Signals & Systems

Spring 2019

https://sites.google.com/site/ntusands/

https://ceiba.ntu.edu.tw/1072EE2011_04

Yu-Chiang Frank Wang 王鈺強, Associate Professor

Dept. Electrical Engineering, National Taiwan University

2019/06/13

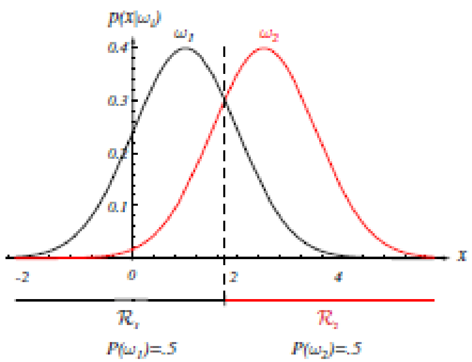
Outline

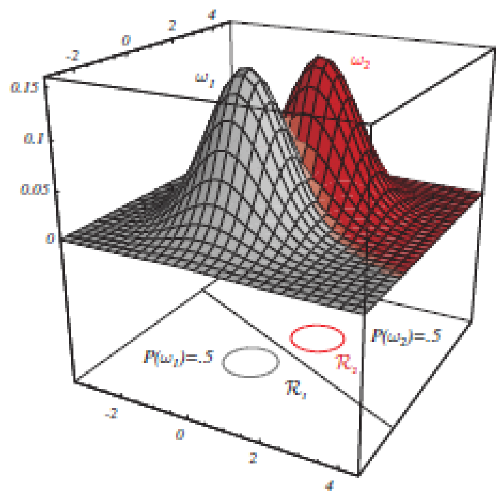
- Ch. 2 LTI Systems
- Ch. 3 FS of Periodic Signals
- Ch. 4 CTFT
- Ch. 5 DTFT
- Ch. 6 Time and Freq. Characterization of Signals and Systems
- Ch. 7 Sampling
- Ch. 9 Laplace Transform
- Ch. 10 z-Transform
- Ch. 8 Comm. Systems

Bonus Lecture:
Machine Learning 101

Machine Learning 101

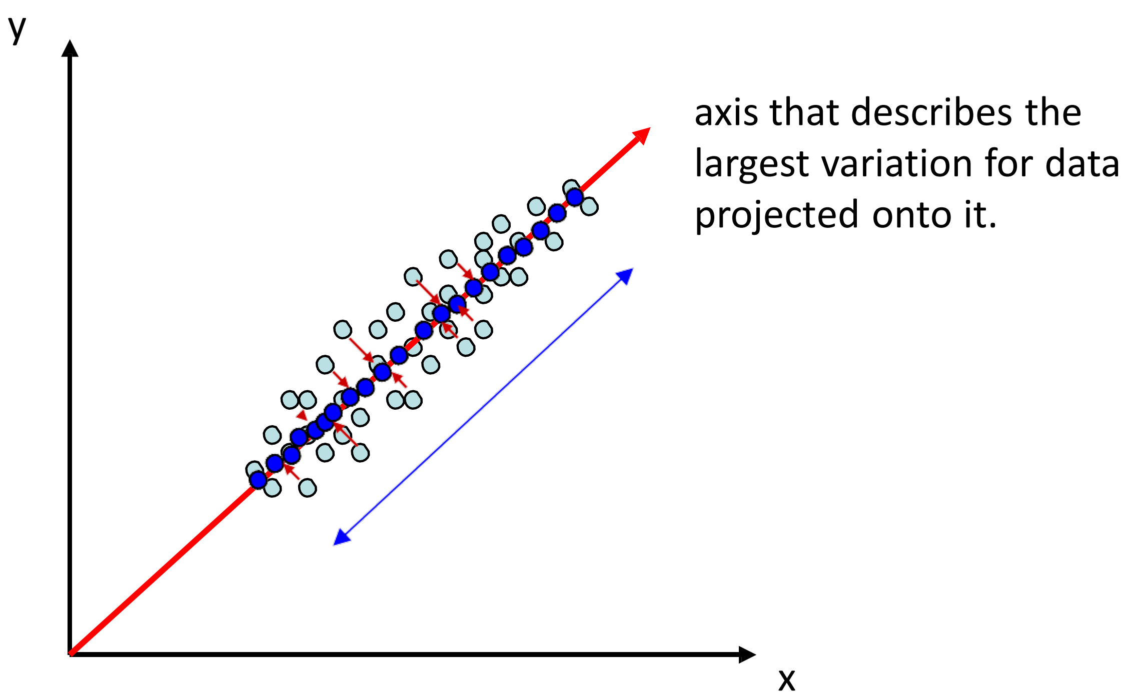
- From Probability to Bayes Decision Rule
- Brief Review of Linear Algebra & Linear System
- Unsupervised vs. Supervised Learning
  - Clustering & Dimension Reduction
  - Training, testing, & validation
  - Linear Classification





