Some Interesting Problems in Pattern Recognition

JEN-MEI CHANG

Department of Mathematics and Statistics
California State University, Long Beach
jchang9@csulb.edu

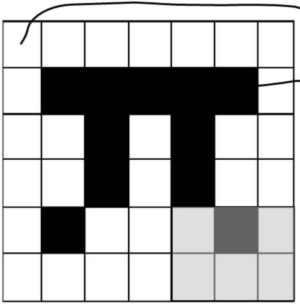
Claremont Math-in-Industry Workshop

Outline

1. Face Recognition - PCA
2. Bankruptcy Prediction - LDA
3. Cocktail Party Problem - BSS
4. Speech Recognition - DFT
5. Handwritten Digit Classification - Tangent
6. Traveling Salesman Problem - Unsupervised Clustering
7. A Challenge Problem For You

Data Matrix

An r-by-c gray scale digital image corresponds to an r-by-c matrix where each entry enumerates one of the 256 possible gray levels of the corresponding pixel.



$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Data Vector

Realize the data matrix by its columns and concatenate columns into a single column vector.

$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\longrightarrow
\begin{bmatrix}
1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \hline
1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ \hline
\vdots \\ \hline
1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1
\end{bmatrix}
$$

$\boxed{\text{IMAGE} \rightarrow \text{MATRIX} \rightarrow \text{VECTOR}}$

Classification/Recognition Problem

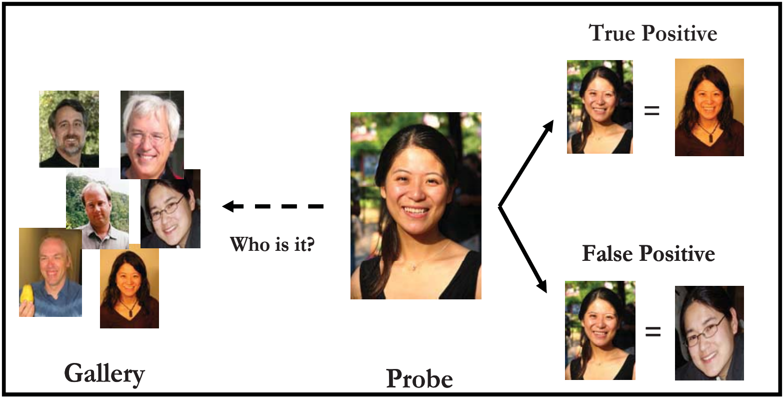
1.5 **Preprocessing**
Geometric normalization,
Feature extraction, etc

