Principal Component Analysis

Keng-Yu Chen

June 9, 2025

Principal Component Analysis (PCA) can be viewed as a linear transform which maps a n-dimensional random vector to a smaller dimensional vector, but preserving most of its information.

In the following we consider a random vector $x = (x_1, x_2, \dots,\ x_n)^T$, where each x_i is a random variable, representing one attribute or property of x. Inner product of two vectors x, y is written as $x^T y$. All vectors are column vectors.

1 Covariance Matrix

The variance of a random variable x_i is defined like

$$Var(x_i) = \mathbb{E}(x_i^2) - \mathbb{E}(x_i)^2$$

where $\mathbb{E}(\cdot)$ is the expectation function. To define the variance of a random vector (a series of random variable), we use the covariance matrix.

The covariance (matrix) of a random vector x can be defined as

$$[Cov(x)]_{ij} = Cov(x_i, x_j) = \mathbb{E}[x_i x_j] - \mathbb{E}[x_i]\mathbb{E}[x_j]$$

Theorem 1. *Let v be any constant m-dimension vector, the statement holds*

$$Var(v^T x) = v^T Cov(x) v$$

Proof.

$$\begin{aligned} Var(v^T x) &= Cov(v^T x, v^T x) \\ &= Cov(v_1 x_1 + \cdots + v_n x_n, v^T x) \\ &= v^T \left(Cov(x_i, v^T x)\right)_{i \le n} \\ &= v^T \left(Cov(x_i, v_1 x_1 + \cdots + v_n x_n)\right)_{i \le n} \\ &= v^T Cov(x) v \end{aligned}$$

□

Theorem 2. *Let $\{\alpha_k\}_n$ be a series of vectors such that*

$$\alpha_k = \arg\max_v \left\{ Var(v^T x) \mid \|v\| = 1, v^T x \perp \{\alpha_i^T x\}_{i<k} \right\}$$

Then α_k is the eigenvector corresponding to the k-th largest eigenvalues for matrix $Cov(x)$.

Proof. Let $\Sigma = Cov(x)$. First consider case $k = 1$. We use the Lagrange multiplier

$$L = Var(v^T x) - \beta(v^T v - 1) = v^T \Sigma v - \beta(v^T v - 1)$$

By differentiation,

$$\frac{\partial}{\partial v} L = 2\Sigma v - 2\beta v = 0 \iff \Sigma v = \beta v$$

We see when the maximum is attained, the vector is the eigenvector of Σ that corresponds to the largest eigenvalue. We derive the first vector α_1.

For $k = 2$, we want $\alpha_2^T x \perp \alpha_1^T x$, and the condition is equivalent to

$$0 = Cov(\alpha_2^T x, \alpha_1^T x) = \alpha_2^T \Sigma \alpha_1 = \beta \alpha_2^T \alpha_1$$

So we just require that $\alpha_2^T\alpha_1 = 0$. Again, by the Lagrange multiplier,

$$L = Var(v^T x) - \beta_1(v^T v - 1) - \beta_2(v^T \alpha_1) = v^T \Sigma v - \beta(v^T v - 1) - \beta_2(v^T \alpha_1)$$

By differentiation,

$$\frac{\partial}{\partial v} L = 2\Sigma v - 2\beta_1 v - \beta_2 \alpha_1 = 0$$

Since $\alpha_2 \perp \alpha_1$, the value for $\beta_2 = 0$ naturally, and we obtain

$$\Sigma v = \beta_1 v$$

When the maximum is attained, the vector is the eigenvector of Σ that corresponds to the second largest eigenvalue. This gives us α_2.

The remaining vectors can be derived in a similar way.

□

2 Principal Component

So far we get a series of independent vectors α_k which, by Theorem 2, maximize the variance of $v^T x$. The first component of x can be defined as $\alpha_1^T x$, and the second component is defined as $\alpha_2^T x$, and so on. We may think it as a series of linear transformations with corresponding matrices A_i such that

$$A_1 = [\alpha_1],\ A_2 = [\alpha_1\ \alpha_2],\ \cdots,\ A_n = [\alpha_1\ \alpha_2\ \cdots \alpha_n]$$

If we want to reduce the vector x to a smaller dimension r by a linear transformation, which is what PCA does, we simply do the corresponding matrix multiplication as

$$y = A_r^T x = (\alpha_1^T x,\ \alpha_2^T x,\ \cdots,\ \alpha_r^T x)$$

Notice that the variance is maximized in the scope of linear transformation.