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

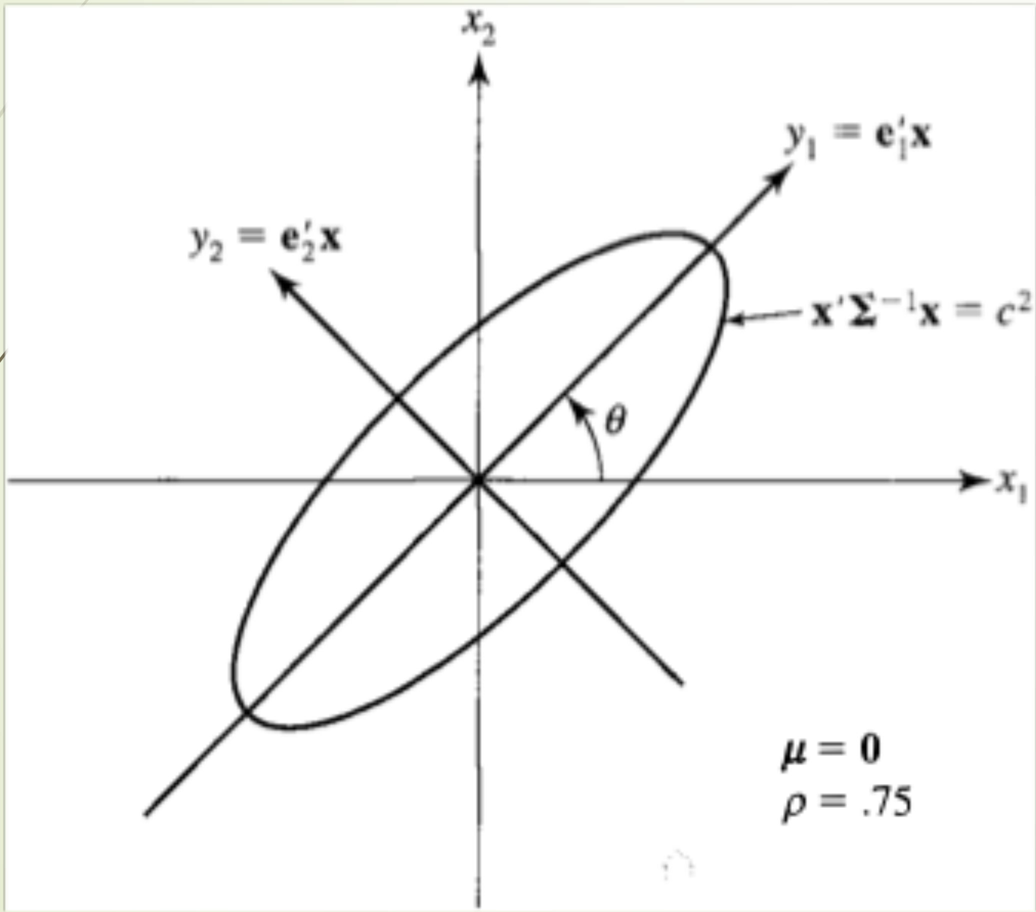


Figure 8.1 The constant density ellipse $\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x} = c^2$ and the principal components y_1, y_2 for a bivariate normal random vector $\mathbf{X}$ having mean $\mathbf{0}$.

