

Outline

- ...
- Feature extraction
- Feature selection / Dim. reduction
- Classification
- ...

Carnegie Mellon

Outline

- ...
- Feature extraction
 - shape
 - texture
- Feature selection / Dim. reduction
- Classification
- ...

Carnegie Mellon

How to extract features?

- Eg.:

Carnegie Mellon

How to extract features?

- Eg.:

Carnegie Mellon

Images - shapes

- distance function: Euclidean, on the area, perimeter, and 20 'moments' [QBIC, '95]
- Q: other 'features' / distance functions?

Carnegie Mellon


Images - shapes

- (A1: turning angle)
- A2: wavelets
- A3: morphology: dilations/erosions
- ...

Carnegie Mellon

Wavelets - example

<http://grail.cs.washington.edu/projects/query/>
Wavelets achieve *great* compression:

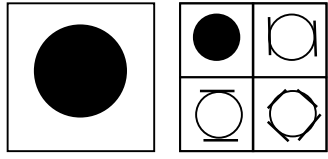


20 100 400 16,000
coefficients

Carnegie Mellon

Wavelets - intuition

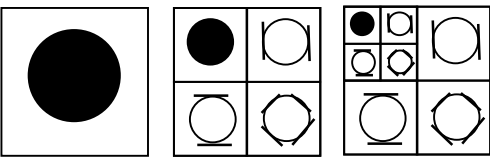
- Edges (horizontal; vertical; diagonal)



Carnegie Mellon

Wavelets - intuition

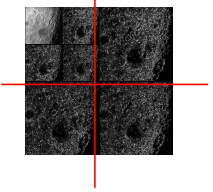
- Edges (horizontal; vertical; diagonal)
- recurse



Carnegie Mellon

Wavelets - intuition

- Edges (horizontal; vertical; diagonal)
- <http://www331.jpl.nasa.gov/public/wave.html>



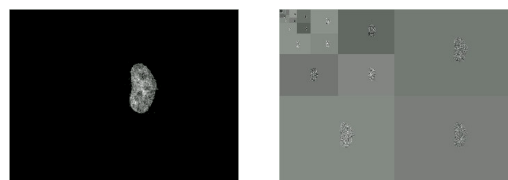
Carnegie Mellon

Wavelets

- Many wavelet basis:
 - Haar
 - Daubechies (-4, -6, -20)
 - Gabor
 - ...

Carnegie Mellon

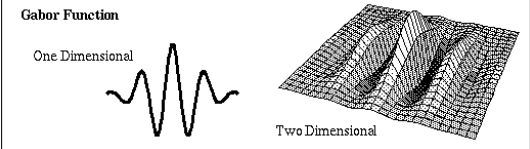
Daubechies D4 decomposition



Original image Wavelet Transformation

Carnegie Mellon

Gabor Function



Gabor Function


One Dimensional

Two Dimensional

We can extend the function to generate Gabor filters by rotating and dilating

Carnegie Mellon


Feature Calculation



- Preprocessing
 - Background subtraction and thresholding
 - Translation and rotation
- Wavelet transformation
 - The Daubechies 4 wavelet
 - 10th level decomposition
 - The average energy of the three high-frequency components

Carnegie Mellon

Feature Calculation



- Preprocessing
- 30 Gabor filters were generated using five different scales and six different orientations
- Convolve an input image with a Gabor filter
- Take the mean and standard deviation of the convolved image
- 60 Gabor texture features

Carnegie Mellon

Wavelets:

- Extremely useful
- Excellent compression / feature extraction, for natural images
- fast to compute ($O(N)$)

Carnegie Mellon

Images - shapes

- (A1: turning angle)
- A2: wavelets
- A3: morphology: dilations/erosions
- ...




