pca

Table des matières

1 Introduction à l'analyse en composantes principales (PCA)

1.1 Qu'est-ce que la PCA ?

L'analyse en composantes principales (PCA) est une méthode statistique qui permet de transformer des données en un ensemble de valeurs de variables non corrélées appelées composantes principales. Elle est utilisée pour réduire la dimensionnalité des données tout en préservant autant que possible la variance présente dans les données d'origine.

Intuition

La PCA peut être vue comme une méthode pour identifier les directions principales dans lesquelles les données varient le plus. Cela permet de simplifier les données tout en conservant leur structure essentielle.

1.2 Historique

La PCA a été inventée en 1901 par Karl Pearson comme une analogie du théorème de l'axe principal en mécanique. Elle a été développée indépendamment et nommée par Harold Hotelling dans les années 1930. Selon le domaine d'application, elle est également appelée transformation discrète de Karhunen-Loève, transformation de Hotelling, décomposition orthogonale propre, décomposition en valeurs singulières, et bien d'autres.

Fiche Récapitulative

Objectif : Résumer le concept principal de la PCA.
Principe central : Réduire la dimensionnalité des données tout en préservant la variance.

Points essentiels : - Réduction de dimension - Préservation de la variance - Transformation des données

Formules clés :
- $X = U\Sigma V^T$
- $C = U\Lambda U^T$

1.3 Perspective géométrique

L'objectif est de trouver les directions principales intrinsèques des données. Cela implique de trouver une base orthonormale le long des directions principales.

Intuition

Iterativement, on cherche la direction principale de "dispersion" des données.

1.4 Perspective algorithmique

L'algorithme commun de la PCA commence par des données $X \in R^D$ prises d'une distribution avec des statistiques d'ordre connues. Les directions principales sont les vecteurs propres triés de la matrice de covariance.

$$E(X) = \mu \tag{1}$$
$$E((X-\mu)(X-\mu)^T) = C \tag{2}$$

À retenir

Les directions principales sont obtenues par la décomposition spectrale de la matrice de covariance.

2 Transformation PCA

Réécrire les données dans la nouvelle base PCA implique un changement de base :

$$Y = U^T(X-\mu) \tag{3}$$

3 PCA et réduction de dimensionnalité

La PCA est une approximation à k termes dans la base PCA. Le choix de d peut être basé sur la couverture de la variance ou une sélection fixe pour les visualisations.

$$Y_d = U_d^T(X-\mu) \tag{4}$$
$$d = \min_k \left(\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{D} \lambda_i} \geq \alpha \right) \tag{5}$$

Vulgarisation simple

La PCA permet de simplifier les données en réduisant le nombre de dimensions tout en conservant l'essentiel de l'information.

4 Exercices

4.1 Exercice 1

Considérons un signal aléatoire $\varphi \in R^3$ avec une matrice d'autocorrélation :

$$R_\varphi = \begin{pmatrix} 1-\alpha & -\frac{\alpha}{2} & -\frac{\alpha}{2} \\ -\frac{\alpha}{2} & 1-\alpha & -\frac{\alpha}{2} \\ -\frac{\alpha}{2} & -\frac{\alpha}{2} & 1-\alpha \end{pmatrix} \tag{6}$$

4.2 Exercice 2

Considérons un signal aléatoire $\varphi \in R^3$ avec une matrice d'autocorrélation :

$$R_\varphi = \begin{pmatrix} 1-\alpha & -\alpha & 0 \\ -\alpha & 1 & -\alpha \\ 0 & -\alpha & 1-\alpha \end{pmatrix} \tag{7}$$

À retenir

Les matrices d'autocorrélation peuvent être utilisées pour déterminer la décomposition PCA d'un signal.