Applied Machine Learning

Adrian Barbu

Syllabus

- Textbooks:
  - Machine Learning by Tom M. Mitchell (publisher: McGraw-Hill)
  - Pattern Recognition and Machine Learning by Christopher M. Bishop (publisher: Springer)
  - The Elements of Statistical Learning by T. Hastie, R. Tibshirani, and J. H. Friedman (publisher: Springer).
- Course Materials:
  - CMU Machine Learning Class: http://www.cs.cmu.edu/~epxing/Class/10701/
  - Tom Michell's ML book website: http://www.cs.cmu.edu/~tom/mlbook.html
  - Nillson's ML book: http://ai.stanford.edu/~nilsson/mlbook.html
  - Blackboard class website

Syllabus

- Homework
  - 12 projects, each worth points depending on difficulty
  - 12 datasets
  - Some projects can be applied to many datasets
    - Pick any appropriate combination
  - One-three weeks for each project
  - Total homework grade capped at 90 points
- Quizzes
  - 5-10 random quizzes worth another 10 points

Syllabus

- Code
  - Students are encouraged to submit the code by email
  - The code for best results for each dataset will be posted on Blackboard for everyone's benefit
  - It is ok to obtain code from the web but

    Appropriate references to the code website or paper should be made in the project

Machine Learning

- Theory and Algorithms for
  - Representing or modeling
  - Classifying, predicting, clustering, recognizing
  - Reasoning under uncertainty
  - Reacting or taking actions to

  complex real world phenomena or information using the system's experience and an explicit model that can be
  - Understood
  - Modified using human prior knowledge
  - Can work automatically, without user interaction

Definition

Learning algorithms

- Given task T
- Improve performance measure P
- With experience E

Well defined learning problem:

- Define task T
- Choose performance measure P for how far we are from the task
- Specify how the experience E is obtained

Also of interest: running time

Text Classification

Task:

- Given a text, assign one or more category types based on contents
  - News
  - Poetry
  - Scientific
  - Political
  - …

A plan to spend up to $1 trillion to stimulate the sagging U.S. economy will not be ready for Obama to sign as soon as he takes office on Jan. 20. Instead, negotiations could stretch into mid-February.

Israel's invasion of Gaza to quell Hamas rocket fire on its cities has pushed the Middle East higher on Obama's already long list of foreign challenges that includes Iraq, Afghanistan and Russia. Obama's silence on the Israeli action has led to some sniping abroad.

U.S. President-elect Barack Obama gestures during remarks during a League of United Latin American Citizens conference in Washington, July 8, 2008. (REUTERS/Jonathan Ernst/Files)

Performance measure:

- % correctly classified

Experience:

- A dataset of texts and their labels (supervised learning)

Similar task without labels given: unsupervised learning.

Object Detection

Task:

- Detect objects of interest (e.g faces)

Performance measure:

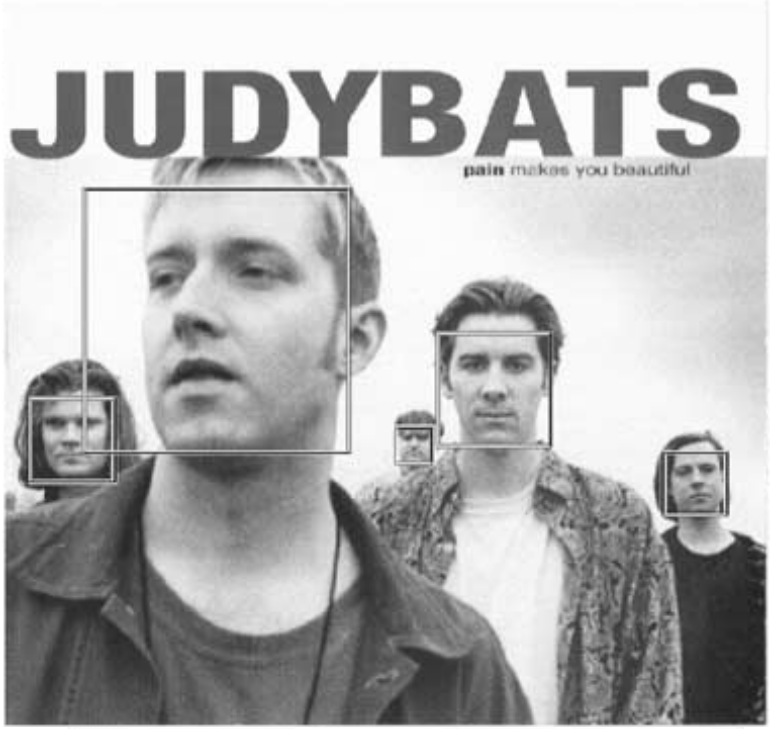
- Detection rate/ false alarm rate

Experience:

- A training set of images containing faces

Viola &Jones, 2001

Schneiderman & Kanade, 2000



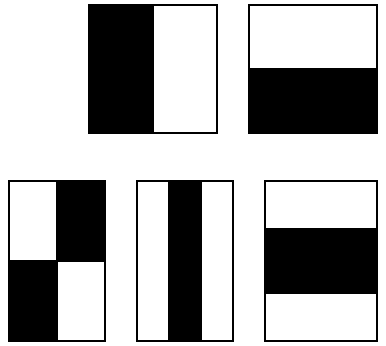
Face Detection Design Decisions

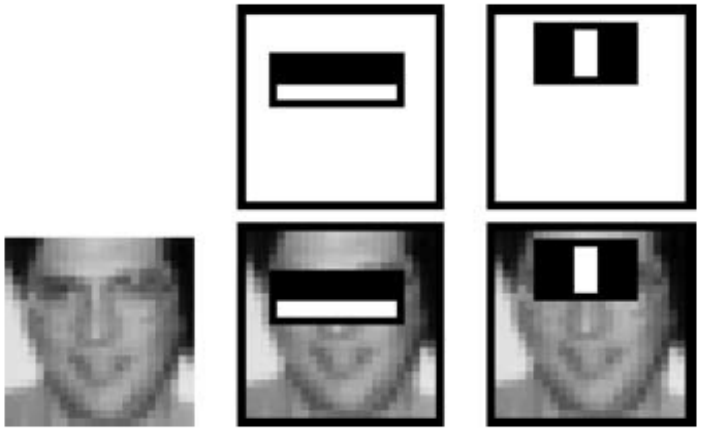
- Type of training experience:
  - Images with manually annotated faces (supervised)
    - Manually labeling data is expensive
  - Images with faces (unsupervised)
  - Mixture of both (semi-supervised)

- Amount of training experience
  - Data should be representative
    - Cover many poses/illuminations
    - Cover face alterations (beard, glasses, eye patch, occlusions, etc)
  - For good results, need 10,000 -100,000 faces

Face Detection Design Decisions

- Target function to be learned:
  - Classification: face/non-face
  - Regression: relative position of closest face
- Number of face parameters: position, scale, etc.
- Feature pool:
  - Type, number
  - E.g. Haar, steerable, etc,
- Type of learning algorithm used:
  - Decision trees
  - Boosting
  - SVM
  - CNN



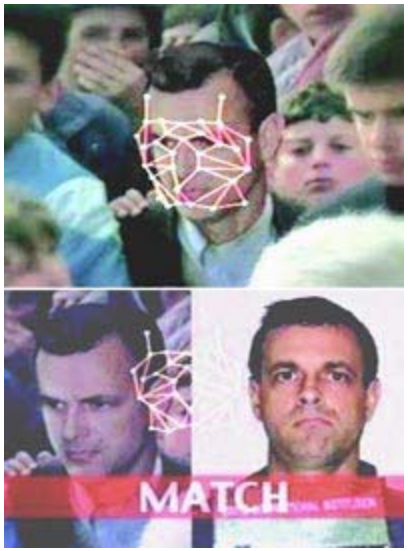


Best two features for face detection

Viola, Jones, 2004

Object Recognition

- Task:
  - Given an image containing a face, check if the face is similar to one from an existing database
- Performance measure:
  - Percentage correctly recognized/error rate
- Experience:
  - Dataset of 2D/3D face images of persons of interest