Carnegie Mellon

Other shape features

- Morphology (dilations, erosions, openings, closings) [Korn+, VLDB96]

shape (B/W)



"structuring element"


R=1

Carnegie Mellon

Other shape features

- Morphology (dilations, erosions, openings, closings) [Korn+, VLDB96]

shape "structuring element"




R=0.5 ●
R=1 ●
R=2 ●

Carnegie Mellon

Other shape features

- Morphology (dilations, erosions, openings, closings) [Korn+, VLDB96]

shape "structuring element"

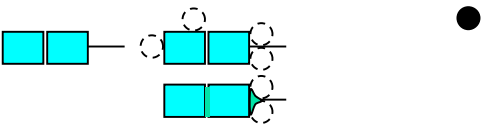


R=0.5 ●
R=1 ●
R=2 ●

Carnegie Mellon

Morphology: closing

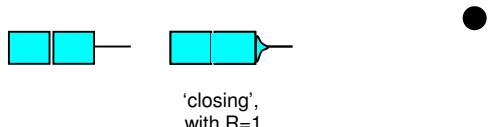
- fill in small gaps
- **very similar** to 'alpha contours'



Carnegie Mellon

Morphology: closing

- fill in small gaps

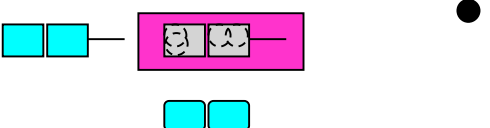


'closing', with R=1

Carnegie Mellon

Morphology: opening

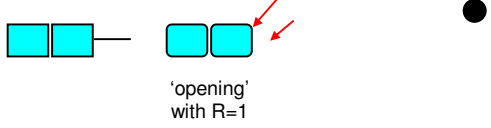
- 'closing', for the complement =
- trim small extremities



Carnegie Mellon

Morphology: opening

- 'closing', for the complement =
- trim small extremities

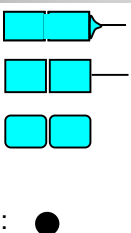


'opening' with R=1

Carnegie Mellon

Morphology

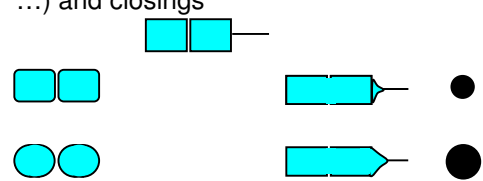
- Closing: fills in gaps
- Opening: trims extremities
- All wrt a structuring element: ●



Carnegie Mellon

Morphology

- Features: areas of openings ($R=1, 2, \dots$) and closings



Carnegie Mellon

Morphology

- resulting areas: 'pattern spectrum'
 - translation (and rotation) independent
- As described: on b/w images
 - can be extended to grayscale ones (eg., by thresholding)

Carnegie Mellon

Conclusions

- Shape: wavelets; math. morphology
- texture: wavelets; Haralick texture features

Carnegie Mellon

References

- Faloutsos, C., R. Barber, et al. (July 1994). "Efficient and Effective Querying by Image Content." J. of Intelligent Information Systems 3(3/4): 231-262.
- Faloutsos, C. and K.-I. D. Lin (May 1995). *FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets*. Proc. of ACM-SIGMOD, San Jose, CA.
- Faloutsos, C., M. Ranganathan, et al. (May 25-27, 1994). *Fast Subsequence Matching in Time-Series Databases*. Proc. ACM SIGMOD, Minneapolis, MN.

Carnegie Mellon

References

- Christos Faloutsos, *Searching Multimedia Databases by Content*, Kluwer 1996

Carnegie Mellon

References

- Flickner, M., H. Sawhney, et al. (Sept. 1995). "Query by Image and Video Content: The QBIC System." IEEE Computer 28(9): 23-32.
- Goldin, D. Q. and P. C. Kanellakis (Sept. 19-22, 1995). *On Similarity Queries for Time-Series Data: Constraint Specification and Implementation (CP95)*, Cassis, France.

Carnegie Mellon

References

- Charles E. Jacobs, Adam Finkelstein, and David H. Salesin. *Fast Multiresolution Image Querying* SIGGRAPH '95, pages 277-286. ACM, New York, 1995.
- Flip Korn, Nikolaos Sidiropoulos, Christos Faloutsos, Eliot Siegel, Zenon Protopapas: *Fast Nearest Neighbor Search in Medical Image Databases*. VLDB 1996: 215-226

Carnegie Mellon

Outline

- ...
- Feature extraction
- Feature selection / Dim. reduction
- Classification
- ...

Carnegie Mellon

Outline

- ...
- Feature selection / Dim. reduction
 - PCA
 - ICA
 - Fractal Dim. reduction
 - variants
- ...