1. **Data collection**
Database creation
a.k.a. **Gallery**

2. **Present novel data**
a.k.a. **Probe**

3. **Classification**
Assign label to probe
and assess accuracy

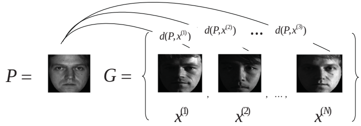
Face Recognition Problem



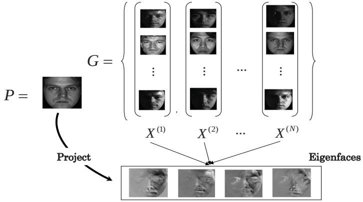
Architectures

Historically

- single-to-single

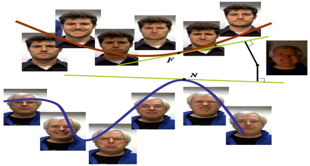


- single-to-many

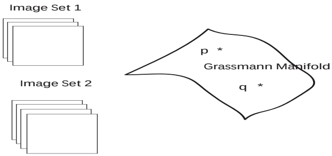


Currently

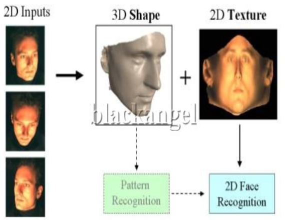
- subspace-to-subspace



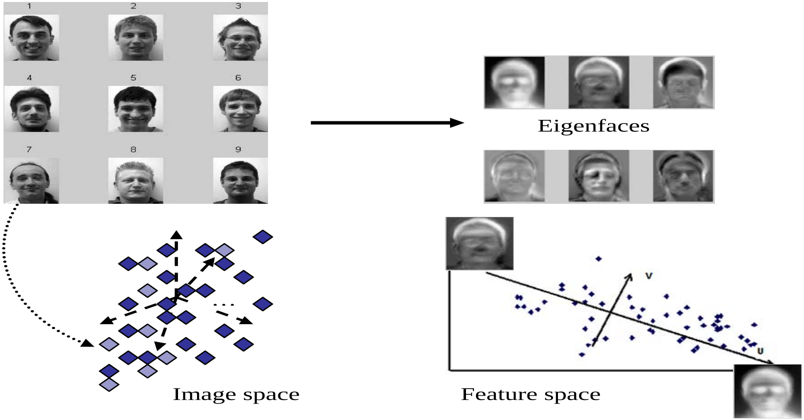
- many-to-many



Commercial Applications

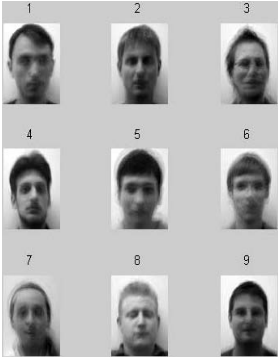


A Possible Mathematical Approach

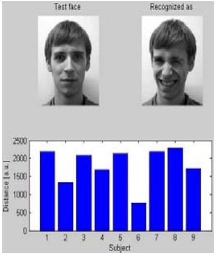


(Adapted from Vladimir Bondarenko at University of Konstanz, ST; http://www.inf.uni-konstanz.de/cgip/lehre/na_08/Lab2/5_FaceRecognition/html/myFaceRecognition.html)

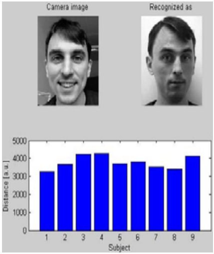
A Possible Mathematical Approach



Database in feature space

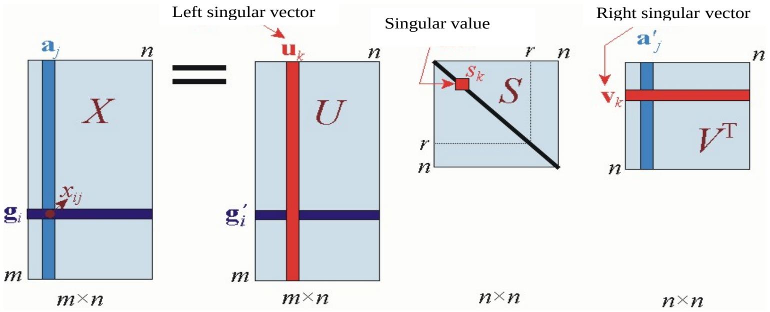


Classification in feature space



(Adapted from Vladimir Bondarenko at University of Konstanz, ST; http://www.inf.uni-konstanz.de/cgip/lehre/na_08/Lab2/5_FaceRecognition/html/myFaceRecognition.html)

$\longrightarrow X = USV^T$



- $V = [v_1, \ldots, v_r, v_{r+1}, \ldots, v_n]$ is orthogonal with v_i's eigenvectors of $X^T X$.
- $S = \text{diag}(s_1, \ldots, s_r, 0, \ldots, 0)$ is diagonal with s_i's square root of eigenvalues of $X^T X$.
- $U = S^{-1} XV$ is orthogonal with u_i's eigenvectors of XX^T.

Bankruptcy prediction is the art of predicting bankruptcy and various measures of financial distress of public firms. It is a vast area of finance and accounting research. The importance of the area is due in part to the relevance for creditors and investors in evaluating the likelihood that a firm may go bankrupt[1].

- Form a feature vector for each firm.
- Two-class classification problem.

[1] adapted from Wikipedia

Bankruptcy prediction is the art of predicting bankruptcy and various measures of financial distress of public firms. It is a vast area of finance and accounting research. The importance of the area is due in part to the relevance for creditors and investors in evaluating the likelihood that a firm may go bankrupt[1].

- Form a feature vector for each firm.
- Two-class classification problem.

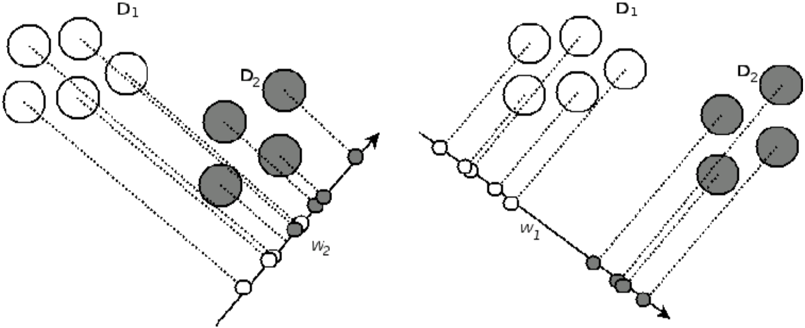
[1] adapted from Wikipedia

Bankruptcy prediction is the art of predicting bankruptcy and various measures of financial distress of public firms. It is a vast area of finance and accounting research. The importance of the area is due in part to the relevance for creditors and investors in evaluating the likelihood that a firm may go bankrupt[1].

- Form a feature vector for each firm.
- Two-class classification problem.