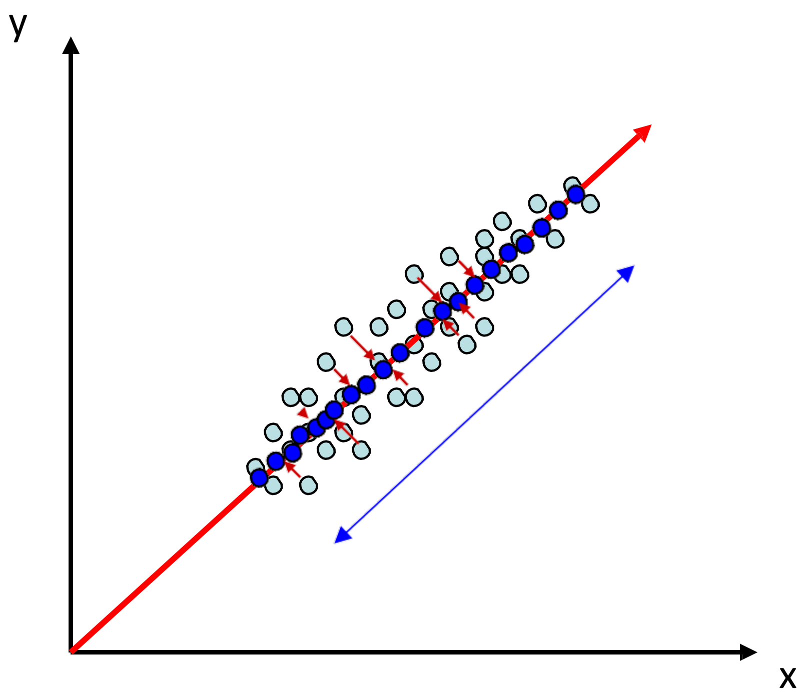
Dimension Reduction

- Principal Component Analysis (PCA)
  - Unsupervised & linear dimension reduction
  - Related to Eigenfaces, etc. feature extraction and classification techniques
  - Still very popular despite of its simplicity and effectiveness.
  - Goal:
    - Determine the projection, so that the variation of projected data is maximized.



Formulation & Derivation for PCA

- Input: a set of instances $\boldsymbol{x}$ without label info
- Output: a projection vector $\boldsymbol{\omega}$ maximizing the variance of the projected data
- In other words, we need to maximize $\mathrm{var}(\boldsymbol{\omega}^T\boldsymbol{x})$ with $\|\boldsymbol{\omega}\| = 1$.



Formulation & Derivation for PCA (cont'd)

- Lagrangian optimization for PCA

Eigenanalysis & PCA

- Find the eigenvectors $\boldsymbol{e}_i$ and the corresponding eigenvalues λ_i
  - The direction $\boldsymbol{e}_i$ captures the variance of λ_i.
  - But, which eigenvectors to use? All of them?
- A d x d covariance matrix contains a maximum of d eigenvector/eigenvalue pairs.
  - Which $\boldsymbol{e}_i$ (and thus λ_i) to consider?
  - Assume N images of size M x M pixels, we have dimension d = M^2.
  - What is the rank of ∑?
  - Thus, at most non-zero eigenvalues can be obtained.

Eigenanalysis & PCA (cont'd)

- Image reconstruction via PCA
  - Expand a signal (e.g., an input image) via eigenvectors as bases
  - With symmetric matrices (e.g., covariance matrix), eigenvectors are orthogonal.
  - They can be regarded as unit basis vectors to span any instance in the d-dim space.