Principal Components

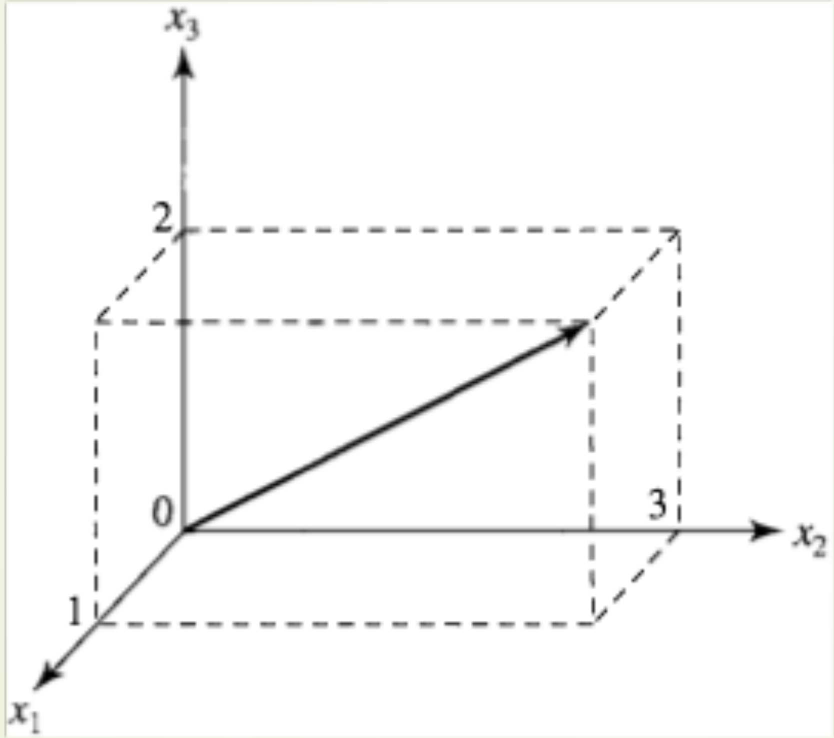
First *sample* principal component = linear combination $\mathbf{a}_1'\mathbf{x}_j$ that maximizes the sample variance of $\mathbf{a}_1'\mathbf{x}_j$ subject to $\mathbf{a}_1'\mathbf{a}_1 = 1$

Second *sample* principal component = linear combination $\mathbf{a}_2'\mathbf{x}_j$ that maximizes the sample variance of $\mathbf{a}_2'\mathbf{x}_j$ subject to $\mathbf{a}_2'\mathbf{a}_2 = 1$ and zero sample covariance for the pairs $(\mathbf{a}_1'\mathbf{x}_j, \mathbf{a}_2'\mathbf{x}_j)$

At the ith step, we have

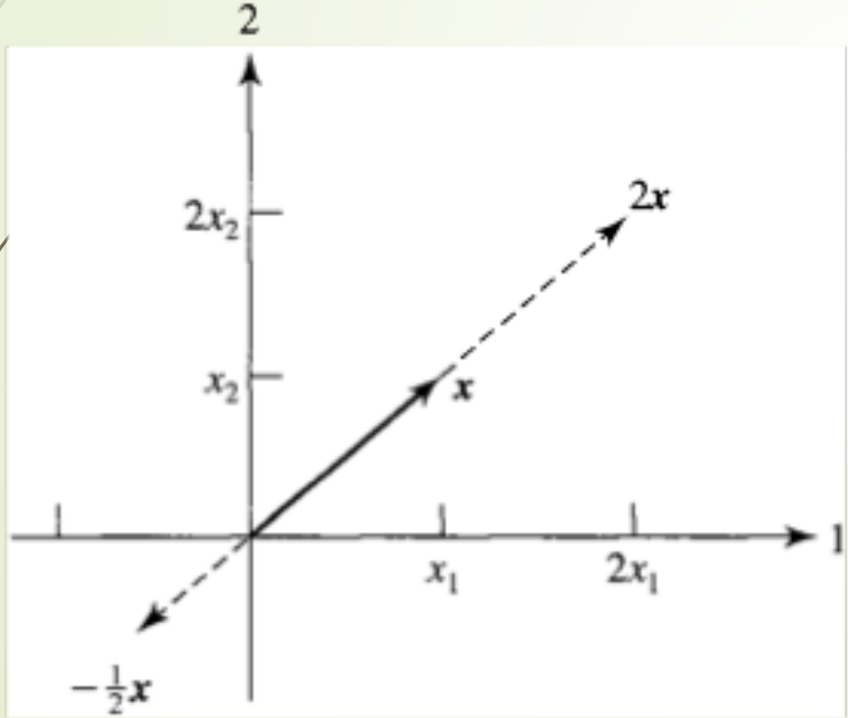
ith *sample* principal component = linear combination $\mathbf{a}_i'\mathbf{x}_j$ that maximizes the sample variance of $\mathbf{a}_i'\mathbf{x}_j$ subject to $\mathbf{a}_i'\mathbf{a}_i = 1$ and zero sample covariance for all pairs $(\mathbf{a}_i'\mathbf{x}_j, \mathbf{a}_k'\mathbf{x}_j)$, $k < i$

