Principal Component Analysis

Grant McNaughton

23 February 2024

1 Problem

Informally put, the problem is to find a "natural" coordinate system for a given collection of points. The following makes it precise.

Problem 1.

In: $X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{R}^{m \times d}$ *where x_i is a point in the d-dim space*

Out: $c = [c_1 \cdots c_d] \in \mathbb{R}^d$ *where c_j is the mean of the j-th coordinate of the m points*

$B = [b_1 \cdots b_d] \in \mathbb{R}^{d \times d}$ *where* $b_k = \arg \max\limits_{\substack{u \in \mathbb{R}^d \\ ||u||=1 \\ b_1 \cdot u = \cdots = b_{k-1} \cdot u = 0}} \sum\limits_{1 \le i \le m} (x_{ci} \cdot u)^2$

where again $\begin{bmatrix} x_{c1} \\ \vdots \\ x_{cm} \end{bmatrix} = \begin{bmatrix} x_1 - c \\ \vdots \\ x_m - c \end{bmatrix}$

Remark 2.

1. *c is the center of the "natural" coordinate system.*
2. *B is the orthonormal basis defining the "natural" coordinate system.*

2 Theory

Notation 3. *Let* $Xc = \begin{bmatrix} x_{c1} \\ \vdots \\ x_{cm} \end{bmatrix}$.

Theorem 4. *Let $S = X_c^t X_c \in \mathbb{R}^{d \times d}$. Let $V = [v_1 \cdots v_d]$, $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} = ED(S)$. Assume that $\lambda_1 \ge \cdots \ge \lambda_d$ and that $||v_1|| = \cdots = ||v_d|| = 1$. Then $B = V$.*

Proof. We will prove by induction that $B = V$. First, we will prove that $B_1, B_2 = v_1, v_2$.

Consider the following set of equations:

$$\begin{aligned}\sum_{i=1}^{m}(x_{ci} \cdot u)^2 &= ||X_c u|| \\ &= (X_c u)^\intercal (X_c u) \\ &= u^\intercal X_c^\intercal X_c u \\ &= u^\intercal S u\end{aligned}$$

Because V is orthonormal, u can be expressed as $a_1 v_1 + \cdots + a_d v_d$ for some set of constants $a_1, \ldots, a_d$. We can continue our computation as follows:

$$\begin{aligned}\sum_{i=1}^{m}(x_{ci} \cdot u)^2 &= u^\intercal S u \\ &= u^\intercal S(a_1 v_1 + \cdots + a_d v_d) \\ &= u^\intercal (a_1 \lambda_1 v_1 + \cdots + a_d \lambda_d v_d) \\ &= (a_1 v_1^\intercal + \cdots + a_d v_d^\intercal)(a_1 \lambda_1 v_1 + \cdots + a_d \lambda_d v_d)\end{aligned}$$

Because all vectors v_i are orthogonal, $v_i^\intercal v_i = 1$ and $v_i^\intercal v_j = 0$ for all $i, j \in [1, n], i \neq j$. Therefore,

$$\begin{aligned}\sum_{i=1}^{m}(x_{ci} \cdot u)^2 &= (a_1 v_1^\intercal + \cdots + a_d v_d^\intercal)(a_1 \lambda_1 v_1 + \cdots + a_d \lambda_d v_d) \\ &= a_1^2 \lambda_1 + \cdots + a_i^2 \lambda_i\end{aligned}$$

From this, we can rewrite our problem as

$$\operatorname*{argmax}_{||u=1||}\Big(\sum_{i=1}^{m}(x_{ci} \cdot u)^2\Big) = \operatorname*{argmax}_{\sum_{i=1}^{d} a_i^2 = 1} \, (a_1^2 \lambda_1 + \cdots + a_d^2 \lambda_d)$$

Without loss of generality, assume the eigendecomposition of S is sorted such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. When this is the case, $a_1 = 1, a_n = 0$ for all $n \in [2, d]$ and $a_1^2 \lambda_1 + \cdots + a_d^2 \lambda_d = \lambda_1$. Thus $b_1 = \operatorname*{argmax}_{||u=1||}(x_{ci} \cdot u)^2 = u = v_1$.