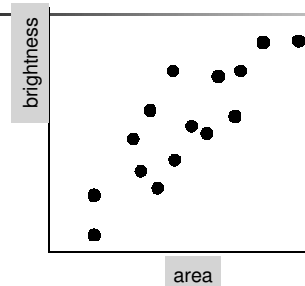
Carnegie Mellon

Feature Reduction

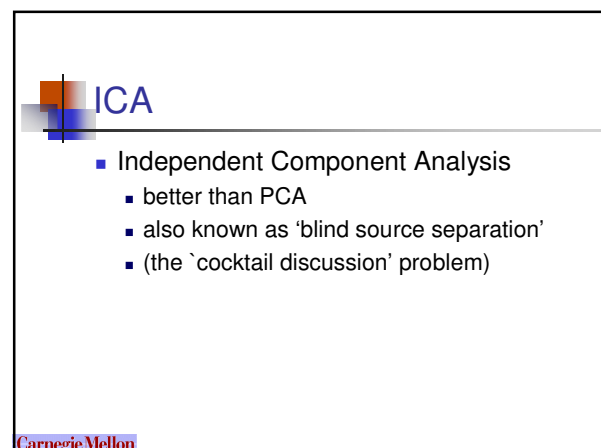
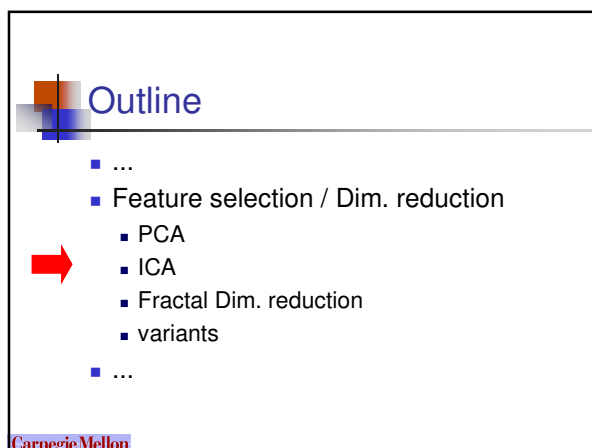
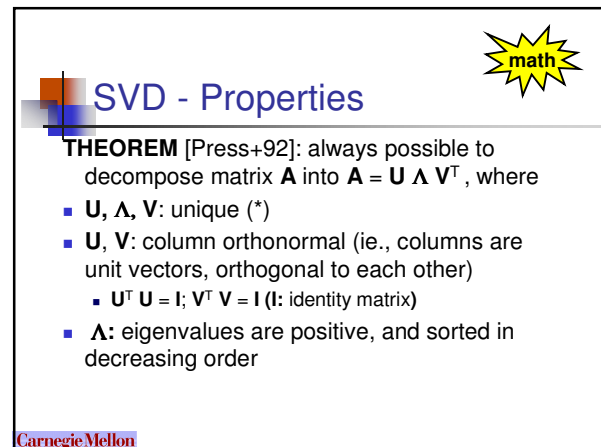
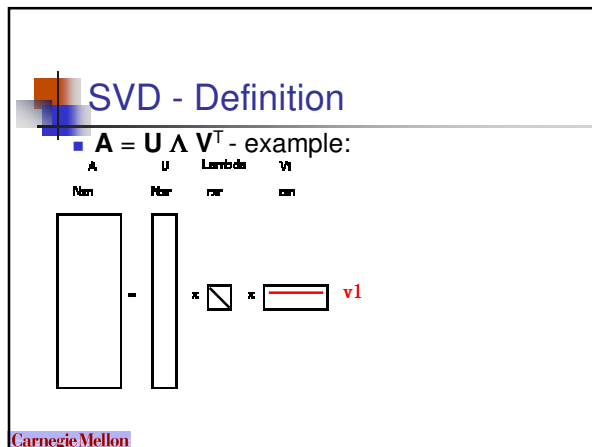
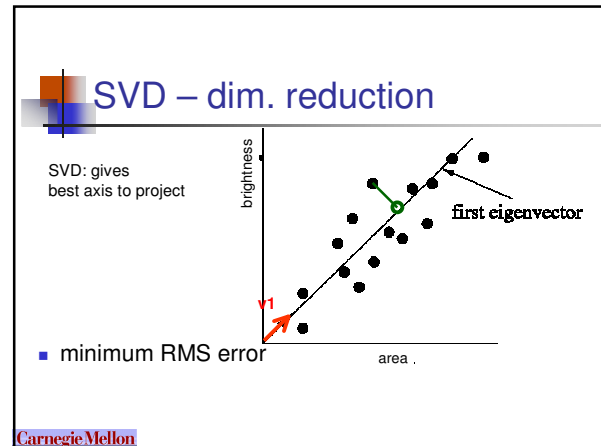
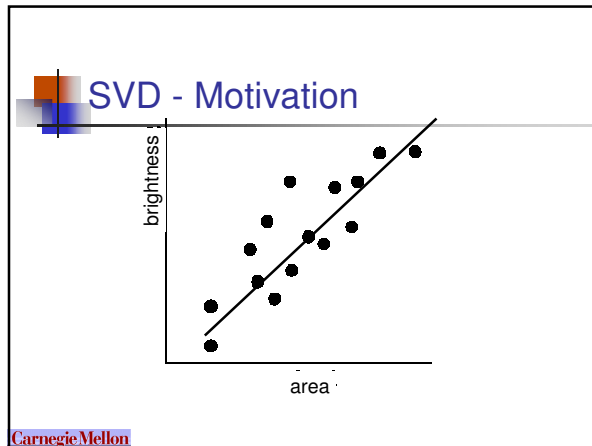
- Remove non-discriminative features
- Remove redundant features
- Benefits :
 - Speed
 - Accuracy
 - Multimedia indexing

