Principle Component Analysis

multivariate data: data describing N different variables

$\vec{w}_1, \vec{w}_2, \ldots \vec{w}_N \rightarrow$ each a vector with m components (representing data points).

→ Principle component Analysis can be used to reduce the dimension of our multivariate data set from N to K $(K < N)$

→ we want to determine the covariance between each of these variables:

→ first subtract the mean of each.

$$\hat{w}_i = \vec{w}_i - \text{mean}(\vec{w}_i)$$

$$A = \begin{bmatrix} - & \hat{w}_1 & - \\ - & \hat{w}_2 & - \\ & \vdots & \\ - & \hat{w}_N & - \end{bmatrix}$$

covariance:

$$\text{cov}(\hat{w}_i, \hat{w}_j) = \hat{w}_i \cdot \hat{w}_j$$

C = covariance matrix, the $N \times N$ matrix such that

$$C_{ij} = \text{cov}(\hat{w}_i, \hat{w}_j)$$

$$C = A A^T = \begin{bmatrix} - & \hat{w}_1 & - \\ - & \hat{w}_2 & - \\ & \vdots & \\ - & \hat{w}_N & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \hat{w}_1^T & \hat{w}_2^T & \cdots & \hat{w}_N^T \\ | & | & & | \end{bmatrix}$$

$$C = \begin{bmatrix} \hat{w}_1 \cdot \hat{w}_1 & \hat{w}_1 \cdot \hat{w}_2 & \cdots & \hat{w}_1 \cdot \hat{w}_N \\ \hat{w}_2 \cdot \hat{w}_1 & \hat{w}_2 \cdot \hat{w}_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \hat{w}_N \cdot \hat{w}_1 & \cdots & \cdots & \hat{w}_N \cdot \hat{w}_N \end{bmatrix}$$

↳ The covariance matrix is always symmetric.

→ The <u>eigenvalues</u> of a covariance matrix will always be real and positive

<u>Theorem: Symmetric matrices</u>

For an $N \times N$ symmetric matrix, eigenvectors from distinct eigenvalues will always be orthogonal to each other.

↳ will prove in Chapter 8.

So for the covariance matrix when we have N distinct eigenvectors, what kind of eigen basis can we form?

→ an <u>orthonormal eigen basis</u>

So can choose eignvectors:

$\vec{u}_1, \vec{u}_2, \ldots \vec{u}_N$ s.t. they are an orthonormal basis of $\mathbb{R}^N$

Example

	Student 1	Student 2	Student 3	Student 4	mean
$\vec{w_1}$ = grade on exam	95	85	75	65	80
$\vec{w_2}$ = # of hours studied	14	10	12	4	10
$\vec{w_3}$ = # of hours watching tv	1	5	5	9	5

$$A = \begin{bmatrix} 15 & 5 & -5 & -15 \\ 4 & 0 & 2 & -6 \\ -4 & 0 & 0 & 4 \end{bmatrix} \qquad A^T = \begin{bmatrix} 15 & 4 & -4 \\ 5 & 0 & 0 \\ -5 & 2 & 0 \\ -15 & -6 & 4 \end{bmatrix}$$

$C = AA^T = 3 \times 3$ matrix

eigenvalues: $\lambda_1 = 503.6 \quad \lambda_2 = 83.3 \quad \lambda_3 = 1.1$

eigenvectors: $u_1 = \begin{bmatrix} -0.996 \\ 0.090 \\ -0.008 \end{bmatrix} \quad u_2 = \begin{bmatrix} -0.076 \\ -0.787 \\ 0.613 \end{bmatrix} \quad u_3 = \begin{bmatrix} 0.049 \\ 0.611 \\ 0.790 \end{bmatrix}$

How much does the first eigenvector explain the total variance in the data?

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = 0.857$$

Consider the vector space spanned by $\vec{u}_1$

$W_1 = \text{span}(\vec{u}_1) \quad \dim(W_1) = 1$

↳ we have reduced the dimension of the data set from $N=3$ to $k=1$ but still have captured 85 % of the variance

if we let $\vec{S}_1 = \begin{bmatrix} 15 \\ 4 \\ -4 \end{bmatrix}$ the data (w/ mean subtracted) for student 1

Then $\text{proj}_{W_1} \vec{S}_1 = (\vec{u}_1 \cdot \vec{S}_1)\vec{u}_1$ is close to the original vector but now lies in W_1

Consider the space spanned by $\vec{u}_1$ and $\vec{u}_2$

$W_2 = \text{span}(\vec{u}_1, \vec{u}_2) \quad \dim(W_2) = 2$

$\text{proj}_{W_2} \vec{S}_1 = (\vec{u}_1 \cdot \vec{S}_1)\vec{u}_1 + (\vec{u}_2 \cdot \vec{S}_1)\vec{u}_2$

is even closer to the original vector ... why?

How much total variance do $\vec{u}_1$ and $\vec{u}_2$ account for?

We can also use Principal component analysis to classify new data.

→ consider data from a new student:

$\vec{S}_T = \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}$

We can consider the projection of $\vec{S}_T$ onto the reduced-dimension space:

$proj_{W_1} \vec{S}_T$

→ then calculate the distance from this to the projection of all the original students

$$d(\vec{S}_T, \vec{S}_i) = \| proj_{W_1} \vec{S}_T - proj_{W_1} \vec{S}_i \| \quad \text{for } i = 1 \ldots 4$$

For the $\vec{S}_i$ that this distance is minimized, we can say that $\vec{S}_T$ is most similar to $\vec{S}_i$