[1] adapted from Wikipedia

A Possible Mathematical Approach

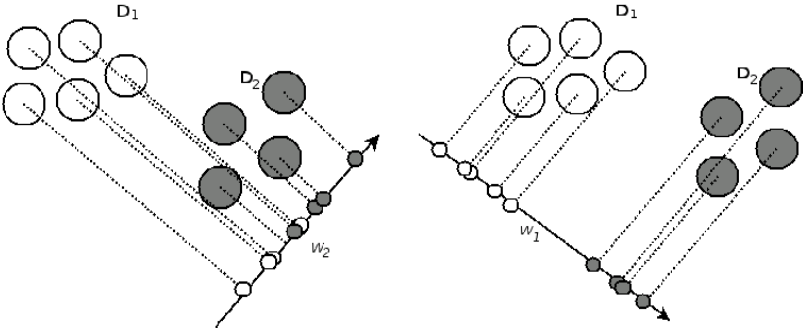


Bad projection

Good projection

Question: Characteristics of a GOOD projection?

A Possible Mathematical Approach

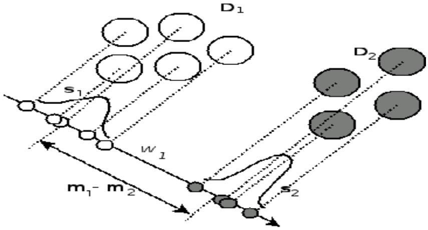


Bad projection   Good projection

Question: Characteristics of a GOOD projection?

Two-Class LDA

$$m_1 = \frac{1}{n_1} \sum_{x \in D_1} w^T x, \quad m_2 = \frac{1}{n_2} \sum_{y \in D_2} w^T y$$



Look for a projection w that

- maximizes (inter-class) distance in the projected space,
- and minimizes the (intra-class) distances in the projected space.

Two-Class LDA

Namely, we desire a w^* such that

$$w^* = \arg\max_w \frac{(m_1 - m_2)^2}{S_1 + S_2},$$

where $S_1 = \sum_{x \in D_1} (w^T x - m_1)^2$ and $S_2 = \sum_{y \in D_2} (w^T y - m_2)^2$.

Alternatively, (with scatter matrices)

$$w^* = \arg\max_w \frac{w^T S_B w}{w^T S_W w}, \tag{1}$$

with $S_W = \sum_{i=1}^{2} \sum_{x \in D_i} (x - \mathbf{m}_i)(x - \mathbf{m}_i)^T$, $S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$.

Two-Class LDA

Namely, we desire a w^* such that

$$w^* = \arg\max_{w} \frac{(m_1 - m_2)^2}{S_1 + S_2},$$

where $S_1 = \sum_{x \in D_1} (w^T x - m_1)^2$ and $S_2 = \sum_{y \in D_2} (w^T y - m_2)^2$.

Alternatively, (with scatter matrices)

$$w^* = \arg\max_{w} \frac{w^T S_B w}{w^T S_W w}, \tag{1}$$

with $S_W = \sum_{i=1}^{2} \sum_{x \in D_i} (x - \mathbf{m}_i)(x - \mathbf{m_i})^T$, $S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$.

LDA

The criterion in Equation (1) is commonly known as the generalized Rayleigh quotient, whose solution can be found via the generalized eigenvalue problem

$$S_B w = \lambda S_W w.$$

LDA for multi-class follows similarly.

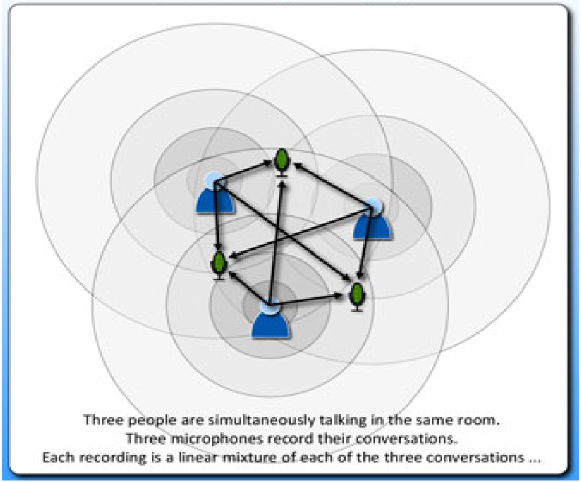
LDA

The criterion in Equation (1) is commonly known as the generalized Rayleigh quotient, whose solution can be found via the generalized eigenvalue problem

$$S_B w = \lambda S_W w.$$

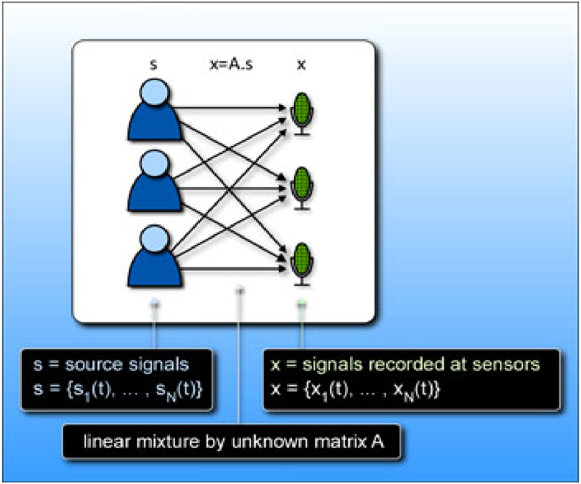
LDA for multi-class follows similarly.