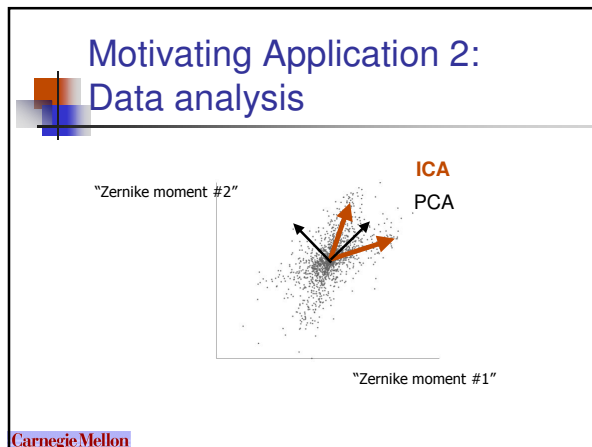
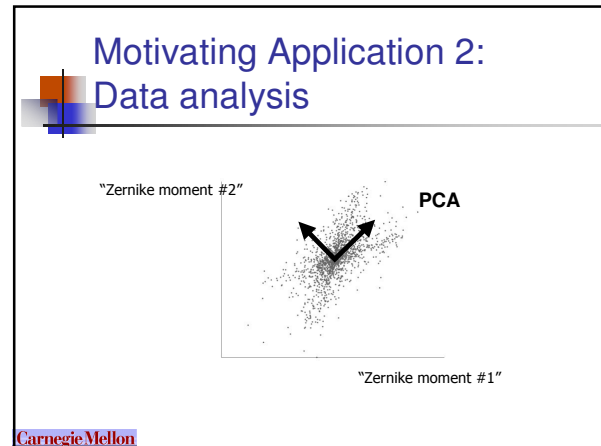
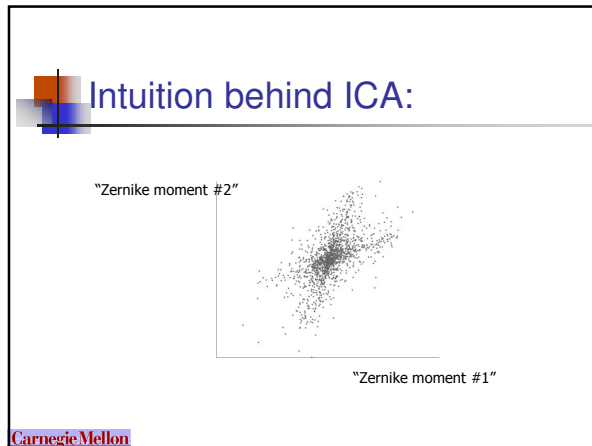
Carnegie Mellon

