CSC-591: Foundations of Data Science
T/Th. 12:50-2:05pm. EBI-1005.

Ranga Raju Vatsavai
Chancellors Faculty Excellence Associate Professor in Geospatial Analytics
Department of Computer Science, North Carolina State University (NCSU)
Associate Director, Center for Geospatial Analytics, NCSU
&
Joint Faculty, Oak Ridge National Laboratory (ORNL)

W11: 10/27/15-10/29/15

Admin: Changes in grading

- Midterm-1: 15%
- Midterm-2: 20%
- Final: 35%

- Midterm-2 Topics
  - Regression (all topics covered in the class)
  - Information theory, attribute selection
  - Dimensionality reduction
  - Nonparametric hypothesis testing

Dimensionality Reduction

- Previously, feature (or attribute) selection
  - Preserves original (reduced) attribute set
    - X_1, X_2, X_3, …X_{d-1}, X_d
- Dimensionality reduction
  - Preserve as much structure as possible
  - Structure: relationships that affects class separability
  - New feature (transformed) space

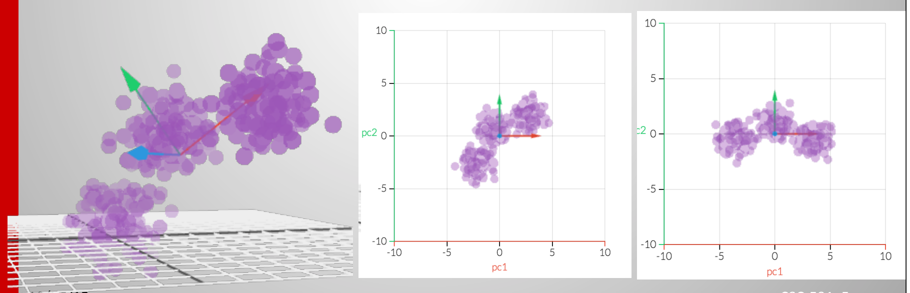
Principal Component Analysis

- Principal component analysis (PCA) is an orthogonal transformation original (correlated data) variables into a set of values of linearly uncorrelated variables called principal components.
  - 1st component: direction of greatest variability in the data
  - 2nd component: orthogonal to 1st, greatest variability of what's left
  - … and so on until d (original dimensionality)

NC STATE UNIVERSITY

PCA

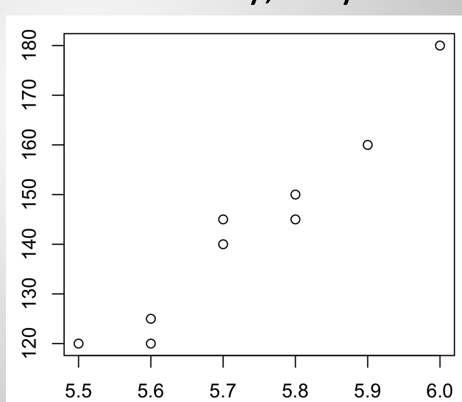
- Choose first “m” components which become m new dimensions (or new coordinate system)
  – Project data onto new dimensions (i.e., change coordinate of every data point to these new dimensions)



NC STATE UNIVERSITY

Greatest Variability

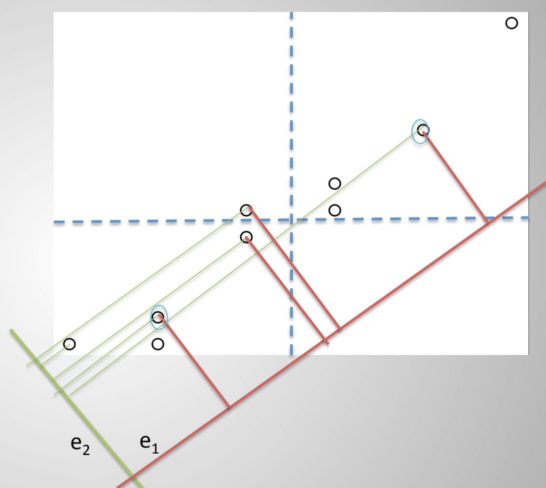
- 2-d example
- Find z that maximize variability, why?



NC STATE UNIVERSITY

Greatest Variability

- 2-d example
- Find z that maximize variability, why?

