Analyse en Composantes Principales (ACP / PCA)

Théorie générale, intuition et démonstrations

1 Introduction

L'Analyse en Composantes Principales (ACP, ou PCA pour *Principal Component Analysis*) est une méthode de réduction de dimension linéaire visant à représenter des données de grande dimension dans un espace de dimension plus faible, tout en conservant un maximum d'information.

Cette information est mesurée par la **variance** des données. PCA cherche donc les directions dans lesquelles les données varient le plus.

2 Cadre mathématique général

2.1 Données

On considère :

- N observations
- chaque observation $x_i \in \mathbb{R}^d$

La matrice des données est :

$$X = \begin{pmatrix} x_1^\top \\ \vdots \\ x_N^\top \end{pmatrix} \in \mathbb{R}^{N \times d}$$

3 Centrage des données

3.1 Définition

La moyenne empirique est :

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Les données centrées sont :

$$\tilde{x}_i = x_i - \mu$$

Sous forme matricielle :

$$\tilde{X} = X - \mathbf{1}\mu^\top$$

3.2 Intuition

Le centrage permet d'étudier la **variabilité autour du barycentre**. Sans centrage, la première composante pourrait simplement capter la moyenne.

4 Matrice de covariance

4.1 Définition

La matrice de covariance empirique est définie par :

$$\Sigma = \frac{1}{N}\tilde{X}^\top \tilde{X} \in \mathbb{R}^{d\times d}$$

4.2 Propriétés

- Σ est symétrique
- Σ est semi-définie positive
- Σ est diagonalisable dans une base orthonormée

5 Problème fondamental de la PCA

5.1 Formulation

On cherche un sous-espace vectoriel $E_k \subset \mathbb{R}^d$ de dimension k maximisant la variance projetée.

Soit $U \in \mathbb{R}^{d\times k}$ telle que :

$$U^\top U = I_k$$

Le problème PCA est :

$$\boxed{\max_{U^\top U = I_k} \operatorname{Tr}(U^\top \Sigma U)}$$

5.2 Intuition

- projeter les données sur E_k
- conserver le maximum d'étalement
- perdre le minimum d'information

6 Cas une dimension

On cherche un vecteur unitaire $u \in \mathbb{R}^d$ tel que :

$$\max_{\|u\|=1} u^\top \Sigma u$$

6.1 Lagrangien

$$\mathcal{L}(u,\lambda) = u^\top \Sigma u - \lambda(u^\top u - 1)$$

Condition stationnaire :

$$\nabla_u \mathcal{L} = 0 \quad \Rightarrow \quad \Sigma u = \lambda u$$

6.2 Conclusion

$$\boxed{\text{Les directions PCA sont les vecteurs propres de } \Sigma}$$

La variance expliquée est la valeur propre associée.

7 Résolution générale (dimension k)

7.1 Décomposition spectrale

Comme Σ est symétrique :

$$\Sigma = Q\Lambda Q^\top$$

avec :

$$\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_d), \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$$

7.2 Solution optimale

$$\boxed{U^\star = \begin{pmatrix} q_1 & \cdots & q_k \end{pmatrix}}$$

où q_i est le vecteur propre associé à λ_i.

7.3 Variance expliquée

$$\text{Variance expliquée} = \sum_{i=1}^{k} \lambda_i$$

$$\text{Proportion expliquée} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{d} \lambda_j}$$

8 Projection et reconstruction

8.1 Projection

$$Z = \tilde{X}U$$

8.2 Reconstruction

$$\hat{X} = ZU^\top + \mu$$

8.3 Erreur minimale

$$\boxed{\min \|X - \hat{X}\|_F^2 = \sum_{i=k+1}^{d} \lambda_i}$$

8.4 Interprétation

La PCA fournit la meilleure approximation linéaire de rang k au sens des moindres carrés.

9 Lien avec la SVD

Décomposition en valeurs singulières :

$$X = WSV^\top$$

Alors :

$$\Sigma = \frac{1}{N} VS^2V^\top$$

Les axes PCA sont les colonnes de V.

