Advanced Concepts in Signal Processing

Lecturer:
Dr Sotirios Tsaftaris

Some slides adopted from: Mike Davies, A. Katsaggelos, S. Theodoridis, A. Ng

Advanced Concepts in Signal Processing

Overview

Advanced statistical models for analysis and processing of signals. Covering: "*Artificial Neural Networks*", "*Machine Learning*" and "*Pattern Recognition*".

Concepts covered:

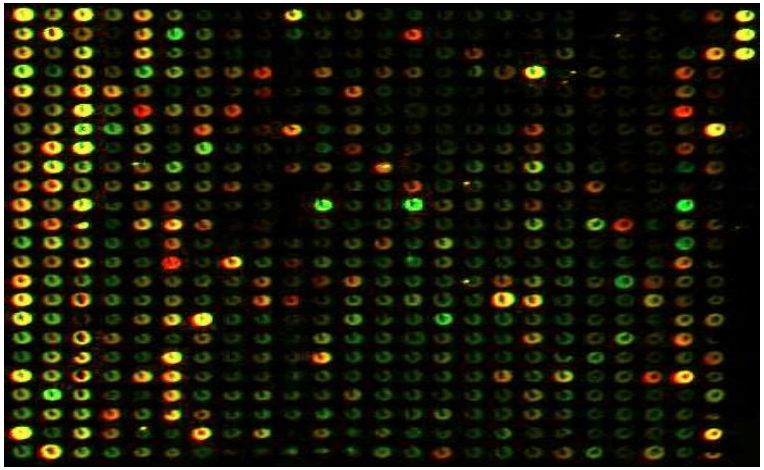
- Classification and recognition
- Statistical Inference and learning
- Clustering
- Data reduction (e.g. PCA)
- Blind signal separation (the "Cocktail Party Problem")

How does this fit our research?

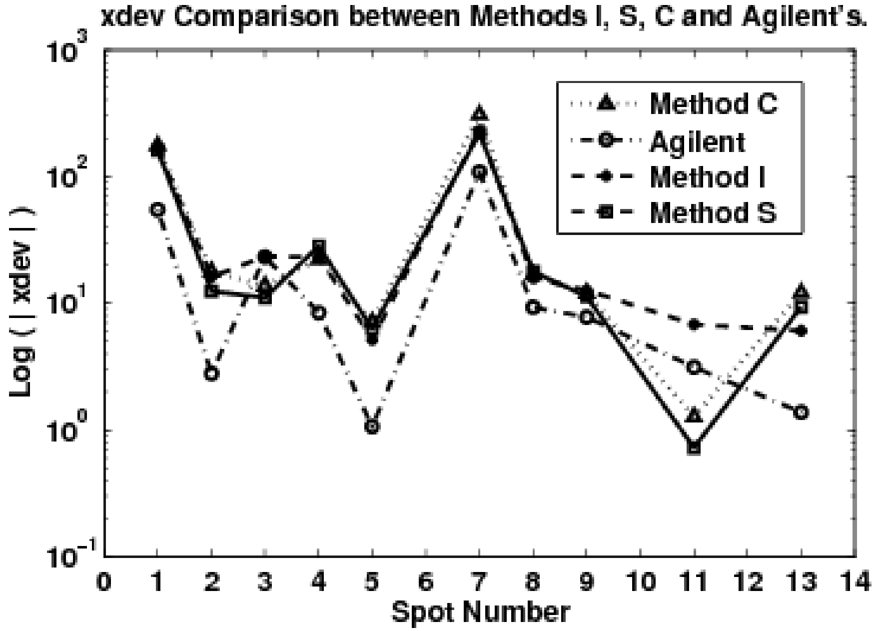
- We build algorithms to analyze **imaging** data (2D, 3D, 2D+t, 3D+t)
- From a **variety** of domains
- Use machine learning throughout