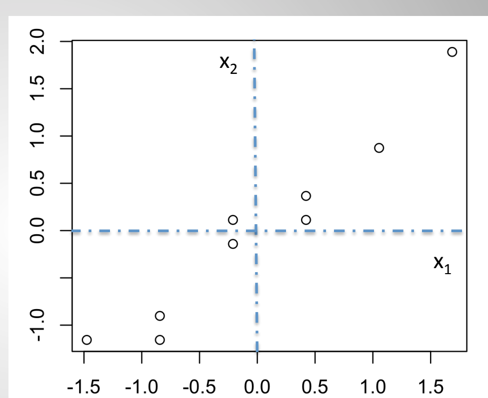


Reduce cases where two points are close in e-space but very far in (x,y)-space
Minimize distances between original points and their projections

NC STATE UNIVERSITY

How to Get Principal Components

- Center the data at 0. That is, $x_{1,j} = x_{i,j} - \mu_1$
- Normalize attributes if needed



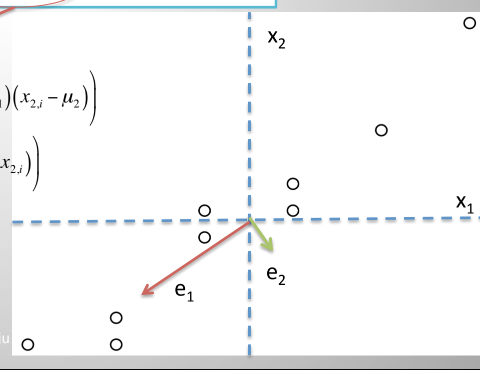
Reduce cases where two points are close in e-space but very far in (x,y)-space
Minimize distances between original points and their projections

Principal Components

- Compute covariance matrix Σ
  - Do x1 and x2 tend to increase (or decrease) together?

	x_1	x_2
x_1	1.0000000	0.9696882
x_2	0.9696882	1.0000000

  - $\mathrm{Cov}(x_1,x_2) = \frac{1}{n}\left(\sum_{i=1}^{n}(x_{1,i}-\mu_1)(x_{2,i}-\mu_2)\right)$
    $= \frac{1}{n}\left(\sum_{i=1}^{n}(x_{1,i})(x_{2,i})\right)$



© Raju

Principal Components

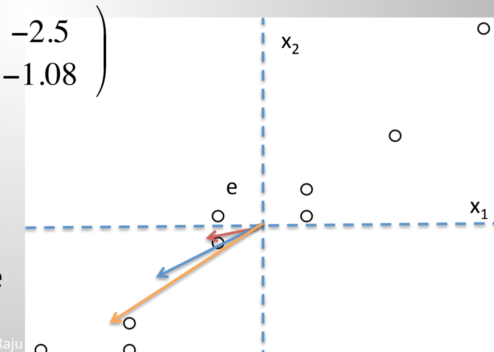
- Multiply covariance matrix Σ by a vector

$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix}\begin{pmatrix} -1 \\ +1 \end{pmatrix} = \begin{pmatrix} -1.2 \\ -0.2 \end{pmatrix}$$

$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix}\begin{pmatrix} -1.2 \\ -0.2 \end{pmatrix} = \begin{pmatrix} -2.5 \\ -1.08 \end{pmatrix}$$

…

- Repeat
  - Turns towards the direction of variance



© Raju

NC STATE UNIVERSITY

Principal Components

- We saw that the slope of vector converging towards maximum variance, but length is growing faster
  - So we want a vector e_i's that aren't turned by Σ
- Finding the basis of Σ
  - Σ e = λ e
  - e's are called eigenvectors
  - λ's are corresponding eigenvalues
- Principal components = eigenvectors with largest eigenvalues

10/27/15 © Raju Vatsavai CSC-591. 11

NC STATE UNIVERSITY

Finding Principal Components

- Find eigenvalues
  - Solve: det(Σ-λI) = 0

$$\det\begin{pmatrix} 2.0-\lambda & 0.8-0 \\ 0.8-0 & 0.6-\lambda \end{pmatrix} = 0$$

$$\det\begin{pmatrix} 2.0-\lambda & 0.8-0 \\ 0.8-0 & 0.6-\lambda \end{pmatrix} = (2.0-\lambda)(0.6-\lambda)-(0.8)(0.8) = 0$$

$$\lambda^2 - 2.6\lambda + 0.56 = 0$$

$$\{\lambda_1, \lambda_2\} = \frac{1}{2}\left(2.6 \pm \sqrt{2.6^2 - 4 * 0.56}\right) = \{2.36, 0.23\}$$

10/27/15 © Raju Vatsavai CSC-591. 12

Finding Principal Components

- Find i^{th} eigenvector by solving $\Sigma\, e_i = \lambda_i\, e_i$

$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix}\begin{pmatrix} e_{1,1} \\ e_{1,2} \end{pmatrix} = 2.36\begin{pmatrix} e_{1,1} \\ e_{1,2} \end{pmatrix}$$

$$2.0e_{1,1} + 0.8e_{1,2} = 2.36e_{1,1}$$

$$0.8e_{1,1} + 0.6e_{1,2} = 2.36e_{1,2}$$

$$0.8e_{1,2} = (2.36 - 2.0)e_{1,1}$$

$$e_{1,1} = \frac{0.8}{0.36}e_{1,2} = 2.2e_{1,2}$$

- Lots of vectors that satisfy this condition
- The simplest is $e_1 = \begin{bmatrix} 2.2 \\ 1.0 \end{bmatrix}$

