PCA * Compression for linear components only

- extract hidden lower dimensional structure from high dimensional dataset
- finds projection of the data onto directions that maximize the variance
- principal components are orthogonal vectors that capture most of the variance

Steps
1. mean subtraction
2. Compute covariance matrix
3. Compute eigenvalues and eigenvectors of covariance matrix
4. Multiply corresponding eigenvectors and eigenvalues

$\mu = \text{mean}(X)$ ← mean across rows
$\Sigma \leftarrow (X - \mu 1^T)^T (X - \mu 1^T)$
$\{\lambda_k, u_k\} \leftarrow$ top k eigenvalues and eigenvectors of Σ
U is a matrix of u_k

return $(X - \mu 1)U$

Reconstruction

$$\hat{X} = \text{mean} + x_1 \times \text{basis}_1 + \dots + x_k \times \text{basis}_k$$

limited to linear projections

Eigenvector and Eigenvalues

Prerequisites

eigenvector v $\quad Av = \lambda v$
eigenvalue λ

when you multiply matrix A with eigenvector λ, you get a multiple of the eigenvector

$Av = \lambda v$

$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

$Av - \lambda I v = 0$
$(A - \lambda I)v = 0$
$|A - \lambda I| = 0$

$\left| \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$

$\begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = 0$

$(3-\lambda)^2 - 1 = 0$
$\lambda^2 - 6\lambda + 9 - 1 = 0$
$(\lambda - 2)(\lambda - 4) = 0$
$\lambda_1 = 2, \quad \lambda_2 = 4$

$\lambda_1 = 2 \quad Av_1 = \lambda_1 I v_1$

$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} v_1 = 2 I v_1$

$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} v_1 - 2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} v_1 = 0$

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 0$

$v_{11} + v_{12} = 0$
$v_{11} + v_{12} = 0$

$v_{11} = -v_{12}$

$v_1 = \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} -v_{12} \\ v_{12} \end{bmatrix}$

$\phi = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$

$\lambda_1 = 4$

$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} v_2 - 4 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} v_2 = 0$

$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = 0$

$-v_{21} + v_{22} = 0$
$v_{21} - v_{22} = 0$

$v_{21} = v_{22}$

$v_2 = \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \begin{bmatrix} v_{22} \\ v_{22} \end{bmatrix}$

Lagrange Multiplier $\quad$ optimization with constraint

max/min of $f(x,y,z)$
subject to $g(x,y,z) = k$

multiply

Form: $F(x,y,z,\lambda) = f(x,y,z) - \lambda(g(x,y,z) - k)$

$F_x = 0$
$F_y = 0$
$F_z = 0$
$F_\lambda = 0$

plug back into original function

$C(x,y) = 6x^2 + 12y^2$
s.t. $g(x,y) = x + y = 90$

$F(x,y,\lambda) = 6x^2 + 12y^2 - \lambda(x + y - 90)$
$= 6x^2 + 12y^2 - \lambda x - \lambda y + 90\lambda$

$F_x = 12x - \lambda = 0 \rightarrow x = \frac{\lambda}{12}$
$F_y = 24y - \lambda = 0 \rightarrow y = \frac{\lambda}{24}$
$F_\lambda = -x + y + 90 = 0$
$-\frac{\lambda}{12} - \frac{\lambda}{24} + 90 = 0$
$-90 = -\frac{3\lambda}{24}$
$-30 = -\frac{\lambda}{24}$
$\lambda = 720$

$x = \frac{720}{12} = 60$
$y = \frac{720}{24} = 30$

PCA wants to find vectors v that save the most variance.

From every point, you want to project the point to v

$z_i = x^T v$

Variance is $(x_i^T v)^2 = z_i^2$

Maximize all of the projection

$\max \sum_{i=1}^n (x_i^T v)^2 = \sum_{i=1}^n z_i^2 = z^T z$

$= \underset{v}{\text{argmax}} \ (Xv)^T Xv$ ← norm v vector

$= \underset{v}{\text{argmax}} \ v^T X^T X v \quad \text{s.t. } v^T v = 1$

---

when we try to optimize with lagrange multiplier, we find that the v's are the eigenvectors of the covariance matrix

$$J = \underset{v, \lambda}{\text{argmax}} \ v^T X^T X v - \lambda(v^T v - 1)$$

$$\frac{\partial J}{\partial v} = 0 = \frac{\partial}{\partial v}\left[v^T X^T X v - \lambda(v^T v - 1)\right] = 0$$

$$= 2X^T X v - \lambda 2 v = 0$$

$$\underbrace{(X^T X)v = \lambda v}_{\text{eigen decomposition}}$$

vectors that capture the most variance are the eigenvectors of the covariance matrix