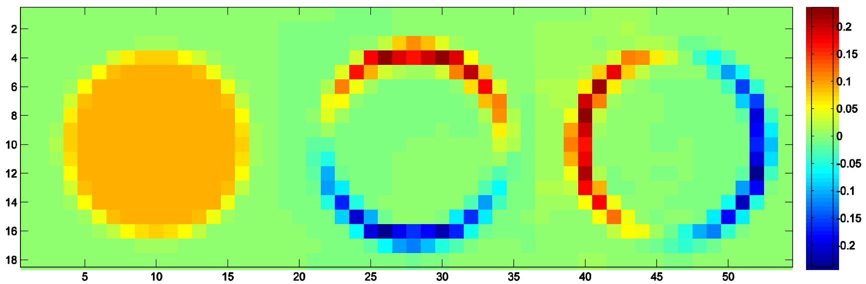
- Some examples…

Microarray Imaging



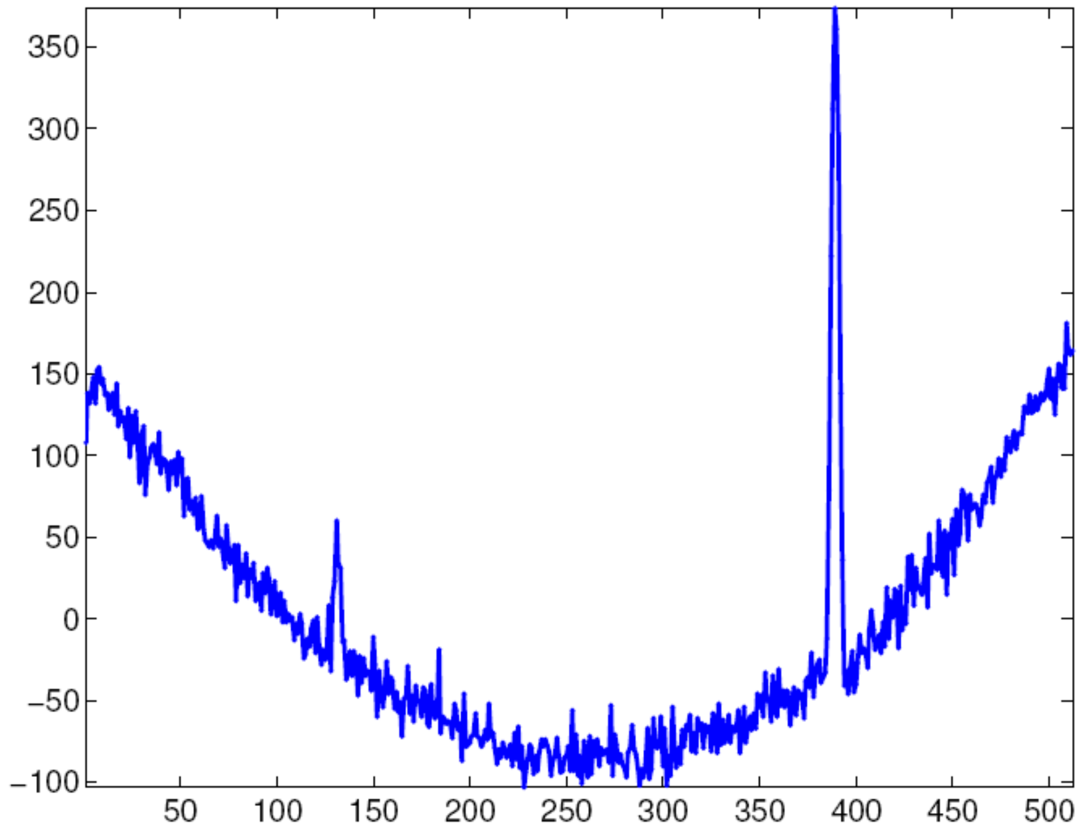
- Use learning methods (**PCA**) for denoising and analysis