Cocktail Party Problem



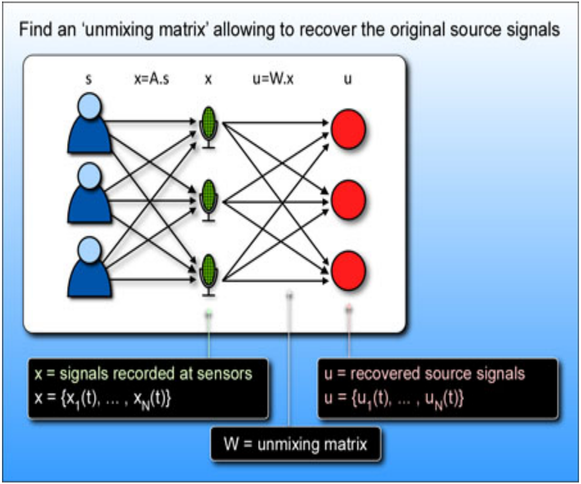
(adapted from André Mouraux)

Cocktail Party Problem



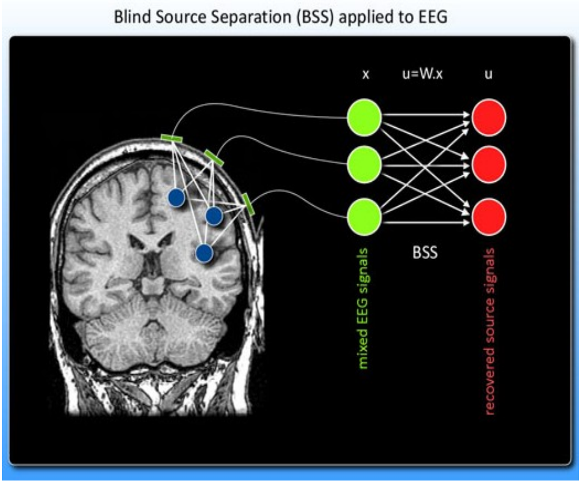
(adapted from André Mouraux)

Cocktail Party Problem



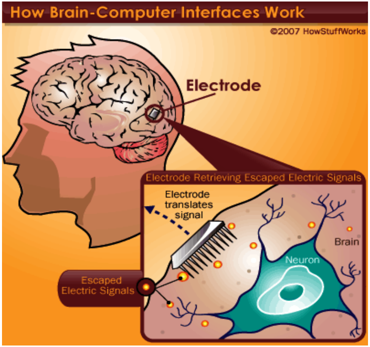
(adapted from André Mouraux)

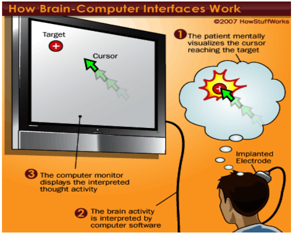
A Similar Problem: EEG

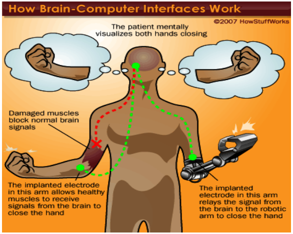


(adapted from André Mouraux)

Commercial Applications







(http://computer.howstuffworks.com/brain-computer-interface1.htm)

A Possible Mathematical Approach

- Decompose observed data into its *noise* and *signal* components:

$$\mathbf{x}^{(\mu)} = \mathbf{s}^{(\mu)} + \mathbf{n}^{(\mu)},$$

or, in terms of data matrices,

$$X = S + N. \quad (S = \text{ signal }, N = \text{ noise })$$

- The optimal first basis vector, ϕ, is taken as a superposition of the data, i.e.,

$$\phi = \psi_1 \mathbf{x}^{(1)} + \cdots + \psi_P \mathbf{x}^{(P)} = X\psi.$$

- May decompose ϕ into signal and noise components

$$\phi = \phi_{\mathbf{n}} + \phi_{\mathbf{s}},$$

where $\phi_{\mathbf{s}} = S\psi$ and $\phi_{\mathbf{n}} = N\psi$.

MNF/BBS

- The basis vector ϕ is said to have maximum noise fraction (MNF) if the ratio

$$D(\phi) = \frac{\phi_{\mathbf{n}}^T \phi_{\mathbf{n}}}{\phi^T \phi}$$

  is a maximum.