Points and Vectors

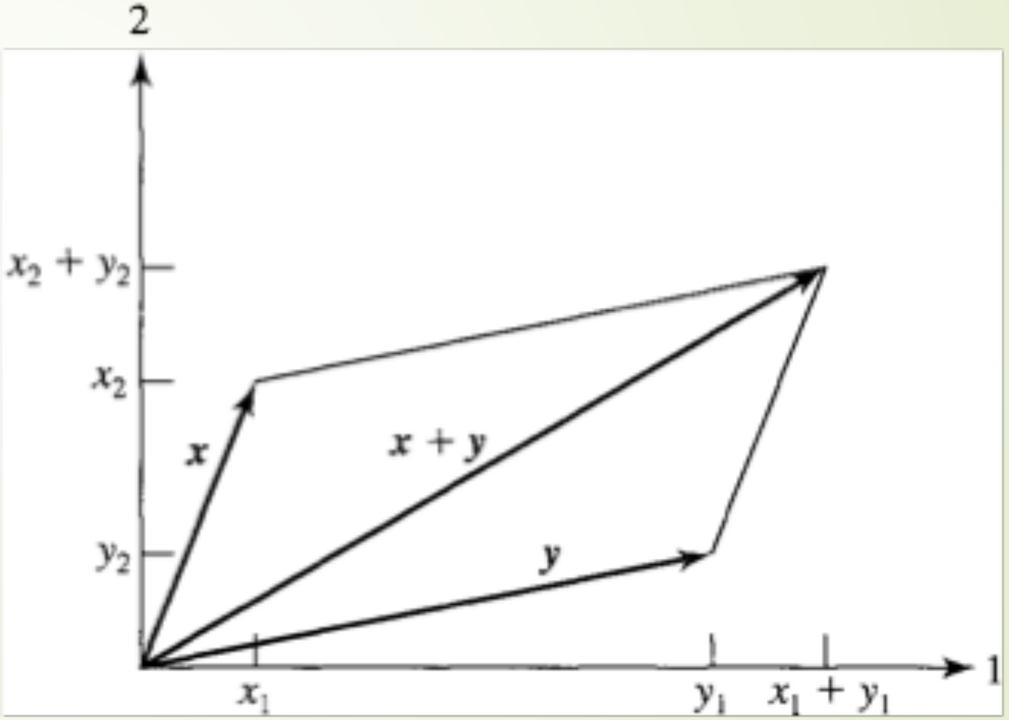


- Every point can be thought of a vector from the origin to that point
- p = (1, 3, 2)

Scalar Multiplication and Vector Addition



(a)