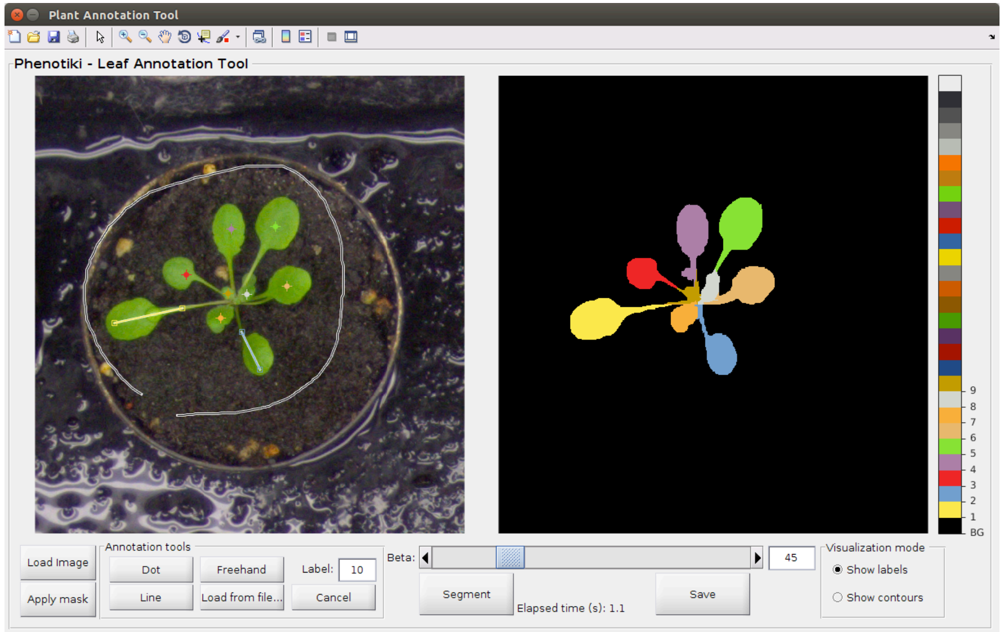


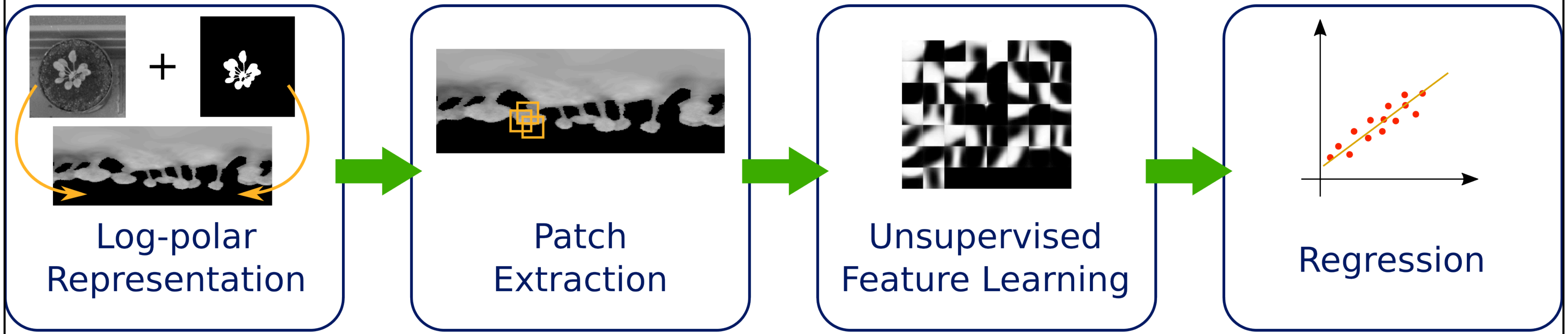
AFM Image Restoration

- Recover radius of DNA carbon NT
- Cantilever distortion ➔errors
- **Iterate** K-means **clustering** in Object / Background points & convex polynomial **fitting** on background



Plant Phenotyping





As you realize…

- We use pattern recognition and machine learning methods all day, every day, …
- We also develop even new pattern recognition algorithms
  - Mostly on representation learning (the general term for learning features from data, think of PCA, ICA, etc)
- But you are not here to learn about me (us) but to learn about machine learning and pattern recognition

What is machine learning?

- What do you think?