Identix face recognition system

Object Segmentation

Task:

- Given an image, find the object and its contour
- E.g. 3D CT heart segmentation

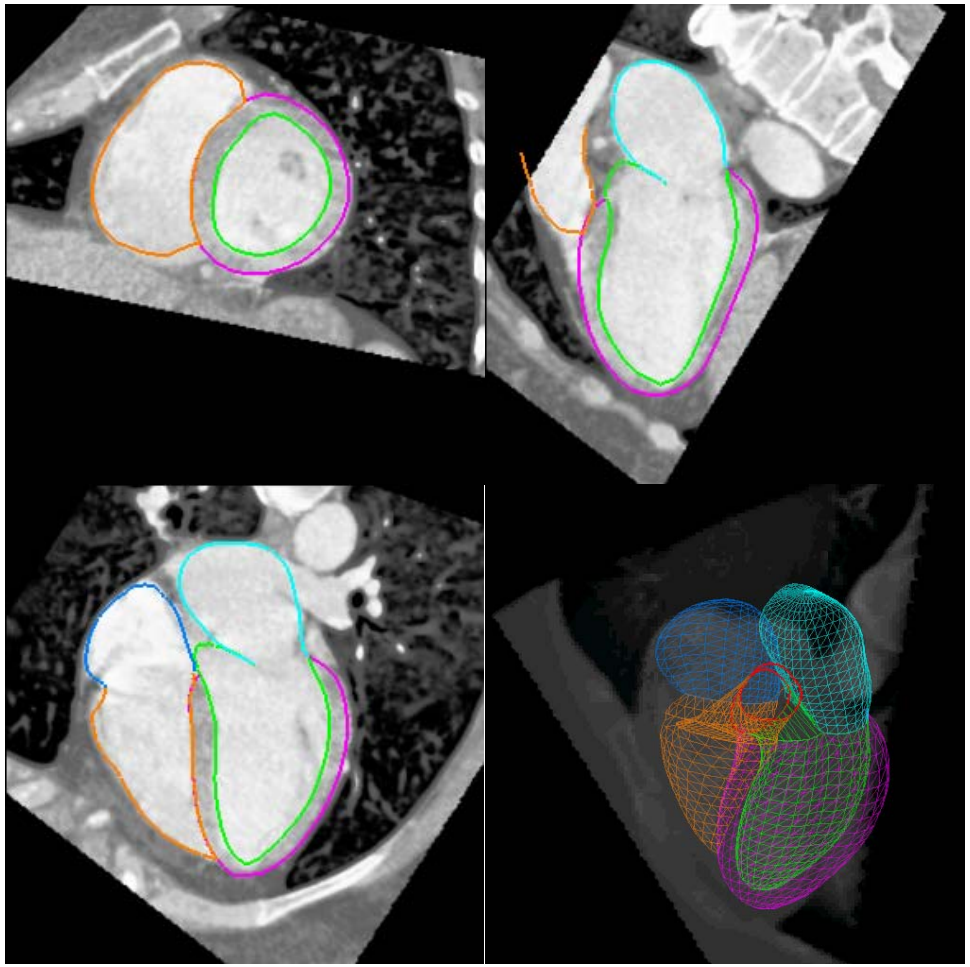
Performance measure:

- Average point to mesh distance

Experience:

- A dataset of images annotated by experts
- E.g. 323 heart CT scans

Running time: 1 sec



Bioinformatics

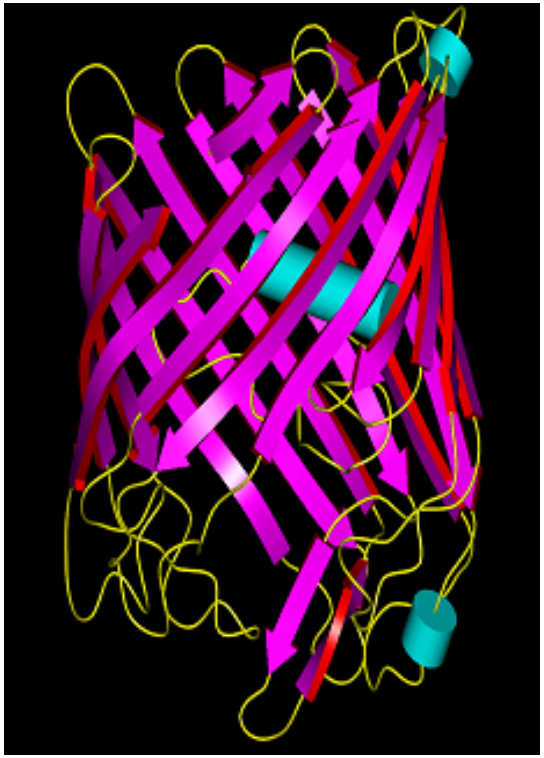
- Predicting protein function from gene sequence
  - Protein shape (folding) $\longleftrightarrow$ its function

```
attcgatcgatcgat
cgatcaggcgcgcta
Cgagcggcgaggacc
tcatcatcgatcag…
```

DNA sequence

```
MRPQAPGSLVDPNEDEL
RMAPWYWGRISREEAKS
ILHGKPDGSFLVRDALS
MKGEYTLTLMKDGQ...
```

Aminoacid sequence



Protein shape (folding)

  - Training examples: pairs of proteins and their folding
- Finding genes related to cancer
  - Training examples: DNA sequence of people with a specific cancer type and of healthy people
  - Very good drugs against certain cancers

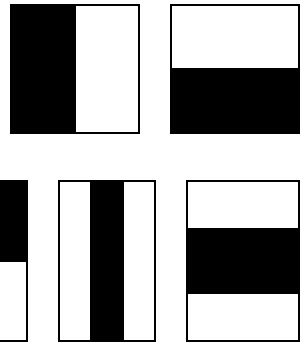
Other Applications of Machine Learning

- Optical character Recognition (OCR)
- Speech Recognition
  - Outperforms all non-learning based methods
  - Speaker dependent/ independent
- Learn to drive an autonomous vehicle
  - DARPA grand challenge (2005)
  - DARPA urban challenge (2007)
  - Google self-driving cars
- Learning to play backgammon
  - Competitive with human world champions
- Learning brain commands
  - Paralyzed people can move a cursor just by thinking about it

Learning a Target Function

Problem setup

- The space of possible instances X
  - E.g. all possible 21x21 pixel windows for face detection



- The space of possible function values Y
  - Discrete for *classification*
  - Continuous for *regression*
- The space of allowable functions $\mathcal{H} \subset \{h : X \rightarrow Y\}$
  - E.g. linear combinations of thresholded features $f_i : X \rightarrow Y$

Experience:

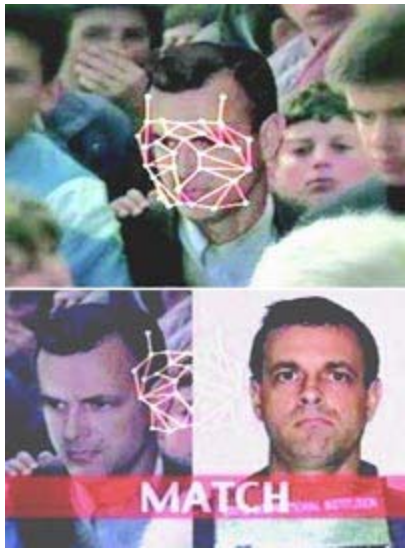
- A set of training examples $E = \{(x_i, y_i) | x_i \in X, y_i \in Y\}$

Learning:

- Find $h \in \mathcal{H}$ that best interpolates E

Classification vs Regression

- Classification = discrete output
  - It is important to find the correct class
  - If it's not correct → it is wrong
    - Misclassification rate
  - E.g. face recognition



- Regression = continuous output
  - It is important to predict the right value
  - Measure of the prediction error
    - Mean Square Error
  - E.g. age regression

Age=?

