Principal Component Analysis

Step 1: Compute Covariance Matrix

In R^2 space

Example 1

$\vec{d_i} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$cov(X, Y) = (\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}))/(n-1)$

Tries to capture the correlation / covariance between two variables

$cov(X, X) = Var(X) = (\sum_{i=1}^{n} (x_i - \bar{x})^2)/(n-1)$

$cov(Y, Y) = Var(Y) = (\sum_{i=1}^{n} (y_i - \bar{y})^2)/(n-1)$

$P(x \in [a, b])$ = $\int_a^b f(x)dx$

Where:

- $f(x) \geq 0$ for all x
- Area under f(x) from -∞ to +∞ = 1

$f(x)$ = normal distribution bell like curve

Example 2

When the points are uniformly distributed across the plane:

$cov(X, Y) = 0$

Create Covariance Matrix for R^2

$COV = \begin{bmatrix} cov(X, X) & cov(X, Y) \\ cov(Y, X) & cov(Y, Y) \end{bmatrix}$

Example in R^3

Stars = $\{\vec{s_1}, \vec{s_2}, \ldots, \vec{s_n}\}$

$$\vec{s_i} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$$

$$COV = \begin{bmatrix} cov(X,X) & cov(X,Y) & cov(X,Z) \\ cov(Y,X) & cov(Y,Y) & cov(Y,Z) \\ cov(Z,X) & cov(Z,Y) & cov(Z,Z) \end{bmatrix}$$

COV matrix is always symmetric

Step 2: Eigen Decomposition:

$det(C - \lambda I) = 0$

Roots of $spectrum(C) = \{\lambda_1, \lambda_2, \lambda_3\}$

Without loss of generality:
$\lambda_1 \geq \lambda_2 \geq \lambda_3$

Principal Components: $E_{\lambda_1}, E_{\lambda_2}, E_{\lambda_3}$

Take the **first 2 eigenspaces** to bring down to R^2

Step 3: Orthogonal Projection
U = span$\{E_{\lambda_1}, E_{\lambda_2}\}$

B = $[E_{\lambda_1}, E_{\lambda_2}]$

Use OP to bring data down from R^3 to R^2:

- Maximising data retention
- Minimising data loss
    - By picking the **first 2 eigenspaces**

Example 3

In R^2
Stars = $\{\vec{s_1}, \vec{s_2}, \vec{s_3}\}$ = $\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}\}$

Step 1: Compute Covariance Matrix:

$C = \begin{bmatrix} cov(X,X) & cov(X,Y) \\ cov(Y,X) & cov(Y,Y) \end{bmatrix}$ = $\begin{bmatrix} 3 & -3/2 \\ -3/2 & 3 \end{bmatrix}$

Step 2: Eigen Decomposition

Characteristic Polynomial = $\lambda^2 - 6\lambda + 27/4$

$Spectrum(C) = \{9/2, 3/2\}$

Pick highest value = 9/2

$E_{9/2}$ = 1st Principal Component = span{$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$}

Step 3: Orthogonal Projection:

U = span{$E_{9/2}$}
B = [$E_{9/2}$] = $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Projection Matrix = $B(B^TB)B^T$ = $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$($\begin{bmatrix} -1 & 1 \end{bmatrix}$ $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$) $\begin{bmatrix} -1 & 1 \end{bmatrix}$ = 1/2$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ = P

$P\vec{s_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$P\vec{s_2} = \begin{bmatrix} 3/2 \\ -3/2 \end{bmatrix}$

$P\vec{s_3} = \begin{bmatrix} -3/2 \\ 3/2 \end{bmatrix}$

Step 2: Eigen Decomposition

Step 3: Orthogonal Projection

Covariance Theorem

$cov(X,Y)^2 \leq Var(X)Var(Y)$