Some applications

- Email spam filtering
- Netflix/Amazon recommendations
- Google suggested queries
- The Google index itself

Extreme(…) applications

- MIT flight
- http://www.youtube.com/watch?v=aiNX-vpDhMo
- Robot in the dessert
- http://www.youtube.com/watch?v=OIOtOmyySQo
- Google car
- http://www.youtube.com/watch?v=cdgQpa1pUUE

A popular with waves…

- Computer world 2007
  - 1) machine learning

- Really?
  - Lets check:
  - https://www.google.com/trends/explore#q=Machine%20learning%2C%20pattern%20recognition&cmpt=q&tz=Etc%2FGMT-1

Big Data

- Obama administration announces $200 million 'big data' research and development initiative, *White House*, March 2012.
- 1000 Genomes on Amazon Cloud, *NIH*, March 2012
- Big data: The next frontier for innovation, competition, and productivity, *McKinsey Global Institute,* May 2011.
- Statisticians and "Big Data" Analysts in High Demand, *BioJobBlog*, March 2012.
- Big Data / Data Mining
  - http://ovum.com/2012/04/05/big-data-creates-demand-for-analytics-skills/

➔ Need to identify relationships in large data

➔ Need machines to do this for us

So it is really popular

- Computer world 2012
  - **5. Business Intelligence/Analytics**
  - ***26%*** *plan to hire for this skill in the next 12 months.*
- Other reasons that contributed to popularity?

Is becoming interdisciplinary

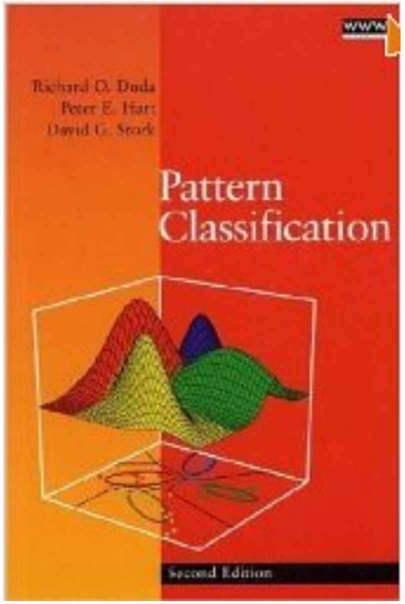
- Examples:
  - Machine learning methods without tears: A primer for ecologists
- With examples even in communications:
  - Learning to Decode Linear Codes Using Deep Learning
  - Convolutional Radio Modulation Recognition Networks

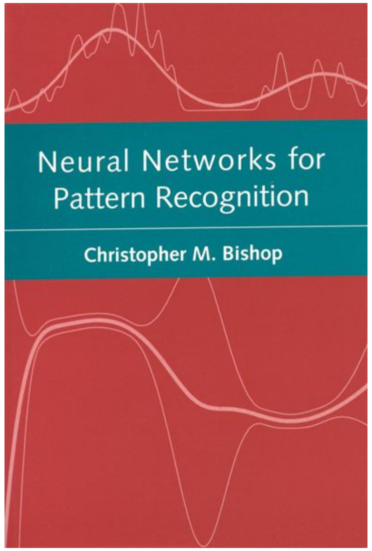
What is machine learning?

- Arthur Samuel [1959] (informal definition) Gives computers ability to learn without being explicitly programmed.
  ➔He built the very first checker's program
- Tom Mitchell [98] (more formal): A well-posed learning problem is defined as follows:
  – A computer program is set to learn from an **experience E** with respect to some **task T** and some **performance measure P** if its performance on **T** as measured by P **improves** with experience E.

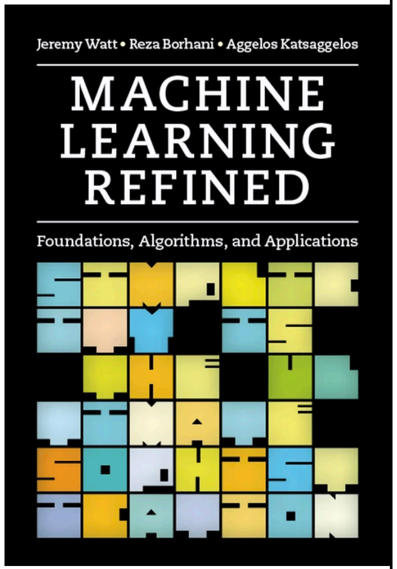
Text Books

- Useful texts ...