Assume the induction hypothesis $b_n = v_n$.

Then b_{n+1} is given by $\operatorname*{argmax}_{\substack{||u=1|| \\ b_1 \cdot u = \cdots = b_n \cdot u = 0}} (x_{ci} \cdot u)^2 = \operatorname*{argmax}_{\substack{\sum_{i=1}^{d} a_i^2 = 1 \\ b_1 \cdot a = \cdots = b_n \cdot a = 0}} (a_1^2 \lambda_1 + \cdots + a_d^2 \lambda_d)$.

Because setting a_x as any nonzero value for any $x \in [1, n]$will make $b_x \cdot a = 0a_1 + \cdots + 0a_{x-1} + 1a_x + 0a_{x+1} + \cdots + 0a_d = a_x \neq 0$, we know that $a_x = 0$ and our new problem can be expressed as $\operatorname*{argmax}_{\substack{\sum_{i=1}^{d} a_i^2 = 1 \\ b_1 \cdot a = \cdots = b_n \cdot a = 0}} (a_{n+1}^2 \lambda_{n+1} + \cdots + a_d^2 \lambda_d)$.

Once again, this is maximized when $a_{n+1} = 1, a_x = 0$ for all $x \in [n+2, d]$ and $a_{n+1}^2 \lambda_{n+1} + \cdots + a_d^2 \lambda_d = \lambda_{n+1}$. Thus $b_{n+1} = \operatorname*{argmax}_{\substack{||u=1|| \\ b_1 \cdot u = \cdots = b_n \cdot u = 0}} (x_{ci} \cdot u)^2 = v_{n+1}$.

By the principle of mathematical induction, we can say that $b_n = v_n$ for all $n \in [1, d]$. $\square$

Corollary 5. *Let $U, \Sigma, V = SVD\,(X_c)$. Then $B = V$.*

Proof. Immediate from the fact that the normalized eigenvector matrix of $S = X_c^t X_c$ is the same as the right singular vector matrix of X_c. □

Algorithm 6. $c, V \leftarrow PCA\,(X)$

1. $c \leftarrow$ *column mean of* X
2. $X_c \leftarrow \begin{bmatrix} x_1 - c \\ \vdots \\ x_m - c \end{bmatrix}$
3. $U, \Sigma, V \leftarrow SVD\,(X_c)$
4. *return* c, V.

Theorem 7. *Let* $X' = \begin{bmatrix} x'_1 \\ \vdots \\ x'_d \end{bmatrix} \in \mathbb{R}^{m \times d}$ *be such that* $x_i - c = x'_{i1}v_1^t + \cdots + x'_{id}v_d^t$. *Then* $X' = X_c V$.

Proof. Note that because V is an orthogonal matrix, $V^{\intercal} = V^{-1}$ and $v_i v_i^{\intercal} = 1, v_i v_j^{\intercal} = 0$ for all $i, j \in \mathbb{N}, i \neq j$. Additionally, we have already defined $X_c = X - c$, so we can say that $x_i - c = x_{ci} = x'_{i1}v_1^{\intercal} + \cdots + x'_{id}v_d^{\intercal}$.

$x_{ci} \in \mathbb{R}^{1 \times d}$ and $V \in \mathbb{R}^{d \times d}$, so $x_{ci}V = (x'_{i1}v_1^{\intercal} + \cdots + x'_{id}v_d^{\intercal})V \in \mathbb{R}^{1 \times d}$. From this, we can show the following:

$$\begin{aligned} x_{ci}V &= (x'_{i1}v_1^{\intercal} + \cdots + x'_{id}v_d^{\intercal})V \\ &= x'_{i1}v_1^{\intercal}v_1 + x'_{i1}v_1^{\intercal}v_2 + \cdots + x'_{i1}v_1^{\intercal}v_d + x'_{i2}v_2^{\intercal}v_1 + \cdots + x'_{id}v_d^{\intercal}v_1 + \cdots + x'_{id}v_d^{\intercal}v_d \\ &= x'_{i1} + x'_{i2} + \cdots + x'_{id} \\ &= x'_i \end{aligned}$$

We can then generalize $x_{ci}V = x'_i$ to say that $X_cV = X'$. □