe.

8 Exemple concret 2D $\to$ 1D (calculs détaillés)

On prend deux classes en 2D :

8.1 Données

Classe 1 ($N_1 = 3$) :

$$x_1 = (1, 1), \quad x_2 = (2, 1), \quad x_3 = (2, 2).$$

Classe 2 ($N_2 = 3$) :

$$x_4 = (5, 4), \quad x_5 = (6, 5), \quad x_6 = (5, 5).$$

8.2 Moyennes de classes

$$\mu_1 = \frac{1}{3}((1, 1) + (2, 1) + (2, 2)) = \left(\frac{5}{3}, \frac{4}{3}\right)$$

$$\mu_2 = \frac{1}{3}((5, 4) + (6, 5) + (5, 5)) = \left(\frac{16}{3}, \frac{14}{3}\right)$$

Moyenne globale ($N = 6$) :

$$\mu = \frac{N_1\mu_1 + N_2\mu_2}{N} = \frac{\mu_1 + \mu_2}{2} = \left(\frac{7}{2}, 3\right).$$

8.3 Calcul de S_W

Pour la classe 1, on calcule $(x_i - \mu_1)$:

$$x_1 - \mu_1 = \left(-\frac{2}{3}, -\frac{1}{3}\right), \quad x_2 - \mu_1 = \left(\frac{1}{3}, -\frac{1}{3}\right), \quad x_3 - \mu_1 = \left(\frac{1}{3}, \frac{2}{3}\right).$$

Donc

$$S_{W,1} = \sum_{i=1}^{3} (x_i - \mu_1)(x_i - \mu_1)^\top = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Pour la classe 2 :

$$x_4 - \mu_2 = \left(-\frac{1}{3}, -\frac{2}{3}\right), \quad x_5 - \mu_2 = \left(\frac{2}{3}, \frac{1}{3}\right), \quad x_6 - \mu_2 = \left(-\frac{1}{3}, \frac{1}{3}\right).$$

Alors

$$S_{W,2} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Ainsi

$$\boxed{S_W = S_{W,1} + S_{W,2} = \begin{pmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{pmatrix}.}$$

8.4 Calcul de S_B

On a $N_1 = N_2 = 3$ et

$$\mu_1 - \mu = \left(\frac{5}{3} - \frac{7}{2}, \frac{4}{3} - 3\right) = \left(-\frac{11}{6}, -\frac{5}{3}\right),$$

$$\mu_2 - \mu = \left(\frac{16}{3} - \frac{7}{2}, \frac{14}{3} - 3\right) = \left(\frac{11}{6}, \frac{5}{3}\right) = -(\mu_1 - \mu).$$

Donc

$$S_B = 3(\mu_1 - \mu)(\mu_1 - \mu)^\top + 3(\mu_2 - \mu)(\mu_2 - \mu)^\top = 6(\mu_1 - \mu)(\mu_1 - \mu)^\top.$$

Or

$$(\mu_1 - \mu)(\mu_1 - \mu)^\top = \begin{pmatrix} \frac{121}{36} & \frac{55}{18} \\ \frac{55}{18} & \frac{25}{9} \end{pmatrix}.$$

Donc

$$\boxed{S_B = \begin{pmatrix} \frac{121}{6} & \frac{55}{3} \\ \frac{55}{3} & \frac{50}{3} \end{pmatrix}.}$$

8.5 Direction LDA : $w \propto S_W^{-1}(\mu_1 - \mu_2)$

Différence des moyennes :

$$\mu_1 - \mu_2 = \left(\frac{5}{3} - \frac{16}{3}, \frac{4}{3} - \frac{14}{3}\right) = \left(-\frac{11}{3}, -\frac{10}{3}\right).$$

Inverse de S_W :

$$S_W = \begin{pmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{pmatrix}, \qquad \det(S_W) = \frac{16}{9} - \frac{4}{9} = \frac{12}{9} = \frac{4}{3}.$$

$$S_W^{-1} = \frac{1}{\det(S_W)} \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix} = \frac{3}{4} \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}.$$

Donc

$$w \propto S_W^{-1}(\mu_1 - \mu_2) = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} -\frac{11}{3} \\ -\frac{10}{3} \end{pmatrix} = \begin{pmatrix} -\frac{11}{3} + \frac{5}{3} \\ \frac{11}{6} - \frac{10}{3} \end{pmatrix} = \begin{pmatrix} -2 \\ -\frac{3}{2} \end{pmatrix}.$$

On peut enlever le signe (même séparation) :

$$\boxed{w \propto (2,\ 1.5).}$$

Normalisation (optionnelle) :

$$\|w\| = \sqrt{2^2 + 1.5^2} = \sqrt{4 + 2.25} = \sqrt{6.25} = 2.5, \quad \Rightarrow \quad \boxed{\hat{w} = (0.8,\ 0.6)\,.}$$

8.6 Projection 1D et séparation

Score 1D :

$$z = \hat{w}^\top x = 0.8x_1 + 0.6x_2.$$

Classe 1 :

$$z(x_1) = 0.8 \cdot 1 + 0.6 \cdot 1 = 1.4, \quad z(x_2) = 0.8 \cdot 2 + 0.6 \cdot 1 = 2.2, \quad z(x_3) = 0.8 \cdot 2 + 0.6 \cdot 2 = 2.8.$$

Classe 2 :

$$z(x_4) = 0.8 \cdot 5 + 0.6 \cdot 4 = 6.4, \quad z(x_5) = 0.8 \cdot 6 + 0.6 \cdot 5 = 7.8, \quad z(x_6) = 0.8 \cdot 5 + 0.6 \cdot 5 = 7.0.$$

On voit une séparation nette entre les deux classes sur l'axe LDA.

9 Lien probabiliste (résumé)

Sous certaines hypothèses classiques, la LDA correspond à un modèle génératif :

- $x|y = c \sim \mathcal{N}(\mu_c, \Sigma)$ (même covariance pour toutes les classes),
- $P(y = c) = \pi_c$.

Alors la frontière de décision est **linéaire** et la règle optimale (Bayes) mène à la LDA.

10 Synthèse

- **PCA** : non supervisée, maximise la variance totale.
- **LDA** : supervisée, maximise la séparation *inter-classe* tout en minimisant la dispersion *intra-classe*.

Formule clé (Fisher) :

$$\boxed{\max_w \frac{w^\top S_B w}{w^\top S_W w} \quad \Rightarrow \quad S_B w = \lambda S_W w.}$$