SVD - Motivation



Carnegie Mellon






- ### Conclusions for ICA
- Better than PCA
 - Actually, uses PCA as a first step!
- Carnegie Mellon


- ### Outline
- ...
 - Feature selection / Dim. reduction
 - PCA
 - ICA
 - Fractal Dim. reduction
 - variants
 - ...
- Carnegie Mellon

- ### Fractal Dimensionality Reduction
1. Calculate the **fractal dimensionality** of the training data.
 2. Forward-Backward select features according to their impact on the fractal dimensionality of the whole data.
- Carnegie Mellon

Dim. reduction 

Spot and drop attributes with strong (non-)linear correlations
Q: how do we do that?

Carnegie Mellon

Dim. reduction - w/ fractals 

(a) Quarter-circle (b) Line (c) Spike


not informative

Carnegie Mellon

Dim. reduction

Spot and drop attributes with strong (non-)linear correlations
Q: how do we do that?
A: compute the **intrinsic** ('**fractal**') dimensionality ~ degrees-of-freedom

Carnegie Mellon

Dim. reduction - w/ fractals 


global FD=1

PFD=1

(c) Spike

PFD=0

Carnegie Mellon

Dim. reduction - w/ fractals 


global FD=1

PFD=1

(b) Line

PFD=1

Carnegie Mellon

Dim. reduction - w/ fractals 

global FD=1

PFD=1


(a) Quarter-circle

PFD=1

Carnegie Mellon

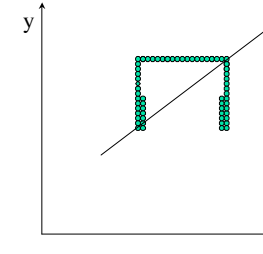
Outline

- ...
- Feature selection / Dim. reduction
 - PCA
 - ICA
 - Fractal Dim. reduction
 - variants
- ...



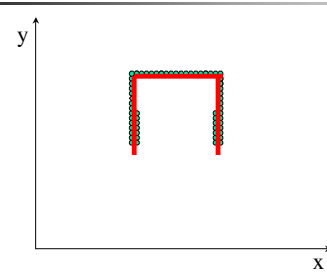
Carnegie Mellon

Nonlinear PCA



Carnegie Mellon

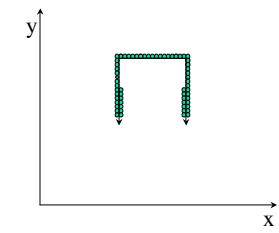
Nonlinear PCA



Carnegie Mellon

Nonlinear PCA

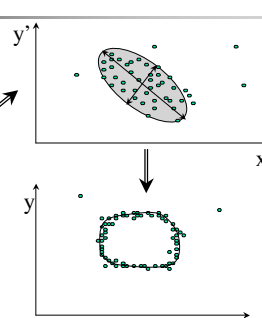
$X_{n \times m}$ is the original data matrix, n points, m dimensions


$$X'_{n \times m'} = F(X_{n \times m})$$

Carnegie Mellon

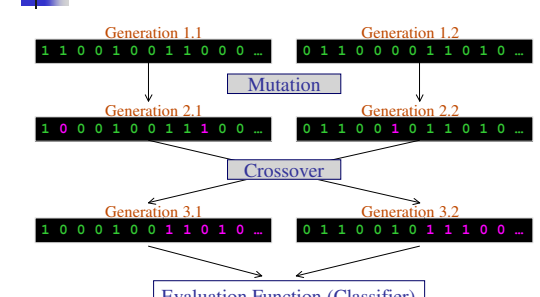
Kernel PCA

Kernel Function

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$
$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$


Carnegie Mellon

Genetic Algorithm



Carnegie Mellon

Stepwise Discriminant Analysis

1. Calculate Wilk's lambda and its corresponding F-statistic of the training data.

$$\Lambda(m) = \frac{|W(X)|}{|T(X)|}, X = [X_1, X_2, \dots, X_m]$$

$$F_{io-enter} = \left(\frac{n-q-m}{q-1} \right) \left(\frac{1-\Lambda(m+1)}{\Lambda(m+1)} \right)$$

2. Forward-Backward selecting features according to the F-statistics.

Carnegie Mellon

References

- Berry, Michael: <http://www.cs.utk.edu/~lsi/>
- Duda, R. O. and P. E. Hart (1973). *Pattern Classification and Scene Analysis*. New York, Wiley.
- Faloutsos, C. (1996). *Searching Multimedia Databases by Content*, Kluwer Academic Inc.
- Foltz, P. W. and S. T. Dumais (Dec. 1992). "Personalized Information Delivery: An Analysis of Information Filtering Methods." *Comm. of ACM (CACM)* 35(12): 51-60.