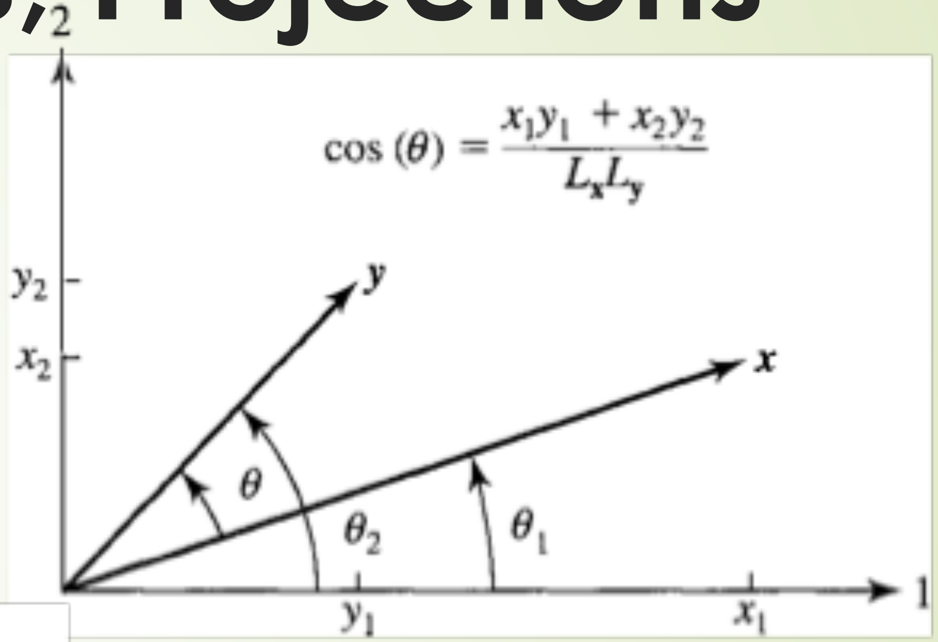


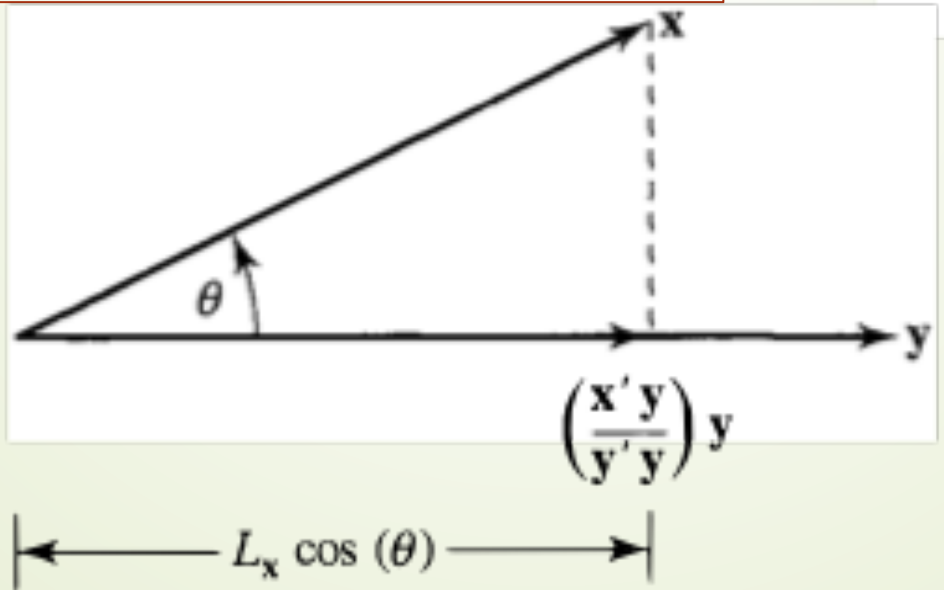
(b)

Dot Product, Angles, Projections

$$\mathbf{x}'\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

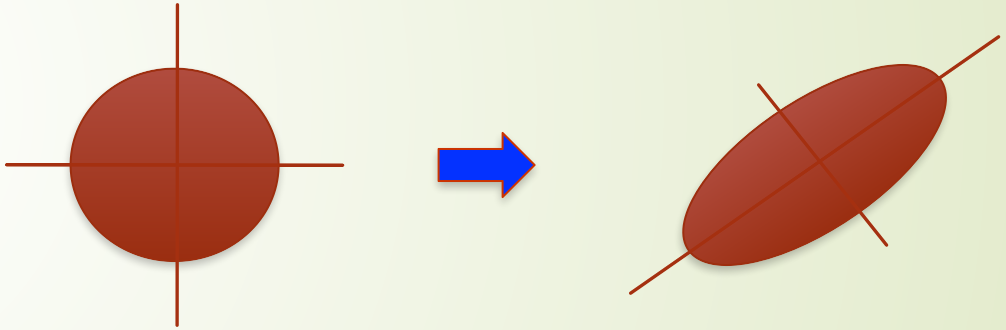
$$\text{Projection of } \mathbf{x} \text{ on } \mathbf{y} = \frac{(\mathbf{x}'\mathbf{y})}{\mathbf{y}'\mathbf{y}}\mathbf{y} = \frac{(\mathbf{x}'\mathbf{y})}{L_{\mathbf{y}}}\frac{1}{L_{\mathbf{y}}}\mathbf{y}$$





