Component Analysis (DHS 10.13)

- Find the right features from given data, in general without any priori information
- Unsupervised approach
- Two popular techniques: (1) Principle Component Analysis and (2) Independent Component Analysis

(1) Principle Component Analysis (DHS 10.13.1)

- A popular technique to project d-dimensional data onto a lower-dimensional subspace in a sum-squared error sense.
- Known also as Karhunen-Loeve Transformation,
- Used in exploratory data analysis
- Uses Eigen Decomposition ($M = V\Sigma V^*$) or Singular Value Decomposition ($M = U\Sigma V^*$).

- Steps
  - Compute *d*-dimensional mean vector μ and *dxd* covariance matrix, Σ from data.
  - Compute eigenvectors and eigenvalues from the covariance matrix
    $D = V^{-1}\Sigma V$
    where D is the diagonal matrix of eigenvalues of Σ. The matrix V holds eigenvectors that diagonalizes the covariance matrix Σ.

    or

    Compute eigenvectors and eigenvalues from the characteristic equation (DHS A.2)
    Given a *dxd* matrix, M, $Mu = \lambda u$
    where $u = e_i$ eigenvectors and λ_i associated eigenvalues.
    Solve $|M - \lambda I| = \lambda^d + a_1\lambda^{d-1} ... + a_{d-1}\lambda + a_d = 0$ for each of its *d* roots λ_j.
    For each root solve for eigenvector e_j.
    Trace $tr[M] = \sum_{i=1}^{d} \lambda_i$ and Determinent $|M| = \prod_{i=1}^{d} \lambda_i$
    - Eigenvector indicate the axis
    - Eigenvalue indicate the amount of variance explained by the axis.
  - Sort eigenvectors and eigenvalues in a descending order.
  - Choose *k* eigenvalues and vectors of the large values.
  - Generally speaking, with *k* dimensionality, the signal subspace can be governed. The rest $d - k$ dimensions generally contain noises.
  - Form a *d* x *k* matrix A whose column consists of the *k* eigenvectors.
  - Re-represent the data by principal components consists of projecting the data onto the k-dimensional subspace $x' = A^T(x - \mu)$