Carnegie Mellon

References

- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, Academic Press.
- Jolliffe, I. T. (1986). *Principal Component Analysis*, Springer Verlag.
- Aapo Hyvarinen, Juha Karhunen, and Erkki Oja *Independent Component Analysis*, John Wiley & Sons, 2001.

Carnegie Mellon

References

- Korn, F., A. Labrinidis, et al. (2000). "Quantifiable Data Mining Using Ratio Rules." *VLDB Journal* 8(3-4): 254-266.
- Jia-Yu Pan, Hiroyuki Kitagawa, Christos Faloutsos, and Masafumi Hamamoto. *AutoSplit: Fast and Scalable Discovery of Hidden Variables in Stream and Multimedia Databases*. PAKDD 2004

Carnegie Mellon

References

- Press, W. H., S. A. Teukolsky, et al. (1992). *Numerical Recipes in C*, Cambridge University Press.
- Strang, G. (1980). *Linear Algebra and Its Applications*, Academic Press.
- Caetano Traina Jr., Agma Traina, Leejay Wu and Christos Faloutsos, *Fast feature selection using the fractal dimension*, XV Brazilian Symposium on Databases (SBBD), Paraiba, Brazil, October 2000

Carnegie Mellon

Classification

Carnegie Mellon

Outline

- ...
- Feature extraction
- Feature selection / Dim. reduction
- ➔ ■ Classification
- ...

Carnegie Mellon

Outline

- ...
- Feature extraction
- Feature selection / Dim. reduction
- ➔ ■ Classification
 - classification trees
 - support vector machines
 - mixture of experts

Carnegie Mellon

Image Classification

Mitoch. Nucleolar Actin

Endosomal Tubulin ???

Carnegie Mellon

Image Classification

+ -

???

Carnegie Mellon

Decision trees - Problem

area	brightness	#objects	...	CLASS-ID
30	150	3		+
				...
				-

??

Carnegie Mellon

Decision trees

■ Pictorially, we have

num. attr#2 (e.g., brightness)

num. attr#1 (e.g., 'area')

Carnegie Mellon

Decision trees

- and we want to label '?'

num. attr#2 (e.g., brightness)

num. attr#1 (e.g., 'area')

Carnegie Mellon

Decision trees

- so we build a decision tree:

num. attr#2 (e.g., brightness)

40

50

num. attr#1 (e.g., 'area')

Carnegie Mellon

Decision trees

- so we build a decision tree:

area < 50

Y

N

bright. < 40

Y

N

...

bright. 40

50 'area'

Carnegie Mellon

Decision trees

- Goal: split address space in (almost) homogeneous regions

area < 50

Y

N

bright. < 40

Y

N

...

bright. 40

50 'area'

Carnegie Mellon

Details - Variations

- Pruning
 - to avoid over-fitting
- AdaBoost
 - (re-train, on the samples that the first classifier failed)
- Bagging
 - draw k samples (with replacement); train k classifiers; majority vote

adv

Carnegie Mellon

AdaBoost

- It creates new and improved base classifiers on its way of training by manipulating the training dataset.
- At each iteration it feeds the base classifier with a different distribution of the data to focus the base classifier on hard examples.


$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$
- Weighted sum of all base classifiers.

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

adv

Carnegie Mellon

Bagging



- Use another strategy to manipulate the training data: Bootstrap resampling with replacement.
- 63.2% of the total original training examples are retained in each bootstrapped set.
- Good for training unstable base classifiers such as neural network and decision tree.
- Weighted sum of all base classifiers.