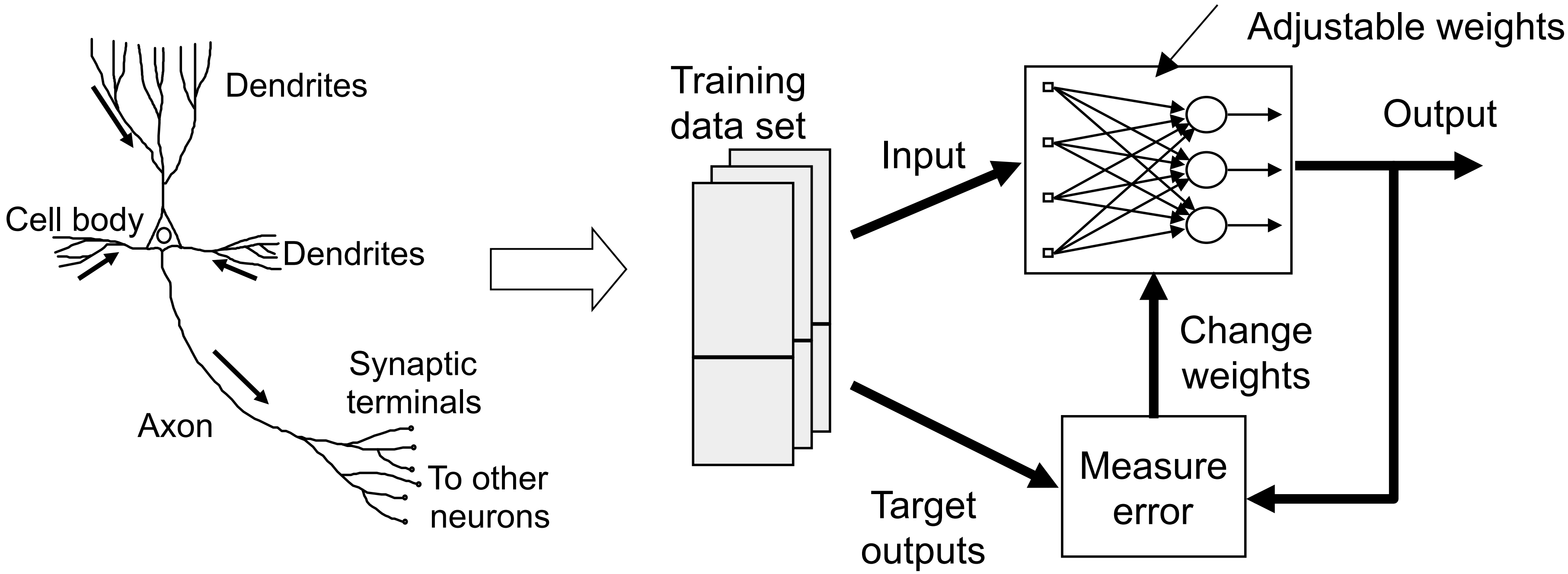


and



Neural Networks

From Biological Neurons to Artificial NNs; Feedforward NNs; NN learning models. MLPs and alternatives (e.g. RBFs)