9.1 Théorème d'Eckart–Young

$$X_k = \sum_{i=1}^{k} s_i w_i v_i^\top$$

est la meilleure approximation de rang k de X.

10 Conclusion

$$\boxed{\text{PCA} = \text{diagonaliser la covariance et projeter sur ses plus grandes directions propres}}$$

Elle permet :

- réduction de dimension
- visualisation
- compression
- débruitage

11 Données d'exemple (quelconques)

On prend $n = 5$ points en $\mathbb{R}^2$:

$$x_1 = (1,1), \quad x_2 = (2,2), \quad x_3 = (3,3), \quad x_4 = (4,4), \quad x_5 = (5,7).$$

On forme la matrice des données (lignes = observations) :

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 4 \\ 5 & 7 \end{pmatrix} \in \mathbb{R}^{5\times 2}.$$

12 Idée / intuition de la PCA

On cherche une droite (un axe) passant par le centre des données telle que :

[itemsep=3pt]

- si on projette tous les points sur cette droite, les projections sont **le plus étalées possible** (variance maximale) ;
- équivalemment, si on reconstruit les points à partir de cette projection (on revient sur la droite), l'**erreur moyenne** est **minimale**.

PCA choisit donc la direction qui capture le plus de variance.

13 Étape 1 — Centrer les données

13.1 Moyenne

$$\mu = \frac{1}{5}\sum_{i=1}^{5} x_i.$$

Coordonnée par coordonnée :

$$\mu_x = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3, \qquad \mu_y = \frac{1+2+3+4+7}{5} = \frac{17}{5} = 3.4.$$

Donc

$$\mu = (3,\ 3.4).$$

13.2 Données centrées

$$\tilde{x}_i = x_i - \mu.$$

On obtient :

$$\tilde{x}_1 = (1-3,\ 1-3.4) = (-2,\ -2.4)$$

$$\tilde{x}_2 = (-1,\ -1.4), \quad \tilde{x}_3 = (0,\ -0.4), \quad \tilde{x}_4 = (1,\ 0.6), \quad \tilde{x}_5 = (2,\ 3.6).$$

La matrice centrée $\tilde{X}$ est donc :

$$\tilde{X} = \begin{pmatrix} -2 & -2.4 \\ -1 & -1.4 \\ 0 & -0.4 \\ 1 & 0.6 \\ 2 & 3.6 \end{pmatrix}.$$

Intuition. La PCA doit décrire la **variation autour du centre**, pas la position absolue.

14 Étape 2 — Covariance (la "forme" du nuage)

On utilise la covariance empirique (version $\frac{1}{n}$) :

$$\Sigma = \frac{1}{n}\tilde{X}^\top \tilde{X}.$$

En 2D :

$$\Sigma = \begin{pmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(x,y) & \mathrm{Var}(y) \end{pmatrix} = \frac{1}{5}\sum_{i=1}^{5} \begin{pmatrix} \tilde{x}_{i1}^2 & \tilde{x}_{i1}\tilde{x}_{i2} \\ \tilde{x}_{i1}\tilde{x}_{i2} & \tilde{x}_{i2}^2 \end{pmatrix}.$$

Calcul des sommes :

14.1 $S_{xx} = \sum \tilde{x}^2$

$$(-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 = 4 + 1 + 0 + 1 + 4 = 10.$$

14.2 $S_{yy} = \sum \tilde{y}^2$

$$(-2.4)^2 + (-1.4)^2 + (-0.4)^2 + (0.6)^2 + (3.6)^2$$

$$= 5.76 + 1.96 + 0.16 + 0.36 + 12.96 = 21.2.$$

14.3 $S_{xy} = \sum \tilde{x}\tilde{y}$

$$(-2)(-2.4) + (-1)(-1.4) + 0(-0.4) + 1(0.6) + 2(3.6)$$
$$= 4.8 + 1.4 + 0 + 0.6 + 7.2 = 14.$$

Donc :

$$\Sigma = \frac{1}{5}\begin{pmatrix} 10 & 14 \\ 14 & 21.2 \end{pmatrix} = \begin{pmatrix} 2 & 2.8 \\ 2.8 & 4.24 \end{pmatrix}.$$

Intuition. Les diagonales mesurent les **variances** sur chaque axe, tandis que le terme hors-diagonale 2.8 mesure à quel point x et y varient ensemble (nuage incliné).