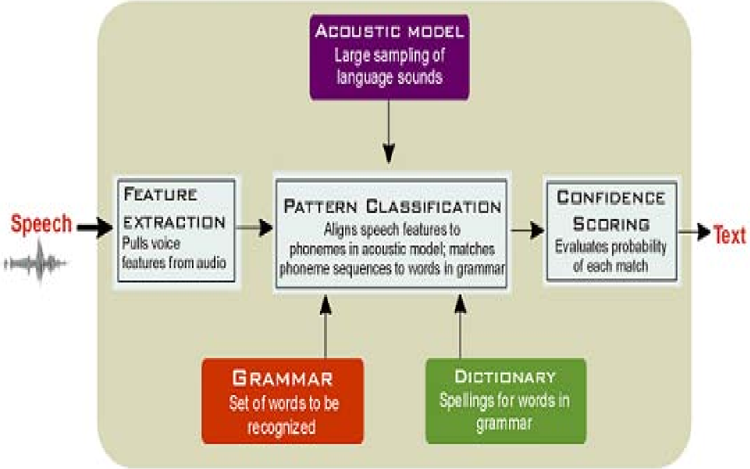
- A steepest descent method yields the *symmetric definite generalized eigenproblem*

$$N^T N\psi = \mu^2 X^T X\psi.$$

  This problem may be solved without actually forming the product matrices $N^T N$ and $X^T X$, using the generalized SVD (gsvd).

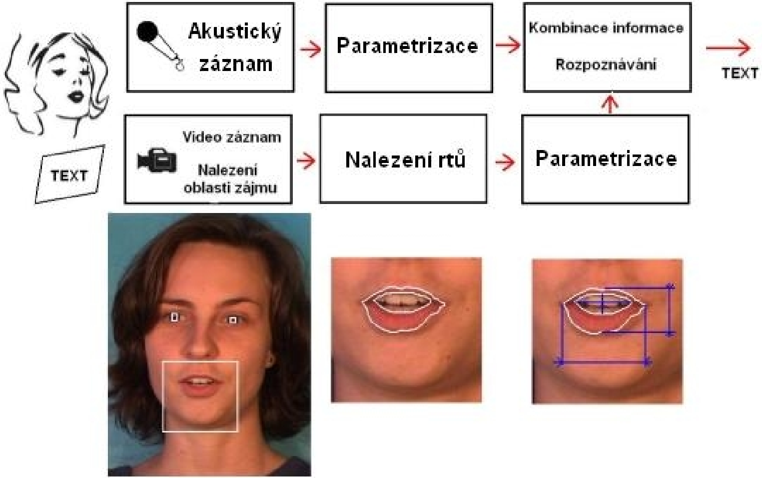
- Note that the same orthonormal basis vector ϕ optimizes the signal-to-noise ratio. And this technique is called **Blind Source Separation (BSS).**

Audio



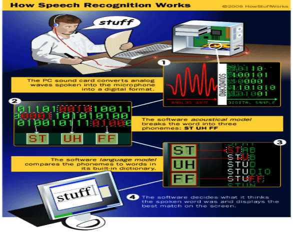
(adapted from AT&T Lab Inc. - http://www.research.att.com/viewProject.cfm?prjID=49)

Audio-Visual



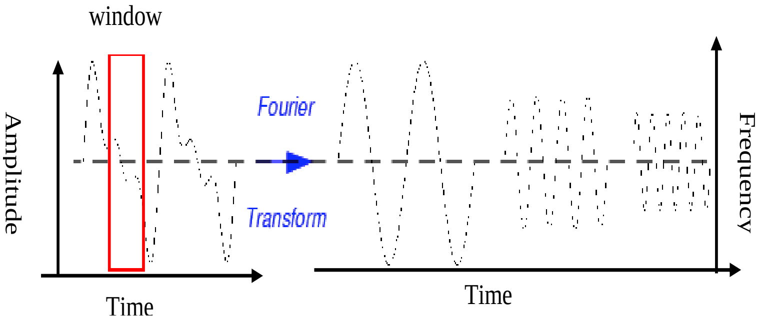
(adapted from Project MUSSLAP - http://musslap.zcu.cz/img/audiovizualni-zpracovani-reci/schema.jpg)