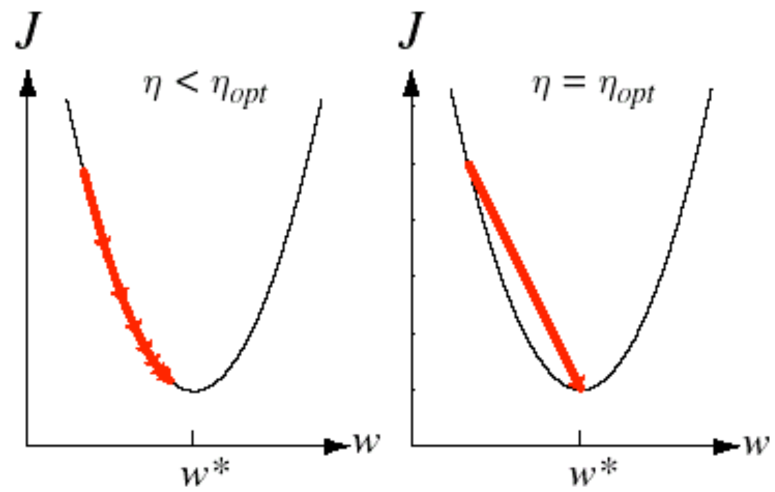


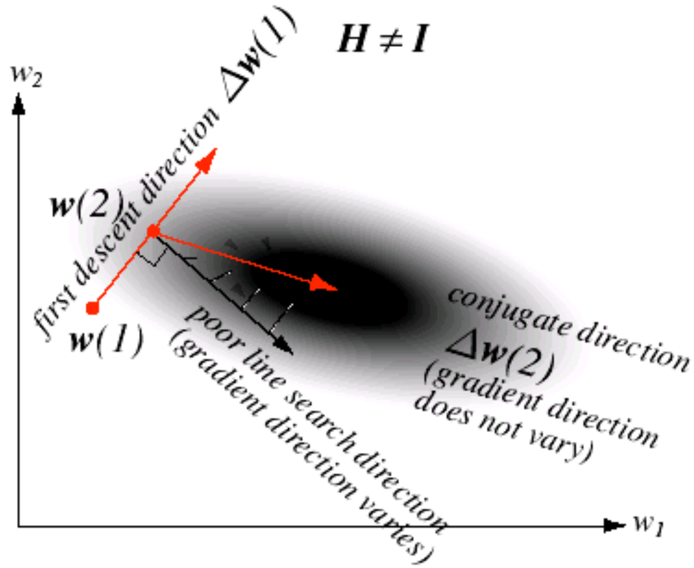
Optimization

Many DSP techniques need *optimization*, e.g.

- Minimizing error in a neural network/adaptive system
- Maximizing probability in Bayesian inference

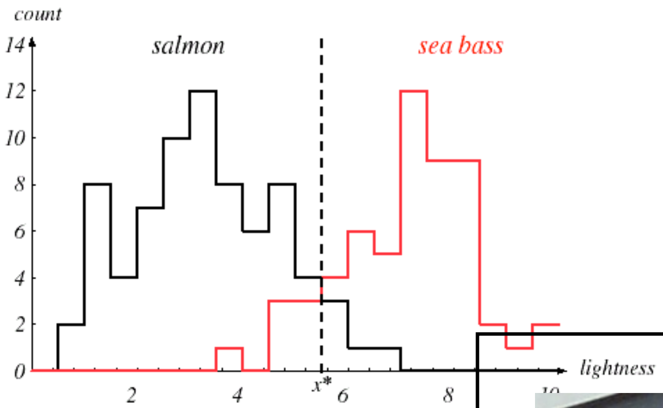
From simple "steepest descent" to more advanced techniques (conjugate gradient, model trust regions,…)





Statistical Inference/Learning

Use of probability theory (e.g. Maximum Likelihood) to estimate the “best” answer to classification problems...