15 Étape 3 — Démonstration du critère PCA (variance maximale)

15.1 Projection sur une direction unitaire u

On prend une direction $u \in \mathbb{R}^2$ telle que $\|u\| = 1$. La projection d'un point centré $\tilde{x}_i$ sur u est :

$$z_i = u^\top \tilde{x}_i.$$

La variance des z_i (avec $\frac{1}{n}$) est :

$$\mathrm{Var}(z) = \frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\sum_{i=1}^{n}(u^\top \tilde{x}_i)^2.$$

Or

$$(u^\top \tilde{x}_i)^2 = u^\top(\tilde{x}_i\tilde{x}_i^\top)u.$$

Donc :

$$\mathrm{Var}(z) = u^\top\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}_i^\top\right)u = u^\top\Sigma u.$$

15.2 Problème PCA (1 composante)

$$\max_{\|u\|=1} \; u^\top \Sigma u.$$

15.3 Lagrange ⇒ équation aux vecteurs propres

Lagrangien :

$$\mathcal{L}(u,\lambda) = u^\top\Sigma u - \lambda(u^\top u - 1).$$

Condition d'optimalité :

$$\nabla_u \mathcal{L} = 2\Sigma u - 2\lambda u = 0 \quad \Rightarrow \quad \Sigma u = \lambda u.$$

Conclusion. Les directions candidates sont les **vecteurs propres** de Σ et la variance projetée vaut exactement la valeur propre λ. La meilleure direction est celle associée au plus grand λ.

16 Étape 4 — Calcul des valeurs propres (complet)

On résout :

$$\det(\Sigma - \lambda I) = 0.$$

On a :

$$\Sigma - \lambda I = \begin{pmatrix} 2-\lambda & 2.8 \\ 2.8 & 4.24-\lambda \end{pmatrix}.$$

Déterminant :

$$(2-\lambda)(4.24-\lambda) - 2.8^2 = 0 \quad \Leftrightarrow \quad (2-\lambda)(4.24-\lambda) - 7.84 = 0.$$

Développons :

$$(2-\lambda)(4.24-\lambda) = 8.48 - 6.24\lambda + \lambda^2.$$

Donc :

$$\lambda^2 - 6.24\lambda + (8.48 - 7.84) = 0 \quad \Rightarrow \quad \lambda^2 - 6.24\lambda + 0.64 = 0.$$

Discriminant :

$$\Delta = 6.24^2 - 4 \cdot 0.64 = 38.9376 - 2.56 = 36.3776, \qquad \sqrt{\Delta} \approx 6.031.$$

Valeurs propres :

$$\lambda_{1,2} = \frac{6.24 \pm 6.031}{2}.$$

Ainsi :

$$\lambda_1 \approx \frac{12.271}{2} = 6.1355, \qquad \lambda_2 \approx \frac{0.209}{2} = 0.1045.$$

Interprétation. La variance totale vaut $\lambda_1 + \lambda_2 \approx 6.24$. La composante 1 explique :

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} \approx \frac{6.1355}{6.24} \approx 98.3\%.$$

17 Étape 5 — Vecteur propre principal u_1

On résout :

$$(\Sigma - \lambda_1 I)u = 0.$$

Avec $\lambda_1 = 6.1355$:

$$\Sigma - \lambda_1 I = \begin{pmatrix} 2 - 6.1355 & 2.8 \\ 2.8 & 4.24 - 6.1355 \end{pmatrix} = \begin{pmatrix} -4.1355 & 2.8 \\ 2.8 & -1.8955 \end{pmatrix}.$$