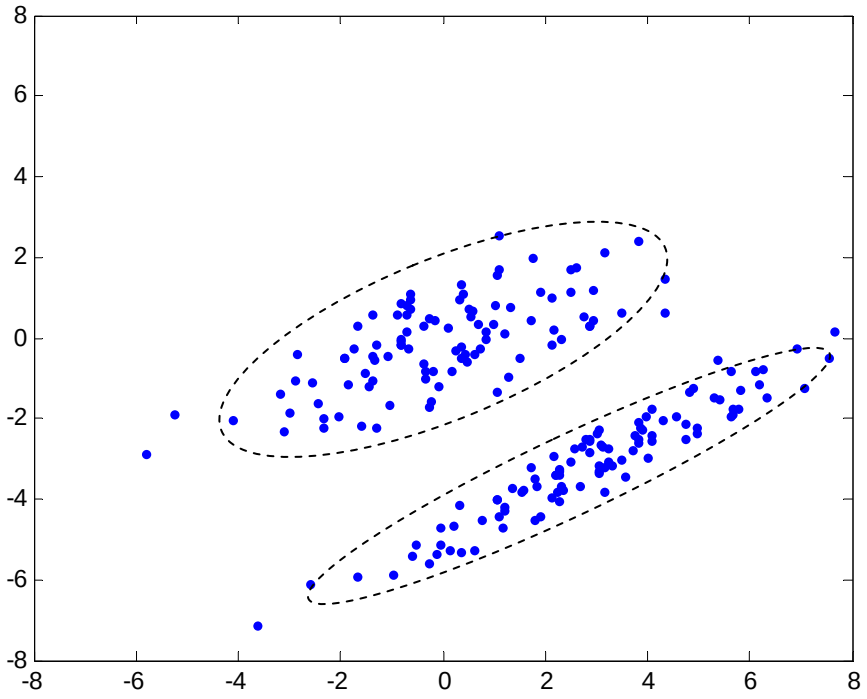
Learning a Probability Density Function

- Aka Density Estimation
- Unsupervised learning
- **Experience:** A set of training examples $E = \{x_i | x_i \in X\}$
- **Learning:** Find a probability density function (pdf)

$$f : X \rightarrow \mathbb{R}_+$$

- Parametric pdf
  - Gaussian, student-t, etc.
  - Mixture models

  → parameter estimation
- Nonparametric pdf:
  - Kernel density estimation
  - Density Trees
  - Bayesian Networks



Conclusions

- Supervised Learning=function approximation:
  - Instance space, X
  - Sample of labeled training data { $<x_i, y_i>$}
  - Hypothesis space, $\mathcal{H} \subset \{h : X \rightarrow Y\}$
  - Y can be discrete for classification or real-valued for regression
- Learning is a search/optimization problem over $\mathcal{H}$
- Various objective functions
  - Minimize training error
  - Generalization power: small error on data that was not used for training
    - Avoid overfitting

Conclusions

- Machine learning works
  - Very popular nowadays
  - Great impact in surveillance, medical imaging, social networks, commerce and beyond
  - Faster computers = more difficult problems can be attacked.
  - Many real-time applications
    - Real-time face detection
    - Autonomous driving
    - Tunnel surveillance
    - Speech recognition
- Many machine learning jobs
  - Google, Facebook, Apple, Tesla, SpaceX, etc.