Commercial Applications



A Possible Mathematical Approach

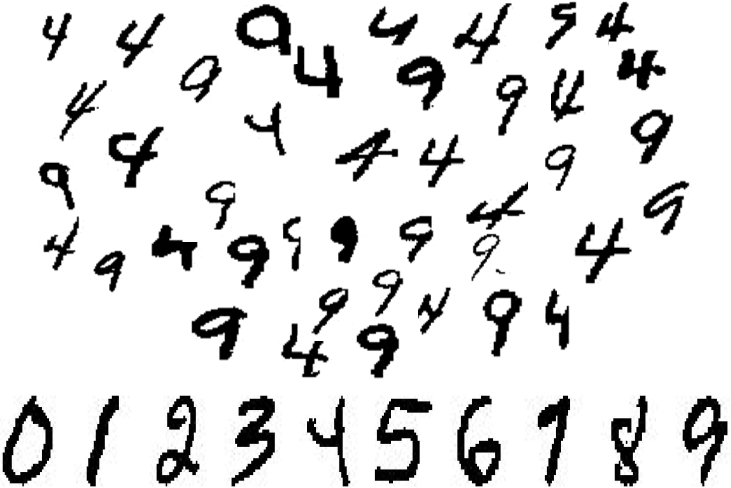
- Continuous $F(\omega) = \int_{-\infty}^{\infty} f(t)e^{i\omega t}\,dt$
- Discrete $f(\omega) = \sum_{k\in\mathbb{Z}} c_k e^{ik\omega}$



How

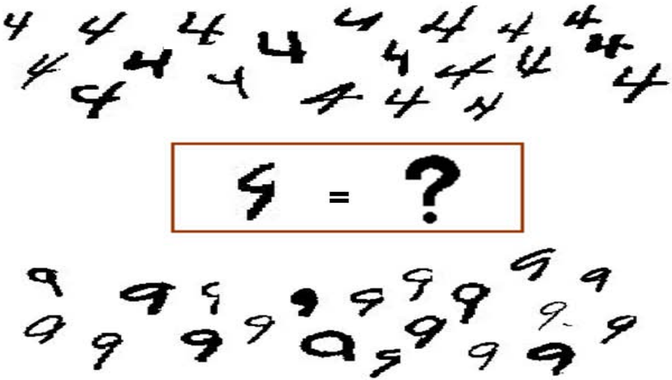
- Fourier analysis is applied to speech waveform in order to discover what frequencies are present at any given moment in the speech signal with time on the horizontal axis and frequency on the vertical.
- The speech recognizer has a database of several thousand such graphs (called a codebook) that identify different types of sounds the human voice can make.
- The sound is “identified” by matching it to its closest entry in the codebook.

Handwritten Digits



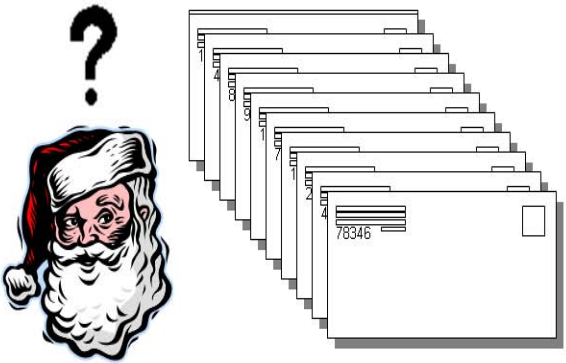
Handwritten Digit Classification

How do we tell whether a new digit is a 4 or a 9?



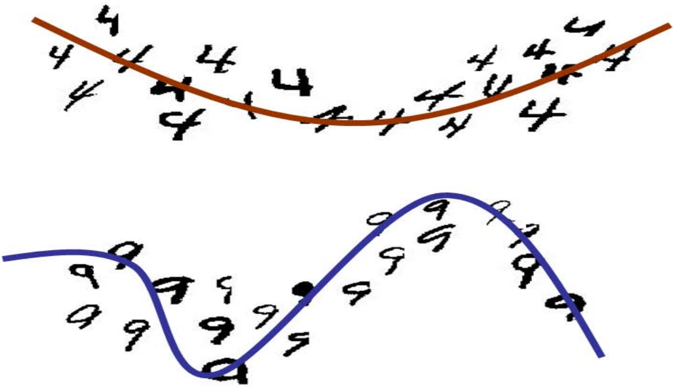
Commercial Applications

Santa thought to himself, "only if these mails can go to the right place according to their zip code".



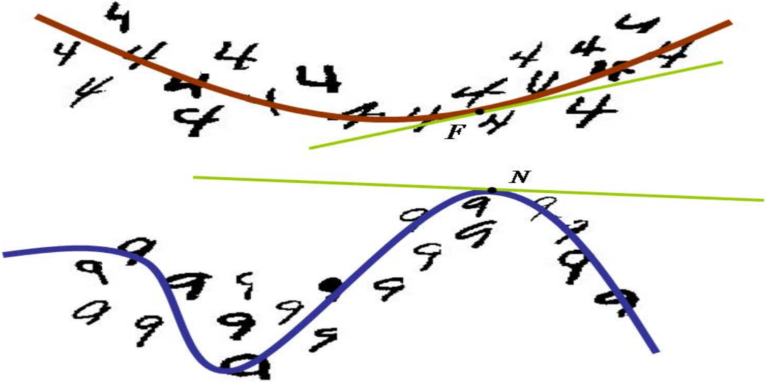
A Possible Mathematical Approach

Imagine a high-D surface (red curve) where all 4's live on and a high-D surface (blue curve) where all 9's live on.