Première ligne :

$$-4.1355\,x + 2.8\,y = 0 \quad \Rightarrow \quad y = \frac{4.1355}{2.8}x \approx 1.47696\,x.$$

On prend donc un vecteur proportionnel :

$$u_1 \propto (1,\ 1.47696).$$

Normalisation :

$$\|u_1\| = \sqrt{1^2 + 1.47696^2} = \sqrt{1 + 2.1804} = \sqrt{3.1804} \approx 1.7834.$$

Donc :

$$u_1 \approx \left(\frac{1}{1.7834}, \frac{1.47696}{1.7834}\right) \approx (0.5607,\ 0.8280).$$

Le second axe PCA est orthogonal :

$$u_2 \approx (-0.8280,\ 0.5607).$$

18 Étape 6 — Projections (réduction 2D → 1D)

Pour chaque point centré $\tilde{x}_i$, la coordonnée PCA (1D) est :

$$z_i = u_1^\top \tilde{x}_i.$$

Calculs :

$$z_1 = (0.5607)(-2) + (0.8280)(-2.4) = -1.1214 - 1.9872 = -3.1086.$$
$$z_2 = (0.5607)(-1) + (0.8280)(-1.4) = -0.5607 - 1.1592 = -1.7199.$$
$$z_3 = (0.5607)(0) + (0.8280)(-0.4) = -0.3312.$$
$$z_4 = (0.5607)(1) + (0.8280)(0.6) = 0.5607 + 0.4968 = 1.0575.$$
$$z_5 = (0.5607)(2) + (0.8280)(3.6) = 1.1214 + 2.9808 = 4.1022.$$

Les z_i sont les données compressées en 1D.

19 Étape 7 — Reconstruction et erreur (intuition "perte minimale")

Reconstruction sur la droite PCA :

$$\hat{x}_i = \mu + z_i u_1.$$

19.1 Exemples

Pour x_1 :

$$\hat{x}_1 = (3, 3.4) + (-3.1086)(0.5607, 0.8280).$$
$$(-3.1086)(0.5607, 0.8280) \approx (-1.743,\ -2.572).$$
$$\hat{x}_1 \approx (1.257,\ 0.828).$$

Pour x_5 :

$$\hat{x}_5 = (3, 3.4) + (4.1022)(0.5607, 0.8280).$$
$$(4.1022)(0.5607, 0.8280) \approx (2.300,\ 3.397).$$
$$\hat{x}_5 \approx (5.300,\ 6.797).$$

Intuition. On remplace chaque point par son ombre sur la meilleure droite possible.

20 Lien direct avec la "variance perdue"

La coordonnée ignorée (sur u_2) est :

$$w_i = u_2^\top \tilde{x}_i,$$

et l'erreur de reconstruction (distance orthogonale à la droite) est $|w_i|$.

Calcul rapide (vérification) :

$$w_1 = (-0.828)(-2) + (0.5607)(-2.4) = 1.656 - 1.3457 = 0.3103.$$
$$w_4 = (-0.828)(1) + (0.5607)(0.6) = -0.828 + 0.3364 = -0.4916.$$

Si on fait la moyenne de w_i^2, on obtient environ 0.104, ce qui correspond à λ_2. Donc :

[itemsep=3pt]

- variance conservée $\approx \lambda_1$,
- variance perdue / erreur moyenne quadratique $\approx \lambda_2$.

21 Synthèse

[itemsep=3pt]

1. Centrer les données.
2. Calculer $\Sigma = \frac{1}{n}\tilde{X}^\top \tilde{X}$.
3. Trouver valeurs propres / vecteurs propres.
4. Prendre les k plus grandes valeurs propres $\Rightarrow$ directions PCA.
5. Projeter : $Z = \tilde{X}U$.
6. Reconstruire : $\hat{X} = ZU^\top + \mu$.