Principal Component Analysis & Data Reduction

Hugh R. Wilson

- Purpose of PCA
- Simple example
- Derivation of PCA
- Dimensionality reduction
- More complex example
- How many dimensions?
- Experiment on learning face PCA
- Nonlinear PCA

YORK U
UNIVERSITÉ
UNIVERSITY

Centre for
Vision Research

Reading

- Today: Principal components: Kutz, pp. 387 (bottom) - 393
- Wed: Linear discriminant analysis: Kutz, pp. 442 - 445

Purpose of PCA

- Simplify data by reducing dimensionality
- Perfect correlation between 2 measurements:
    - Redundancy
    - Example: nose length = eye separation
    - One dimension is irrelevant
- Remove least significant dimensions
- All PCs mutually orthogonal: independent information
- Remove noise (small variances in highest PCs)
- Equivalent to rotation of axes

Gram-Schmidt Orthogonalization

- Two linearly independent column vectors A & B
- Assume vectors normalized to unit length
- Covariance: $A'B = \rho_{AB}$
- A' is transpose of A
- Make dimensions independent (orthogonal)
- $\cos(AB\_angle) = \rho_{AB}$
- $B_{new} = B - \rho_{AB}A$
- $A' B_{new} = A'B - A'A\rho = \rho - \rho = 0$
- Can repeat with vectors C, D, etc. as long as all are linearly independent
- $C_{new} = C - \rho_{AC}A - \rho_{BC}B_{new}$