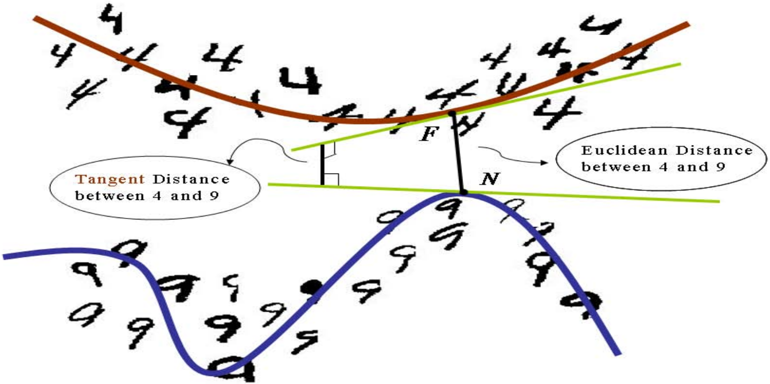


Manifold Learning

Create a Tangent Space of the 4's at F and create a Tangent Space of the 9's at N.

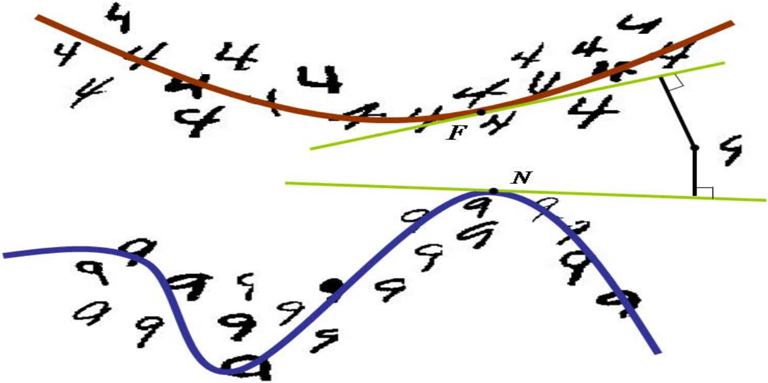


Which Distance?



Classification

So, is it a 4 or a 9?



TSP

Given a list of cities and their pairwise distances, the goal is to find a shortest route that visits each city exactly once.

(Adapted from Wikipedia: http://en.wikipedia.org/wiki/Travelling_salesman_problem)

A Possible Mathematical Approach

The SOFM (Kohonen's Self-Organizing Feature Map) Algorithm

Given a data set $X = \{\mathbf{x}^{(\mu)}\}$,

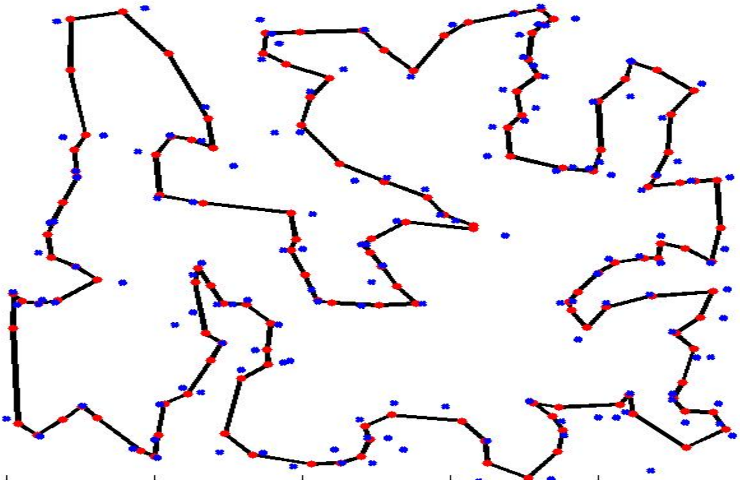
1. Initialize a set of center vectors $\{c_i\}$, $i \in \mathcal{I}$.
2. Present $\mathbf{x}^{(\mu)}$ to the network.
3. Determine the winning center vector $c_{i'}$.
4. Update *all* the center vectors using

