# Principal Component Analysis (PCA)

- **Pattern recognition in high-dimensional spaces**
  - Problems arise when performing recognition in a high-dimensional space (curse of dimensionality).
  - Significant improvements can be achieved by first mapping the data into a *lower-dimensional sub*-space.

$$x = \begin{bmatrix} a_1 \\ a_2 \\ \ldots \\ a_N \end{bmatrix} --> reduce\ dimensionality --> y = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_K \end{bmatrix} \quad (K << N)$$

  - The goal of PCA is to reduce the dimensionality of the data **while retaining as much information as possible** in the original dataset.

# Principal Component Analysis (PCA)

- **Dimensionality reduction**
  - PCA allows us to compute a linear transformation that maps data from a high dimensional space to a lower dimensional sub-space.

***K x N***

$$y = Tx \text{ where } T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ t_{21} & t_{22} & \cdots & t_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ t_{K1} & t_{K2} & \cdots & t_{KN} \end{bmatrix}$$

$$b_1 = t_{11}a_1 + t_{12}a_2 + \ldots + t_{1n}a_N$$
$$b_2 = t_{21}a_1 + t_{22}a_2 + \ldots + t_{2n}a_N$$
$$\ldots$$
$$b_K = t_{K1}a_1 + t_{K2}a_2 + \ldots + t_{KN}a_N$$

# Principal Component Analysis (PCA)

- **Methodology**
  - Suppose $x_1$, $x_2$, ..., $x_M$ are $N$ x 1 vectors

Step 1: $\bar{x} = \frac{1}{M} \sum_{i=1}^{M} x_i$

Step 2: subtract the mean: $\Phi_i = x_i - \bar{x}$ (i.e., center at zero)

Step 3: form the matrix $A = [\Phi_1 \; \Phi_2 \cdots \Phi_M]$ ($N$x$M$ matrix), then compute:

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T = \frac{1}{M} A A^T$$

(sample **covariance** matrix, $N$x$N$, characterizes the *scatter* of the data)

Step 4: compute the eigenvalues of $C$: $\lambda_1 > \lambda_2 > \cdots > \lambda_N$

Step 5: compute the eigenvectors of $C$: $u_1, u_2, \ldots, u_N$

# Principal Component Analysis (PCA)

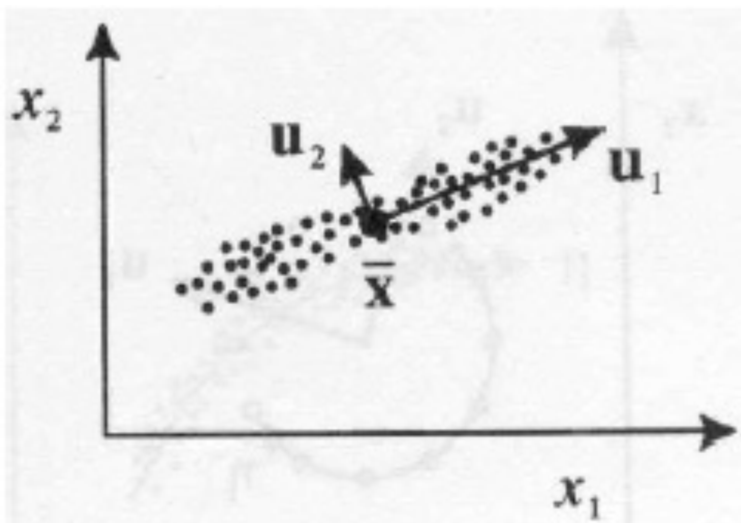
- **Linear transformation implied by PCA**
  - The effective linear transformation $R^N \rightarrow R^K$ that performs the dimensionality reduction is:

$$\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \dots \\ u_K^T \end{bmatrix} (x - \bar{x}) = U^T (x - \bar{x})$$

(i.e., simply computing coefficients of linear expansion)

# Principal Component Analysis (PCA)

- **Geometric interpretation**
  - PCA projects the data along the directions where the data varies the most.
  - These directions are determined by the eigenvectors of the covariance matrix corresponding to the largest eigenvalues.
  - The magnitude of the eigenvalues corresponds to the variance of the data along the eigenvector directions.



# Principal Component Analysis (PCA)

- **How to choose K (i.e., number of principal components) ?**
  - To choose $K$, use the following criterion:

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > Threshold \quad \text{(e.g., 0.9 or 0.95)}$$

  - In this case, we say that we "preserve" 90% or 95% of the information in our data. If K=N, then we "preserve" 100% of the information in our data.
  - PCA minimizes the reconstruction error, $e$, and it can be shown that the error is equal to:

$$e = ||x - \hat{x}|| \quad \rightarrow \quad e = 1/2 \sum_{i=K+1}^{N} \lambda_i$$

# PCA **Standardization**

- **Standardization**
  - The principal components are dependent on the *units* used to measure the original variables as well as on the *range* of values they assume.
  - Always standardize the data prior to using PCA.
  - A common standardization method is to transform all the data to have zero mean and unit standard deviation:

$$\frac{x_i - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \quad (\boldsymbol{\mu} \text{ and } \boldsymbol{\sigma} \text{ are the mean and standard deviation of } x_i\text{'s})$$