Finding Principal Components

- To avoid multiple solutions, we want e_i's to be unit vectors, i.e. $\|e_1\| = 1$

$$e_1 = \begin{bmatrix} 0.91 \\ 0.41 \end{bmatrix}$$

- Now solve for 2^{nd} eigenvector

$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix}\begin{pmatrix} e_{1,1} \\ e_{1,2} \end{pmatrix} = 0.23\begin{pmatrix} e_{1,1} \\ e_{1,2} \end{pmatrix}$$

$$e_2 = \begin{bmatrix} -0.41 \\ 0.91 \end{bmatrix}$$

NC STATE UNIVERSITY

Projection

- After finding principal components, we need to project our data onto new dimensions
- e_1, … e_m are new (first m) dimensions
- We have instance x = {x_1, … x_d} (original coordinates)
- We want new coordinates x' = {x'_1, … x'_m}
  - Center each instance: (x' – μ)
  - Project to each dimension: $(x' - \mu)^T e_j$ for j=1,…,m

NC STATE UNIVERSITY

Projection

- After finding principal components, we need to project our data onto new dimensions
- e_1, … e_m are new (first m) dimensions
- We have instance x = {x_1, … x_d} (original coordinates)
- We want new coordinates x' = {x'_1, … x'_m}
  - Center each instance: (x' – μ)
  - Project to each dimension: $(x' - \mu)^T e_j$ for j=1,…,m

NC STATE UNIVERSITY

Key Properties

- Eigenvectors (e) maximizes the variance
- What is the variance along eigenvector
- Variance of projected points $(x^T e)=\lambda$

NC STATE UNIVERSITY

How many dimensions

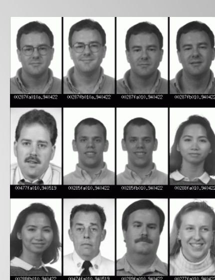
- Of all eigenvectors $e_1, \ldots, e_d$, we want e_m, m<<d
- We know, eigenvalue λ_i = variance along ei
  - Sort eigenvector s.t. $\lambda_1 \le \lambda_2 \ldots \le \lambda_d$
  - Pick first m eigenvectors which explain 90% or (95%) of total variance
- Or, plot eigenvalues as function of dimensions
  - (like K-means)

Example: Eigen Faces

- Face recognition using PCA (Eigenfaces)

Representation of Images

- Data: Set of N image of size KxK
- Convert the images into K^2xN matrix
- Each NxN image becomes a N^2 vector in the matrix

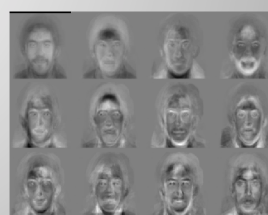


Single Image

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & -1 & 2 \\ 4 & 5 & 1 \end{bmatrix}_{3\times 3} \rightarrow \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \\ -1 \\ 2 \\ 4 \\ 5 \\ 1 \end{bmatrix}_{9\times 1} \rightarrow \begin{bmatrix} 1 & 5 & \ldots & & \ldots & 8 \\ 2 & 1 & .. & & .. & 1 \\ \ldots & \ldots & \ldots & & \ldots & \ldots \\ 3 & 3 & .. & & .. & -2 \\ -1 & 2 & \ldots & \ldots & \ldots & 10 \\ \ldots & \ldots & \ldots & & \ldots & \ldots \\ 4 & 6 & \ldots & & \ldots & 3 \\ 5 & 2 & \ldots & & \ldots & 1 \\ 1 & 7 & \ldots & & \ldots & -2 \end{bmatrix}_{N^2 M}$$

Representation of Images

1. Average dataset
2. Center the data
   1. Each face differs from the average by vector
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues (set of M eigenvectors each K^2 dim)
5. Convert into images (take each column (Eigenvector) and convert it KxK image)



Nonlinear Dimensionality Reduction

- Its difficult to represent more than 3-d data
- Simplify by assuming that the data of interest lie on an embedded non-linear manifold within the higher-dimensional space. If the manifold is of low enough dimension, the data can be visualized in the low-dimensional space.
- Several approaches
  – ISOMAP
  – LLE

© Raju Vatsavai

NC STATE UNIVERSITY

Additional Resources and Acknowledgements

- PCA Tutorials/References
  - A Tutorial on Principal Component Analysis, by Jonathon Shlens, arXiv
  - Principal Component Analysis, by H. Abdi, L. Williams, Weily, 2010
  - (search on web for these articles)
- D. Mladenic

© Raju Vatsavai