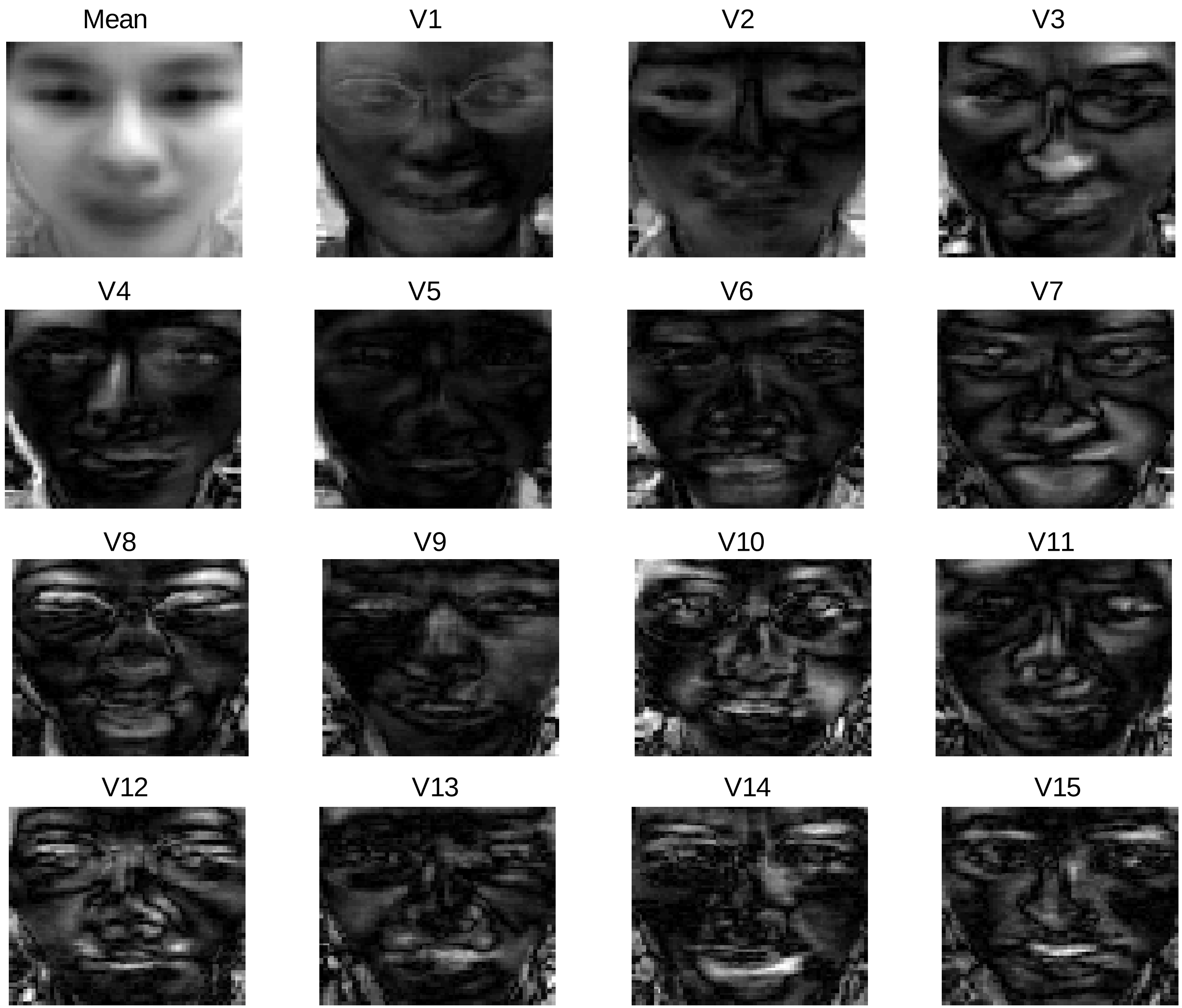
Practical Issues in PCA

- Assume we have N = 100 images of size 200 x 200 pixels (i.e., d = 40000).
- What is the size of the covariance matrix? What's its rank?
- What can we do? Gram Matrix Trick!

Let’s See an Example (CMU AMP Face Database)

- Let’s take 5 face images x 13 people = 65 images, each is of size 64 x 64 = 4096 pixels.
- # of eigenvectors are expected to use for perfectly reconstructing the input = 64.
- Let’s check it out!

What Do the Eigenvectors/Eigenfaces Look Like?



All 64 Eigenvectors, do we need them all?

