CAP 5768: Introduction to Data Science

Giri NARASIMHAN

www.cis.fiu.edu/~giri/teach/5768.html
PCA and Matrices

From Johnson & Wichern, Applied multivariate statistical analysis, 6th Ed
PCA

- 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

![Diagram of principal components](image)

**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 $a'_i a_i = 1$

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

$i$th 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 $(a'_i x_j, a'_k x_j)$, $k < i$
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

(b) Vector addition

\[
\begin{align*}
2x_2 & = 2x \\
x_2 & = x \\
\frac{1}{2}x & = -\frac{1}{2}x
\end{align*}
\]

\[
\begin{align*}
x_1 & = x \\
y_1 & = y \\
x + y & = x + y
\end{align*}
\]
Dot Product, Angles, Projections

\[ x' y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \]

Projection of \( \mathbf{x} \) on \( \mathbf{y} \):
\[ \text{Projection of } \mathbf{x} \text{ on } \mathbf{y} = \frac{(x' y)}{y' y} \mathbf{y} = \frac{(x' y)}{L_y} \frac{1}{L_y} \mathbf{y} \]
Matrices & Transformations

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

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots + \lambda_k e_k e_k'$$

where $\lambda_i$ are the eigenvalues of $A$, and $e_i$ are the corresponding eigenvectors.
Quadratic Form

- 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

\[ \mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i' = \mathbf{P} \Lambda \mathbf{P}' \]

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

\[ \mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i} \mathbf{e}_i \mathbf{e}_i' = \mathbf{P} \Lambda^{1/2} \mathbf{P}' \]

\[ \mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k] \]

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

- If we take \( r \) eigenvectors, then
  \[
P_r = [e_1, e_2, ..., e_r], \text{ and}
  \]

- A can be approximated by taking \( r \) eigenvectors
  \[
P \Lambda P' = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_r \\
\end{pmatrix}
\]
Random Matrices

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

\[ E(X + Y) = E(X) + E(Y) \]
\[ E(AXB) = A E(X) B \]
Covariance Matrix

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

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

\[ \mathbf{v}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix} \]

\[ \rho = (\mathbf{v}^{1/2})^{-1} \Sigma (\mathbf{v}^{1/2})^{-1} \]

\[ \mathbf{v}^{1/2} \rho \mathbf{v}^{1/2} = \Sigma \]