Lecture 13: Principal Components Analysis (PCA)

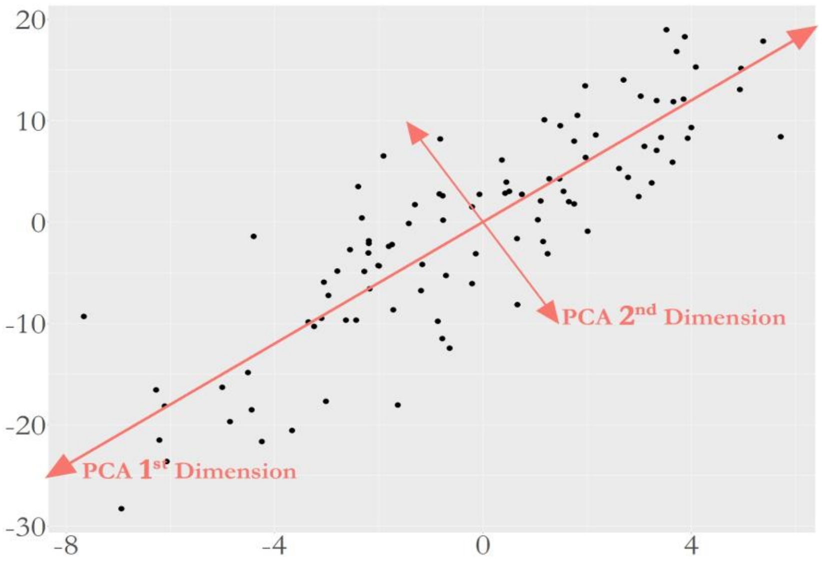
Instructor: Prof. Shuai Huang

Industrial and Systems Engineering

University of Washington

The idea of PCA

- The PCA method is built on the assumption that, for a multivariate dataset that has many variables, the dimensionality of the dataset is smaller than it appears to be.
- Another assumption of PCA is that the relationship between the underlying dimensions and the variables (surface dimensions) is linear.



Formulation of PCA

- To guide the estimation of the linear weights to combine the variables, the following formulation is used in PCA:

$$\boldsymbol{w}_{(1)} = \arg \max_{\boldsymbol{w}_{(1)}^T \boldsymbol{w}_{(1)}=1} \left\{ \sum_{n=1}^{N} \boldsymbol{x}_{(i)} \cdot \boldsymbol{w}_{(1)} \right\},$$

- where there are N samples and p variables, $\boldsymbol{x}_{(n)} \in R^{1\times p}$ is the nth sample, and $\boldsymbol{w}_{(1)} \in R^{p\times 1}$ is the linear weights vector of the first PC.
- Note that the constraint $\boldsymbol{w}_{(1)}^T \boldsymbol{w}_{(1)} = 1$ is to control the scale of the vector – without which an infinite number of solutions would exist. This also indicates that the absolute magnitudes of the weights are meaningless. Only the relative magnitudes are useful.

Formulation of PCA (cont'd)

- A more succinct form could be:

$$\boldsymbol{w}_{(1)} = \arg \max_{\boldsymbol{w}_{(1)}^T \boldsymbol{w}_{(1)}=1} \left\{ \boldsymbol{w}_{(1)}^T \mathbf{X}^T \mathbf{X} \boldsymbol{w}_{(1)} \right\},$$

- where $\mathbf{X} \in R^{N \times p}$ is usually called the data matrix that concatenate all the N samples into a matrix. $\mathbf{X}^T \mathbf{X}$ is actually the sample covariance matrix.
- The procedure for finding $\boldsymbol{w}_{(1)}$ could be readily used for finding $\boldsymbol{w}_{(2)}$, since with $\boldsymbol{w}_{(1)}$ removed, $\boldsymbol{w}_{(2)}$ is the largest variance source now.
- This process could be generalized as:

For the kth PC, create a dataset as $\mathbf{X}_{(k)} = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X}\, \boldsymbol{w}_{(s)} \boldsymbol{w}_{(s)}^T$.

Then, solve $\boldsymbol{w}_{(k)} = \arg \max_{\boldsymbol{w}_{(k)}^T \boldsymbol{w}_{(k)}=1} \left\{ \boldsymbol{w}_{(k)}^T \mathbf{X}_{(k)}^T \mathbf{X}_{(k)} \boldsymbol{w}_{(k)} \right\}$ for $\boldsymbol{w}_{(k)}$.

A small dataset example

X_1	X_2
-1	0
3	3
3	5
-3	-2
3	4
5	6
7	6
2	2

- First, we can calculate the sample covariance matrix as

$$\mathbf{S} = \mathbf{X}^T\mathbf{X} = \begin{bmatrix} 115 & 118 \\ 118 & 130 \end{bmatrix}.$$

- We can obtain the $\boldsymbol{w}_{(1)}$ by

$$\boldsymbol{w}_{(1)} = \arg\max_{\boldsymbol{w}_{(1)}^T\boldsymbol{w}_{(1)}=1} \left\{\boldsymbol{w}_{(1)}^T\mathbf{S}\boldsymbol{w}_{(1)}\right\}.$$

- The Lagrange form is

$$\boldsymbol{w}_{(1)}^T\mathbf{S}\boldsymbol{w}_{(1)} - \lambda_1\boldsymbol{w}_{(1)}^T\boldsymbol{w}_{(1)}.$$

- By taking the derivative of the lagrangian form with regards to $\boldsymbol{w}_{(1)}$, it is not hard to arrive at the equation:

$$\mathbf{S}\boldsymbol{w}_{(1)} - \lambda_1\boldsymbol{w}_{(1)} = 0.$$

- Thus, this is an eigenvalue problem of the matrix $\mathbf{S}$. We can solve it as $\lambda_1 = 240.74$ and $\boldsymbol{w}_{(1)} = [0.68, 0.73]$. Further, we can get that $\lambda_2 = 4.26$ and $\boldsymbol{w}_{(2)} = [-0.73, 0.68]$.

R lab

- Download the markdown code from course website
- Conduct the experiments
- Interpret the results
- Repeat the analysis on other datasets