Use only 1 eigenvector, MSE = 1233

Use 2 eigenvectors, MSE = 1027

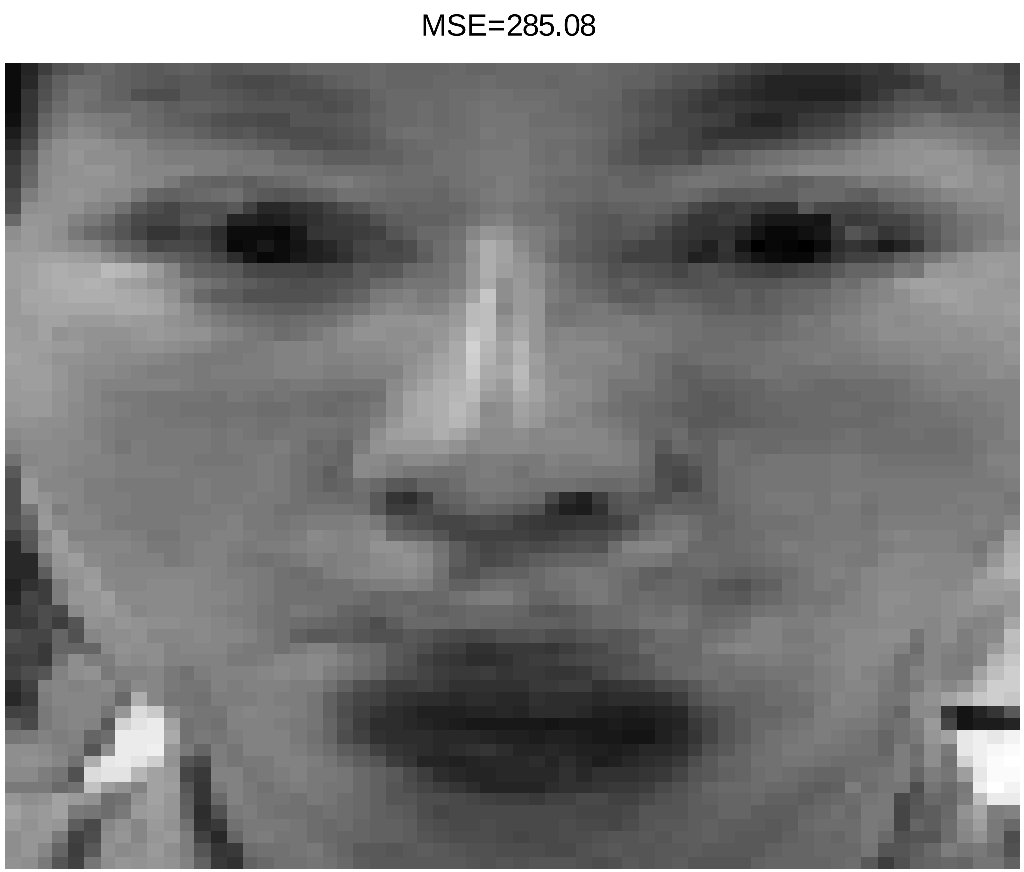
MSE=1027.63

Use 3 eigenvectors, MSE = 758

Use 4 eigenvectors, MSE = 634

MSE=634.54

Use 8 eigenvectors, MSE = 285



With 20 eigenvectors, MSE = 87

MSE=87.93

With 30 eigenvectors, MSE = 20

MSE=20.55

With 50 eigenvectors, MSE = 2.14

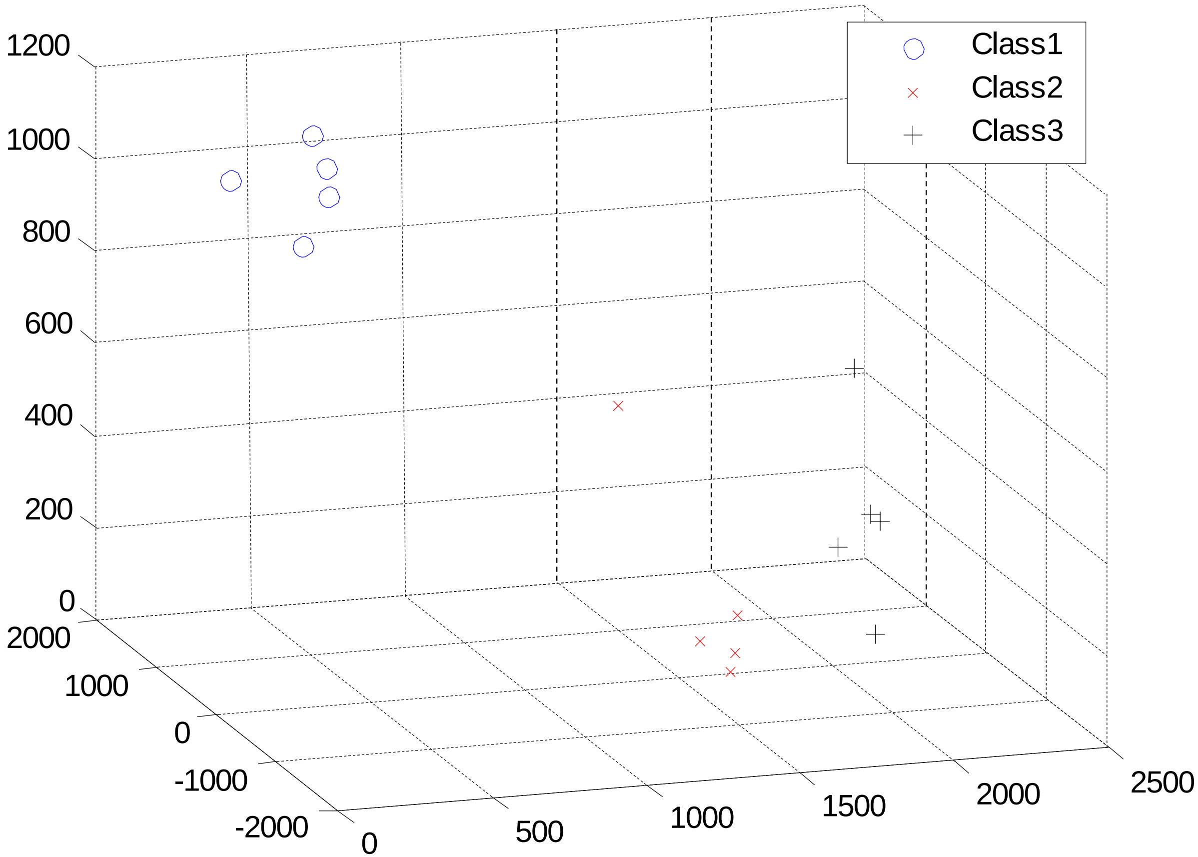
MSE=2.14

With 60 eigenvectors, MSE = 0.06

All 64 eigenvectors, MSE = 0

Final Remarks

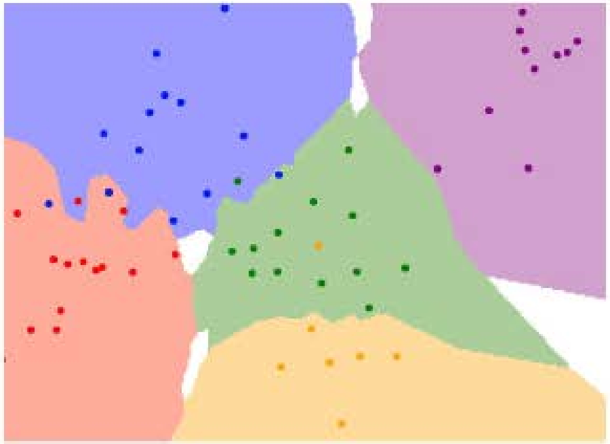
- Linear & unsupervised dimension reduction
- PCA can be applied as a feature extraction/preprocessing technique.
    - E.g,, Use the top 3 eigenvectors to project data into a 3D space for classification.



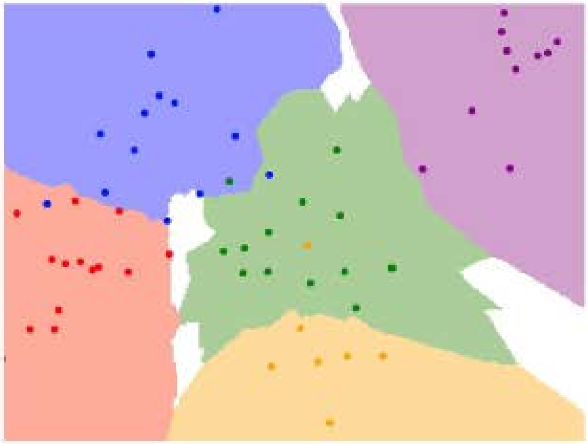
Final Remarks (cont'd)

- How do we classify? For example…
  - Given a test face input, project into the same 3D space (by the same 3 eigenvectors).
  - The resulting vector in the 3D space is the feature for this test input.
  - We can do a simple Nearest Neighbor (NN) classification with Euclidean distance, which calculates the distance to all the projected training data in this space.
  - If NN, then the label of the closest training instance determines the classification output.
  - If k-nearest neighbors (k-NN), then k-nearest neighbors need to vote for the decision.

k = 1



k = 3



k = 5

Demo available at http://vision.stanford.edu/teaching/cs231n-demos/knn/

Image credit: Stanford CS231n

Final Remarks (cont'd)

- If labels for each data is provided → Linear Discriminant Analysis (LDA)
  - LDA is also known as Fisher's discriminant analysis.
  - Eigenface vs. Fisherface (IEEE Trans. PAMI 1997)
- If linear DR is not sufficient, and non-linear DR is of interest…
  - Isomap, locally linear embedding (LLE), etc.
  - **t-distributed stochastic neighbor embedding (t-SNE)** (by G. Hinton & L. van der Maaten)