$$c_i^{n+1} = c_i^n + \epsilon h(d(i, i'))(\mathbf{x}^{(\mu)} - c_{i'}).$$

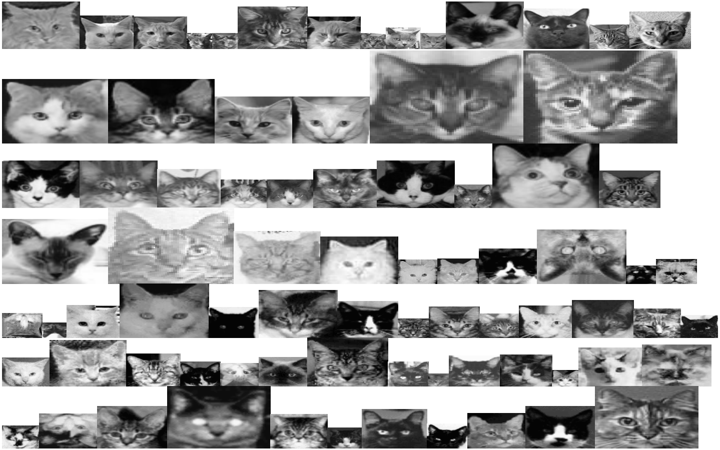
5. Repeat

$d(i, i')$ is a (topological) metric on the indices. Typically, $h(x)$ is taken to be a Gaussian, i.e., $h(x) = exp(-x^2/r^2)$.

SOFM Result on 150 Cities

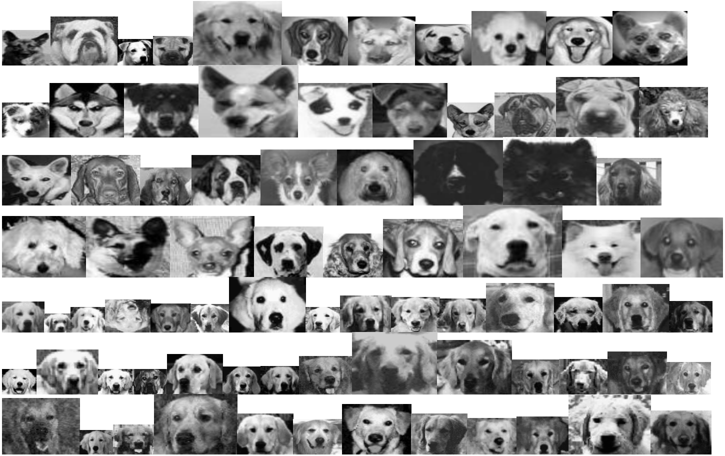


Example Cats



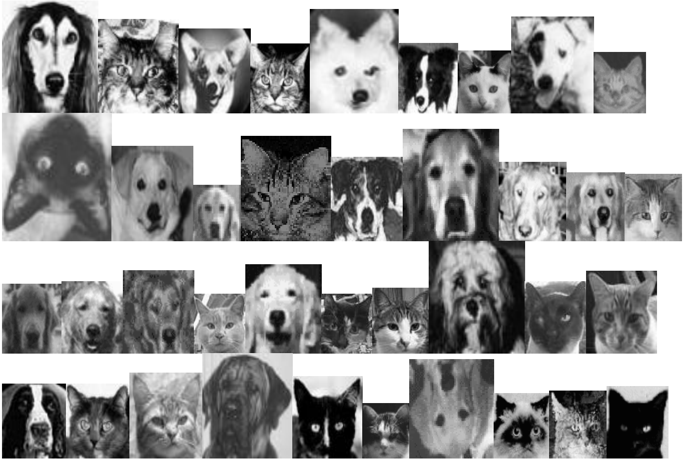
Courtesy of Dr. J.R. Beveridge in the Department of Computer Science at CSU.

Example Dogs



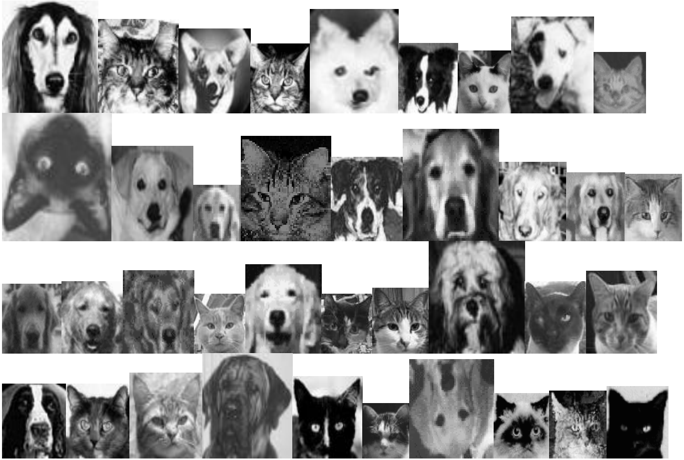
Courtesy of Dr. J.R. Beveridge in the Department of Computer Science at CSU.

Probe Set



Why is the problem of pattern recognition challenging?

Probe Set



Why is the problem of pattern recognition challenging?

Get Started!

- This talk is available via
  http://www.csulb.edu/∼jchang9/files/patternRecTalk_Claremont_JMC.pdf
- The problem data set can be accessed from
  http://www.csulb.edu/∼jchang9/files/PatternRecData.mat