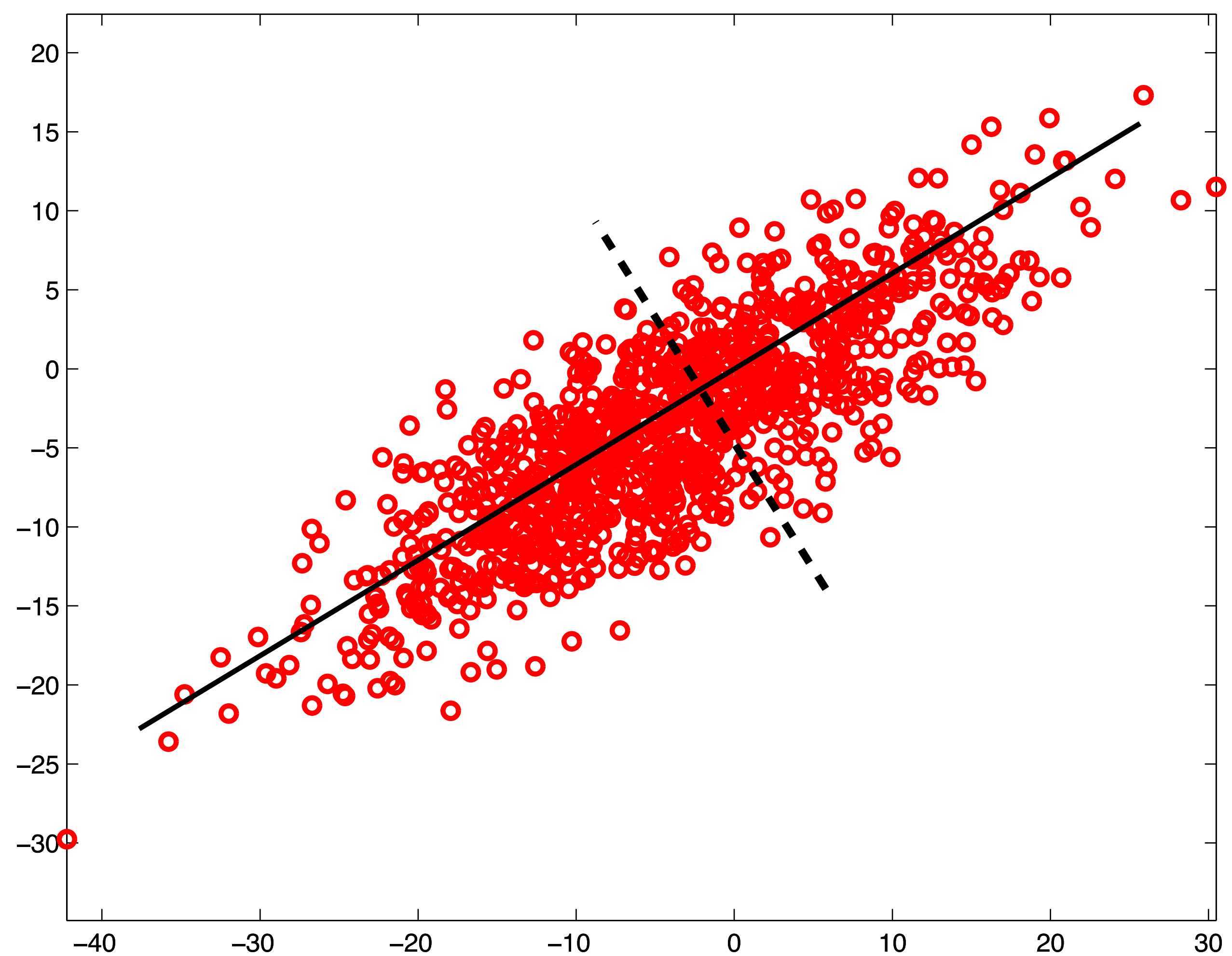
YORK UNIVERSITÉ UNIVERSITY

Data for PCA

- Make N measurements on each of M specimens, $M \geq N$
- Eg: nose length, eye separation, mouth width, nose width
- Eg: multiple disease symptom measurements on people
- Create data matrix:
    - M rows, one for each specimen
    - N columns, one for each measurement
- Data represented as M points in N dimensional space
- Can we simplify?

PCA in 2 Dimensions

- Two measurements that are correlated
- PC1 explains 92.5% of the variance
- PC2 explains 7.5% of the variance



PCA Approach

- Use covariance (or correlation) of measurements to reduce dimensionality
- Require PCA dimensions to be orthogonal: independent info
- First PC: accounts for largest % of variance
- Second PC: accounts for next largest % of variance

PCA Derivation

- Subtract means from data, **D**
- Find vector **v** to maximize variance of projected data:
  - var(**Dv**) = (1/(N-1))(**Dv**)'(**Dv**) = (1/(N-1))(**v'D'**)(**Dv**) = **v'Sv**
  - **S** is the data covariance matrix: (1/(N-1))**D'D**
- Normalization constraint: **v'v = 1**
- Use Lagrange multiplier for constraint & maximize:
  - **v'Sv +** λ(**1 - v'v**)
- (d/d**v'**)(**v'Sv +** λ(**1 - v'v**)) = 0
- **Sv =** λ**v**
- PCs are eigenvectors of covariance matrix
- Variance explained is λ
- Vector with largest λ is PC1, second largest λ is PC2, etc.
- Equivalent to minimizing mean squared error of representation in lower dimensional space

YORK UNIVERSITÉ UNIVERSITY

PCA in Matlab™

- Data: M rows of subjects **x** N measurements on each (M > N)
- Subtract mean
- Calculate eigenvalues of data covariance
- Project data onto P < N PCs (must flip PC matrix left-right)

