Introduction to Data Science

GIRI NARASIMHAN, SCIS, FIU
PCA and Matrices

FROM JOHNSON & WICHERN, APPLIED MULTIVARIATE STATISTICAL ANALYSIS, 6TH ED
PCA: Principal Component Analysis

- Tool for Dimensionality Reduction
  - Reduces impact of curse of dimensionality
- Tool for finding Subspace in which data lies
- Summarization of data to find important variables
- Compares relative importance of variables
- Explains the most amount of variation in data
Principal Components

Figure 8.1 The constant density ellipse $x'\Sigma^{-1}x = c^2$ and the principal components $y_1, y_2$ for a bivariate normal random vector $X$ having mean 0.
Principal Components

First sample

principal component = linear combination $a'_i x_j$ that maximizes

the sample variance of $a'_i x_j$ subject
to $a'_1 a_1 = 1$

Second sample

principal component = linear combination $a'_2 x_j$ that maximizes the sample

variance of $a'_2 x_j$ subject to $a'_2 a_2 = 1$ and zero sample
covariance for the pairs ($a'_1 x_j$, $a'_2 x_j$)

At the $i$th step, we have

$ith$ sample

principal component = linear combination $a'_i x_j$ that maximizes the sample

variance of $a'_i x_j$ subject to $a'_i a_i = 1$ and zero sample
covariance for all pairs ($a'_{i} x_j$, $a'_k x_j$), $k < i$
PCA Animation

https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues
Points and Vectors

- Every point can be thought of a vector from the origin to that point
- \( p = (1, 3, 2) \)
Scalar Multiplication and Vector Addition

(a) Scalar multiplication of a vector by a scalar $2$.

(b) Vector addition of two vectors $x$ and $y$.

$2x_2$, $-\frac{1}{2}x$, $x_1$, $2x_1$, $x_2$, $x_1 + y_1$. 
Dot Product, Angles, Projections

\[ x'y = x_1y_1 + x_2y_2 + \ldots + x_ny_n \]

Projection of \( x \) on \( y = \frac{(x'y)}{y'y} y = \frac{(x'y)}{L_y} \frac{1}{L_y} y \]

\[ \cos(\theta) = \frac{x_1y_1 + x_2y_2}{L_x L_y} \]
Matrices & Transformations

- Arrays of Values, $A$
- Linear Transformations
  - $Ax = y$
- Matrix Product
  - Composing transforms
- Matrix Inverse: $AB = I \rightarrow B = A^{-1}$
Data as Matrices

\[
\mathbf{X} = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
= \begin{bmatrix}
  \mathbf{x}_1' \\
  \mathbf{x}_2' \\
  \vdots \\
  \mathbf{x}_n'
\end{bmatrix}
\leftarrow \text{1st (multivariate) observation}
\]

\[
= \begin{bmatrix}
  \mathbf{x}_1' \\
  \mathbf{x}_2' \\
  \vdots \\
  \mathbf{x}_n'
\end{bmatrix}
\leftarrow \text{nth (multivariate) observation}
\]
Eigenvalues and Eigenvectors

- Under transform $A$, eigenvectors experience change in magnitude only, but not direction
- $Ax = \lambda x; (A - \lambda I)x = 0$
- Characteristic Eq: $|A - \lambda I| = 0$
- Eigenvalues: $\lambda$
- Eigenvectors: $x$, $e$
Eigenvalues and Eigenvectors
If $A$ is symmetric, then the following decomposition holds true:

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots + \lambda_k e_k e_k'$$
The scalar $x'Ax$ is called **quadratic form**

A is **positive definite** if $x'Ax > 0$, whenever $x$ is a nonzero vector.

Equivalently, A is **positive definite** if all its eigenvalues are positive.
Matrix Inverse & Square Root

\[ A = \sum_{i=1}^{k} \lambda_i \, e_i e_i' = P \Lambda P' \]

\[ P = \begin{bmatrix} e_1, e_2, \ldots, e_k \end{bmatrix} \]

\[ \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \]

\[ A^{-1} = P \Lambda^{-1} P' = \sum_{i=1}^{k} \frac{1}{\lambda_i} e_i e_i' \]

\[ A^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i} \, e_i e_i' = P \Lambda^{1/2} P' \]
Dimension Reduction Revisited

- If we take $r$ eigenvectors, then
  - $P_r = [e_1, e_2, \ldots, e_r]$, and
  - $A$ can be approximated by taking $r$ eigenvectors.

$$
\Lambda = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_r
\end{bmatrix}_{(r \times r)}
$$

$$
P' = \begin{bmatrix}
P & \Lambda & P' \\
(k \times r) & (r \times r) & (r \times k)
\end{bmatrix}
$$
Random Matrices

\[
E(\mathbf{X}) = \begin{bmatrix}
E(X_{11}) & E(X_{12}) & \cdots & E(X_{1p}) \\
E(X_{21}) & E(X_{22}) & \cdots & E(X_{2p}) \\
\vdots & \vdots & \ddots & \vdots \\
E(X_{n1}) & E(X_{n2}) & \cdots & E(X_{np})
\end{bmatrix}
\]

\[
E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})
\]

\[
E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}
\]
Covariance Matrix

\[ \Sigma = E(\mathbf{X} - \mu)(\mathbf{X} - \mu)' \]

\[ \Sigma = \text{Cov}(\mathbf{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \]
Correlation Matrix, $\rho$

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

$$\rho = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1}$$

$$V^{1/2} \rho V^{1/2} = \Sigma$$
Singular Value Decomposition

- Spectral Decomp. for sq. symm. matrices
- Non-sq. asymmetric matrices?
  - Use sq. root of eigenvalues of $AA'$
  - Singular values of $A$

\[
A = U \Lambda V' \quad \begin{array}{ccc}
(k \times r) & (r \times r) & (r \times k)
\end{array}
\]
Dimensionality Reduction

- Given an $m \times k$ matrix $A$, we can approximate it by an $m \times s$ matrix $B$ with $s < k = \text{rank}(A)$. Then

$$B = \sum_{i=1}^{s} \lambda_i u_i v_i'$$

- Here we are picking $s$ singular values from SVD