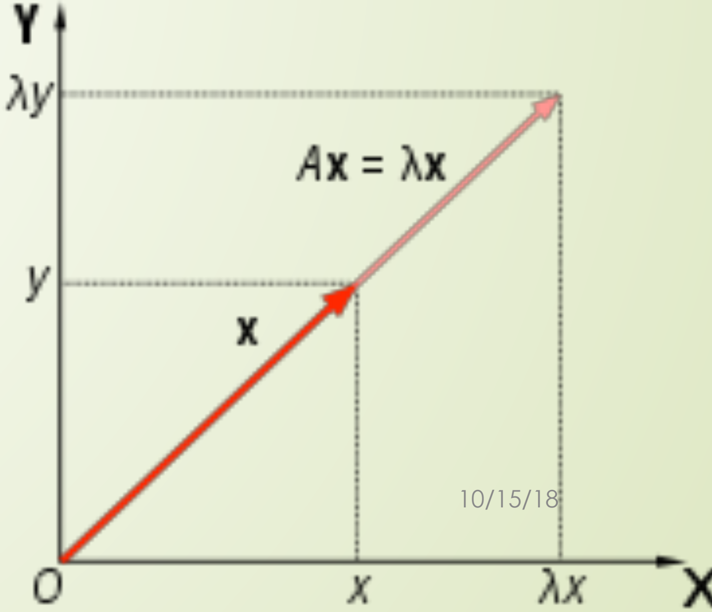
Matrices & Transformations

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

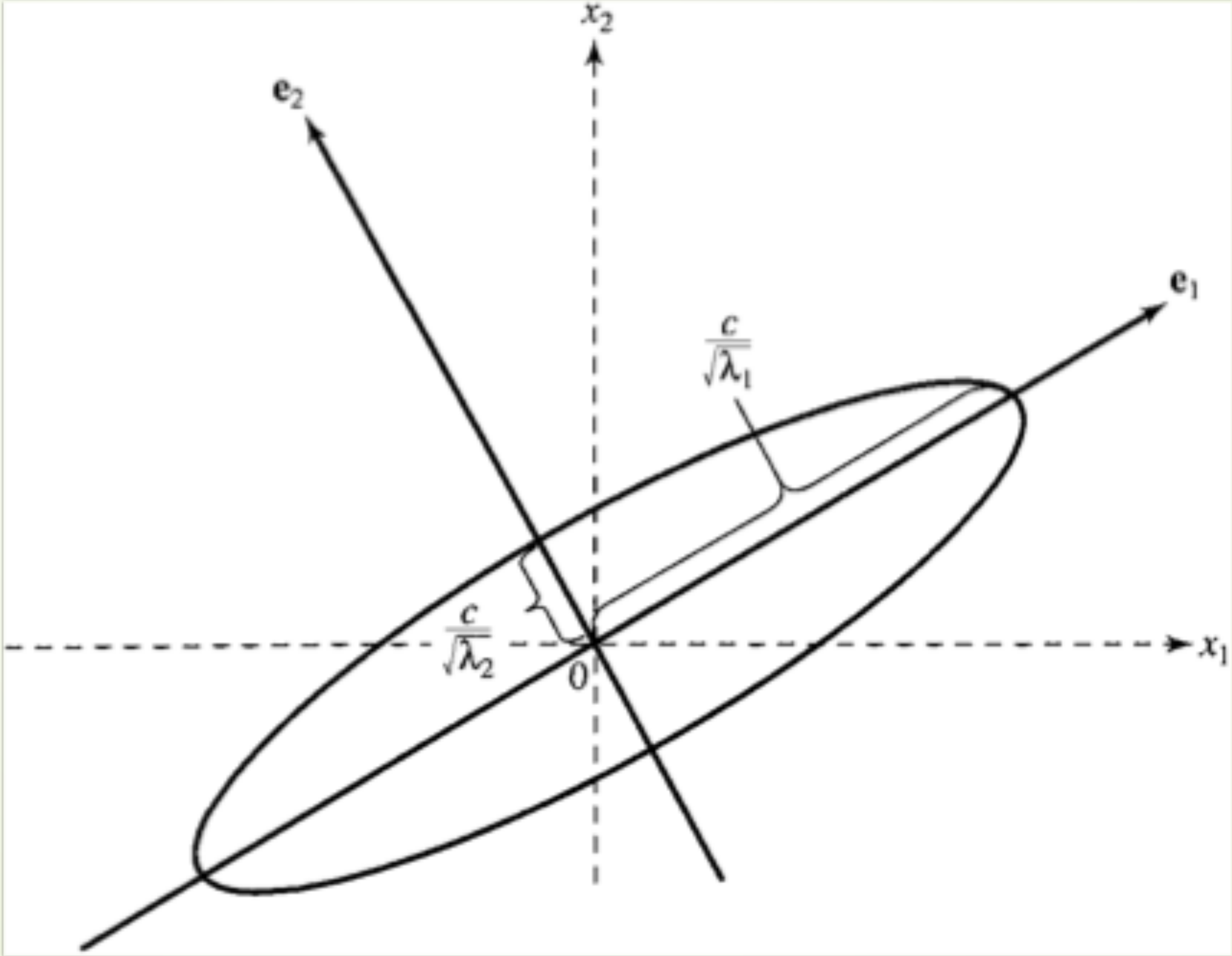


Eigenvalues and Eigenvectors

- 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: λ
- Eigenvectors: x, e



Eigenvalues and Eigenvectors



Spectral Decomposition

- If A is symmetric, then the following decomposition holds true:

$$\underset{(k\times k)}{\mathbf{A}} = \lambda_1 \underset{(k\times 1)}{\mathbf{e}_1} \underset{(1\times k)}{\mathbf{e}_1'} + \lambda_2 \underset{(k\times 1)}{\mathbf{e}_2} \underset{(1\times k)}{\mathbf{e}_2'} + \cdots + \lambda_k \underset{(k\times 1)}{\mathbf{e}_k} \underset{(1\times k)}{\mathbf{e}_k'}$$

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

$$\underset{(k\times k)}{\mathbf{A}} = \sum_{i=1}^{k} \lambda_i \underset{(k\times 1)}{\mathbf{e}_i} \underset{(1\times k)}{\mathbf{e}_i'} = \underset{(k\times k)}{\mathbf{P}} \underset{(k\times k)}{\mathbf{\Lambda}} \underset{(k\times k)}{\mathbf{P}'}$$

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

$$\underset{(k\times k)}{\mathbf{\Lambda}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

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

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

Dimension Reduction Revisited

- If we take r eigenvectors, then
  - $P_r = [e_1, e_2, \ldots, e_r]$, and

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

  - A can be approximated by taking r eigenvectors

$$\underset{(k \times r)}{P} \; \underset{(r \times r)}{\Lambda} \; \underset{(r \times k)}{P'}$$

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

$$\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$$

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Correlation Matrix, ρ

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

$$\boldsymbol{\rho} = \left(\mathbf{V}^{1/2}\right)^{-1} \boldsymbol{\Sigma} \left(\mathbf{V}^{1/2}\right)^{-1}$$

$$\mathbf{V}^{1/2} \boldsymbol{\rho} \mathbf{V}^{1/2} = \boldsymbol{\Sigma}$$