$$Dta = Dta - ones(M,1) * mean(Dta)$$

$$\left[PC, Var\right] = eig\left(\mathrm{cov}(Data)\right)$$

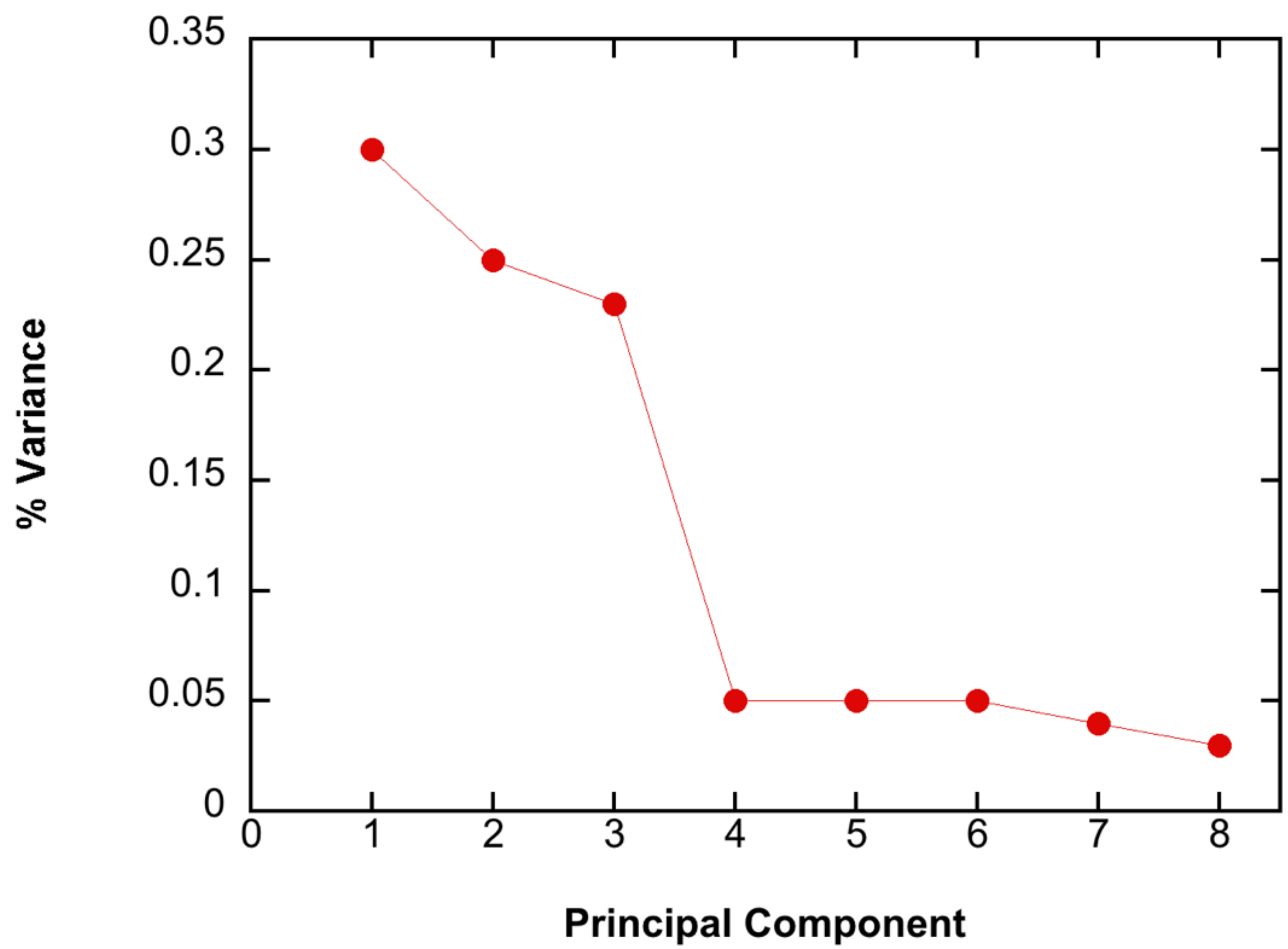
$$PClr = fliplr(PC)$$

$$PCproj = Dta * PClr(:,1:P)$$

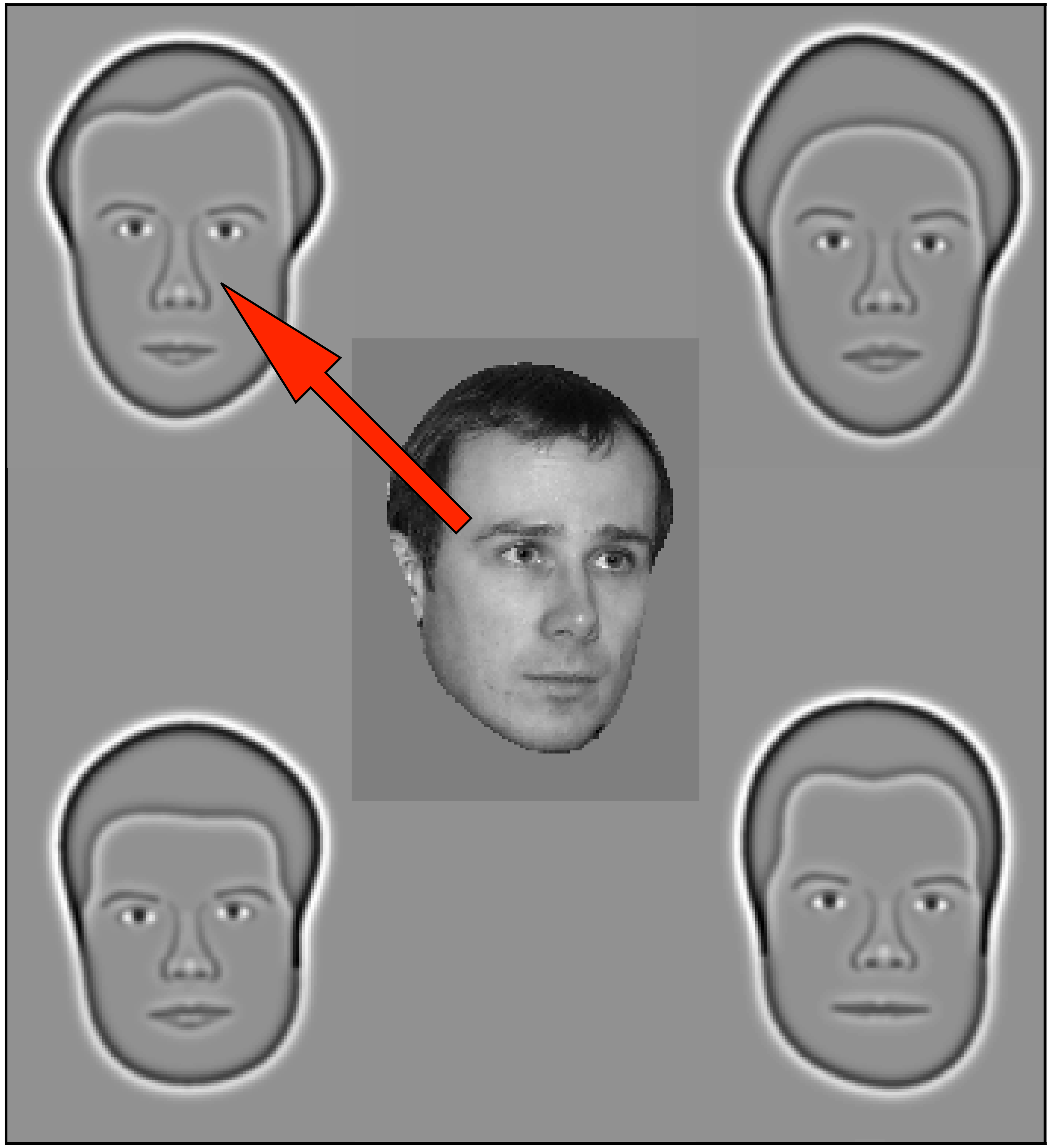
YORK UNIVERSITÉ UNIVERSITY

How Many PCs?

- Decide how much variance is important (eg. 75%, 90%)
- Cut off when variance explained drops below average/PC
- Plot variance per component & look for break (scree plot)



Face Identification



YORK UNIVERSITÉ UNIVERSITY

Centre for Vision Research

Synthetic Faces

Face geometry represented by 39 measurements

Synthetic Face PCA

- Each face: 39 measurements

10 components:
90% of variance

PC#	Variance	Cumulative
1.0000	0.2345	0.2345
2.0000	0.2192	0.4537
3.0000	0.1013	0.5549
4.0000	0.0818	0.6368
5.0000	0.0654	0.7022
6.0000	0.0637	0.7658
7.0000	0.0465	0.8123
8.0000	0.0384	0.8507
9.0000	0.0292	0.8799
10.0000	0.0211	0.9010
11.0000	0.0198	0.9208
12.0000	0.0151	0.9359
13.0000	0.0125	0.9484

YORK U UNIVERSITÉ UNIVERSITY

Remaining 29: noise or irrelevant

Summary

- PCA reveals data structure determined by covariance
- Calculation for N measurements, M samples (M > N):
  - Subtract means from measurements
  - Data covariance matrix CV
  - [Vector, EV] = eig(CV)
- Dimensionality reduction
- May facilitate separating categories (next lecture)
- Can be implemented by neural learning networks
- Nonlinear manifolds in data require embellishments