Carnegie Mellon

Conclusions -Practitioner's guide:

- **Many** available implementations
 - e.g., C4.5 (freeware), C5.0
 - Also, inside larger stat. packages
- Advanced ideas: boosting, bagging
- Recent, scalable methods
 - see [Mitchell] or [Han+Kamber] for details

Carnegie Mellon


References

- Tom Mitchell, *Machine Learning*, McGraw Hill, 1997.
- Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2000.

Carnegie Mellon

Outline

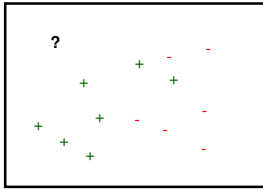
- ...
- Feature extraction
- Feature selection / Dim. reduction
- Classification
 - classification trees
 - support vector machines
 - mixture of experts



Carnegie Mellon

Problem: Classification

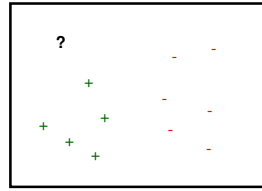
- we want to label '?'



Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??

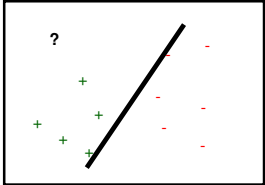


Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??

bright.



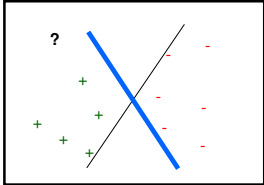
area

Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??

bright.



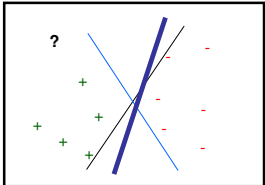
area

Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??

bright.



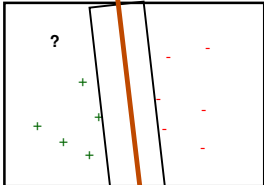
area

Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??

bright.



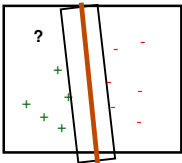
area

Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??
- A: the one with the widest corridor!

bright.



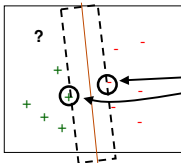
area

Carnegie Mellon

Support Vector Machines (SVMs)

- we want to label '?' - linear separator??
- A: the one with the widest corridor!

bright.



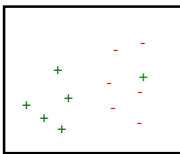
area

Carnegie Mellon

Support Vector Machines (SVMs)

- Q: what if + and - are not separable?
- A: penalize mis-classifications

bright.



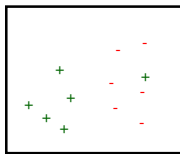
area

Carnegie Mellon

Support Vector Machines (SVMs)

- Q: how about non-linear separators?
- A:

bright.



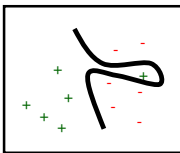
area

Carnegie Mellon

Support Vector Machines (SVMs)

- Q: how about non-linear separators?
- A: possible (but need human)

bright.



area

Carnegie Mellon

Performance

- training:
 - $O(N_s^3 + N_s^2 L + N_s L d)$ to
 - $O(d * L^2)$
- where
 - N_s : # of support vectors
 - L : size of training set
 - d : dimensionality

Carnegie Mellon

Performance

- classification
 - $O(M N_s)$
- where
 - N_s : # of support vectors
 - M : # of operations to compute similarity (~ inner product ~ 'kernel')

Carnegie Mellon

References

- C.J.C. Burges: *A Tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery 2, 121-167, 1998
- Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge, UK, 2000.

software:

- <http://svmlight.joachims.org/>
- <http://www.kernel-machines.org/>

Carnegie Mellon

Outline

- ...
- Feature extraction
- Feature selection / Dim. reduction
- Classification
 - classification trees
 - support vector machines
 - mixture of experts

Carnegie Mellon

Mixture of experts

- Train several classifiers
- use a (weighted) majority vote scheme

Carnegie Mellon

Conclusions: 6 powerful tools:

- shape & texture features:
 - wavelets
 - mathematical morphology
- Dim. reduction:
 - SVD/PCA
 - ICA
- Classification:
 - decision trees
 - SVMs

Carnegie Mellon

Back to Bob!

Carnegie Mellon