Chapitre 5 : Réduction de dimension

24 février 2026

Plan du chapitre

Motivations

Sélection vs Extraction

Mathématiques de l'ACP

Pourquoi réduire la dimension ?

Les données en grande dimension posent de nombreux problèmes. Quatre motivations principales incitent à réduire cette dimension :

1. **La malédiction de la dimension** (le fléau).
2. **Le compromis Biais-Variance :** Un modèle plus simple conduit généralement à une variance plus faible.
3. **L'explicabilité :** Le besoin de justifier simplement le résultat produit par un algorithme.
4. **La visualisation**.

Comment réduire la dimension ?

Deux grandes approches existent : la sélection et l'extraction.

A. Sélection de variables

- **Filtrage :** Évaluation univariée (corrélation, information mutuelle). Rapide mais ignore les interactions.
- **Conteneur (Wrapper) :** Recherche heuristique de sous-ensembles (Ascendante/Descendante). Teste la performance globale.
- **Méthodes embarquées (Embedded) :** Apprentissage conjoint du modèle et des variables utiles (ex : pénalisation LASSO).

Extraction de variables et ACP

B. Extraction de variables

Créer de **nouvelles** variables (combinaisons des anciennes) qui capturent l'information pertinente.

Analyse en Composantes Principales (ACP)

Cherche une transformation linéaire $\mathbf{x} \in \mathbb{R}^d \to z \in \mathbb{R}^p$ $(p \ll d)$ qui maximise la variance conservée.

- $z_i = \langle v_i, \mathbf{x} \rangle$
- Les z_i doivent être décorrélées.
- $\text{Var}(z_1) \geq \text{Var}(z_2) \geq \cdots \geq \text{Var}(z_p)$

ACP : Première composante principale

Trouver le vecteur v_1 qui maximise $\text{Var}(z_1)$ sous la contrainte $\|v_1\|^2 = 1$.

$$\text{Var}(z_1) = \text{Var}(\langle v_1, \mathbf{x} \rangle) = \mathbb{E}(v_1^T \mathbf{x}\mathbf{x}^T v_1) = v_1^T \Gamma v_1$$

où Γ est la matrice de covariance.

Formulons le Lagrangien :

$$L(v_1, \lambda_1) = v_1^T \Gamma v_1 + \lambda_1 (1 - v_1^T v_1)$$

En annulant le gradient par rapport à v_1 :

$$\nabla_{v_1} L = 2\Gamma v_1 - 2\lambda_1 v_1 = 0 \implies \Gamma v_1 = \lambda_1 v_1$$

Résultat : v_1 est le vecteur propre associé à la plus grande valeur propre λ_1 de Γ. De plus, $\text{Var}(z_1) = \lambda_1$.

ACP : Deuxième composante principale

Trouver v_2 maximisant Var(z_2) avec les contraintes $\|v_2\|^2 = 1$ et $\langle v_1, v_2 \rangle = 0$ (décorrélation).

$$L(v_2, \lambda_2) = v_2^T \Gamma v_2 + \lambda_2(1 - v_2^T v_2) + \beta(0 - v_2^T v_1)$$

$$\nabla_{v_2} L = 2\Gamma v_2 - 2\lambda_2 v_2 - \beta v_1 = 0 \implies \Gamma v_2 = \lambda_2 v_2 + \frac{\beta}{2} v_1$$

En multipliant à gauche par v_1^T :

$$v_1^T \Gamma v_2 = \lambda_2 v_1^T v_2 + \frac{\beta}{2} v_1^T v_1 = \frac{\beta}{2}$$

Or, $v_1^T \Gamma v_2 = (\Gamma v_1)^T v_2 = \lambda_1 v_1^T v_2 = 0$. Donc $\beta = 0$.
Résultat : $\Gamma v_2 = \lambda_2 v_2$. v_2 est le 2ème plus grand vecteur propre.

Remarques pratiques sur l'ACP

L'implémentation de l'ACP nécessite quelques précautions en pratique :

- **Standardisation :** Les variables doivent être mises à la même échelle (centrées-réduites), sinon la variable avec la plus grande unité dictera artificiellement la variance.
- **Estimation :** On utilise la matrice de covariance *empirique*.
- **Choix de p :** On choisit souvent p en regardant la proportion de variance expliquée cumulée :

$$\frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{d} \lambda_k} \geq \text{Seuil (ex : 90\%)}$$