Principal Component Analysis

ECE/CS 498 DS U/G

Lecture 12

Ravi K. Iyer

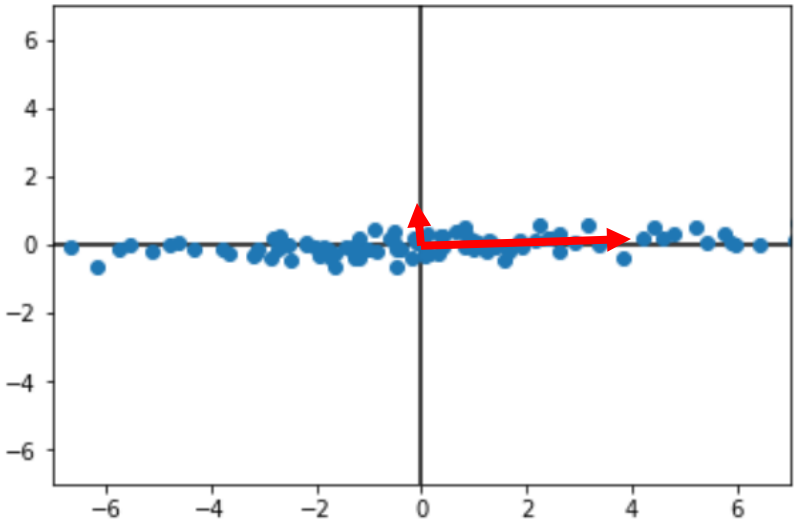
Dept. of Electrical and Computer Engineering

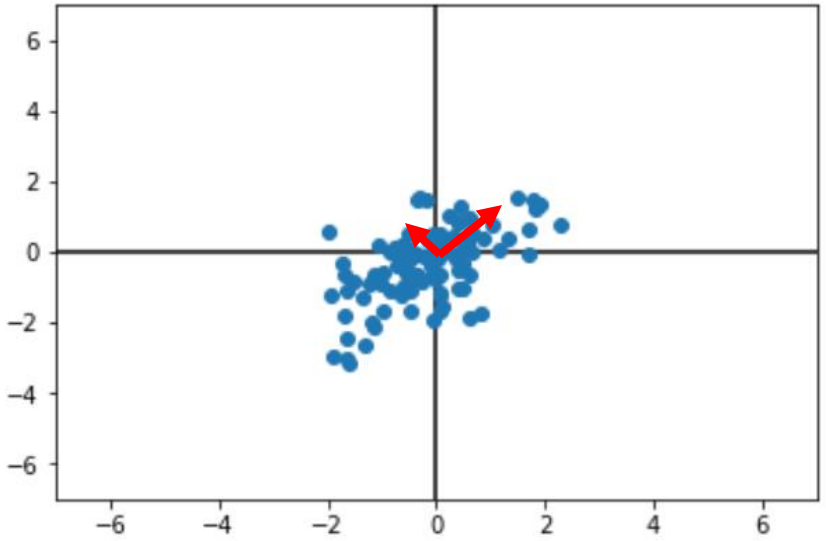
University of Illinois at Urbana Champaign

Announcements

- Homework 3 will be released on Friday, Mar 1
  - Due on Friday, Mar 8
  - No late submission will be allowed for this HW
- In class activity 3 today
  - Principal Component Analysis, Clustering
- Please submit half page write of Dr. Weinshilboum's lecture today
- MP1 grades released
  - Contact TAs if you have any questions
- No discussion section on Friday, Mar 1
  - Additional office hours

The Algebra of PCA: Covariance/Correlation Matrix





$$\begin{bmatrix} 10 & 0.5 \\ 0.5 & 0.1 \end{bmatrix}$$

Covariance matrix

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

Correlation matrix

- If variance of features is not on comparable scale, then principal components have high contribution from features with large variance

The Algebra of PCA: Covariance/Correlation Matrix

- PCA can be found using the covariance matrix OR the correlation matrix
- **Covariance Matrix:**
  - Variables must be in same units
  - Emphasizes variables with most variance
  - Generally, using covariance's among variables only makes sense if they are measured in the same units
- **Correlation Matrix:**
  - Variables are standardized (mean 0.0, SD 1.0)
  - Variables can be in different units
  - All variables have same impact on analysis

$$r_{ij} = \frac{C_{ij}}{\sqrt{V_i V_j}}$$

r_{ij}: Correlation between variables i and j

C_{ij}: Covariance of variables i and j

V_j: Variance of variable j

	X_1	X_2
X_1	6.6707	3.4170
X_2	3.4170	6.2384

Variance-covariance Matrix

Trace (sum of diagonals): 12.9091

	X_1	X_2
X_1	1.0000	0.5297
X_2	0.5297	1.0000

Correlation Matrix

Trace (sum of diagonals): 2.0

PCA with Correlation Matrix

- Compute correlation matrix from covariance matrix:

$$r_{ij} = \frac{C_{ij}}{\sqrt{V_i V_j}}$$

Correlation between variables i and j

Covariance of variables i and j

Variance of variable j

- Solve eigenvalue equation: $S_{cor} a = \lambda a$

Correlation Matrix

- Compute eigenvalues by solving: $|S_{cor} - \lambda I| = 0$

- Compute eigenvectors (principal components) by solving the following for each eigenvalue λ_i: $(S_{cor} - \lambda_i I) a_i = 0$

- Principal components may be different for correlation matrix and covariance matrix