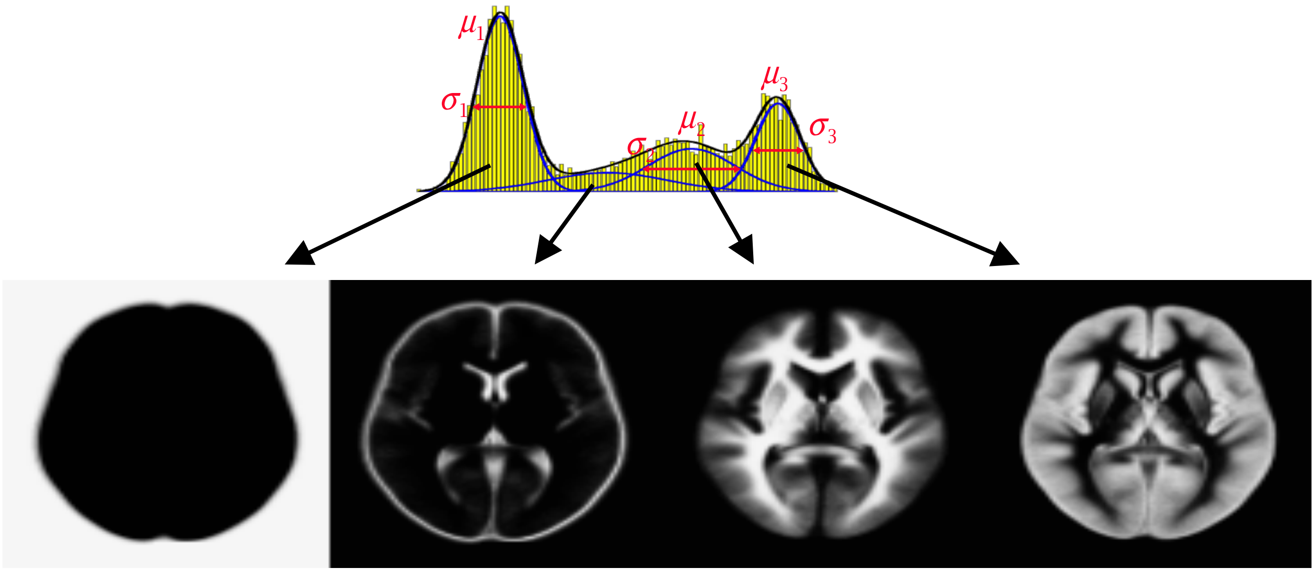


3.3.1. .A.V.M. .4.

Clustering

Collecting together "similar" observations or signals.

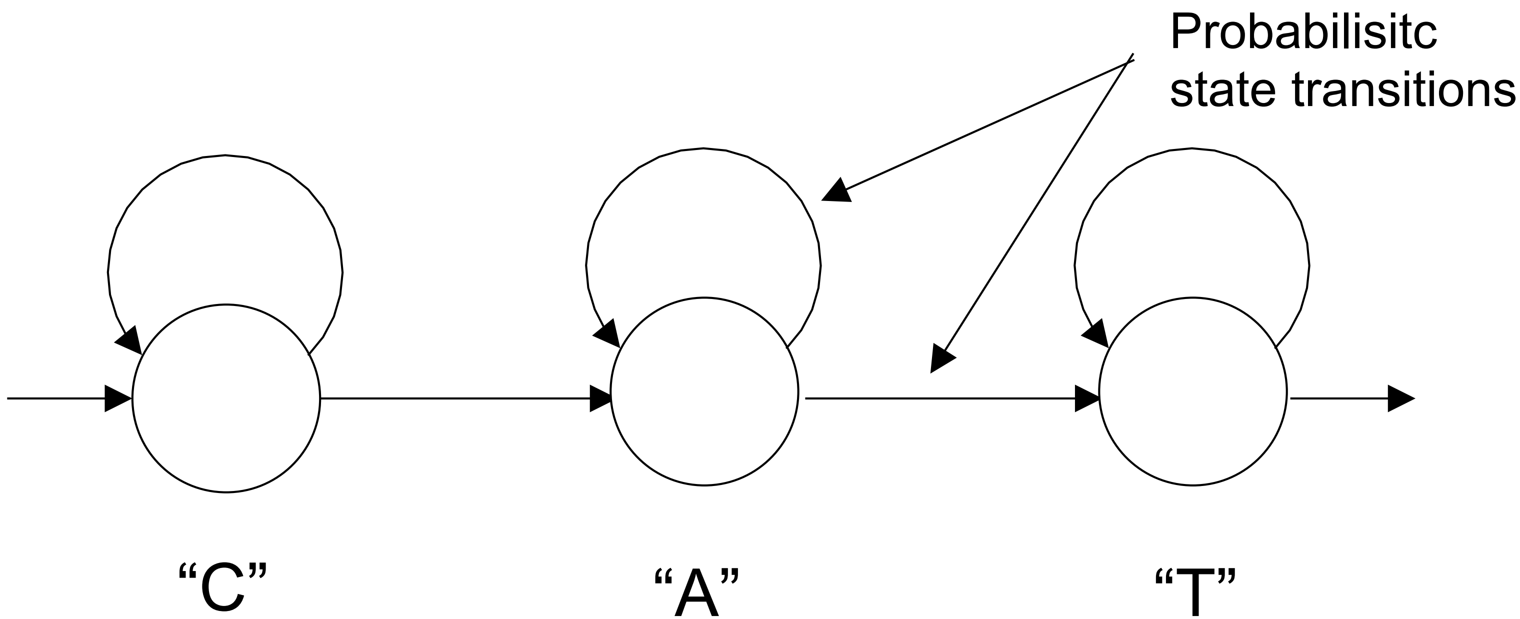
- Gaussian Mixture Models: learning (EM) and issues;
- K-means algorithm: coding optimality + links with GMMs



Hidden Markov Models (HMMs)

Dynamic Classification Problems using Hidden Markov Models (HMMs)

e.g. application to statistical modelling of speech

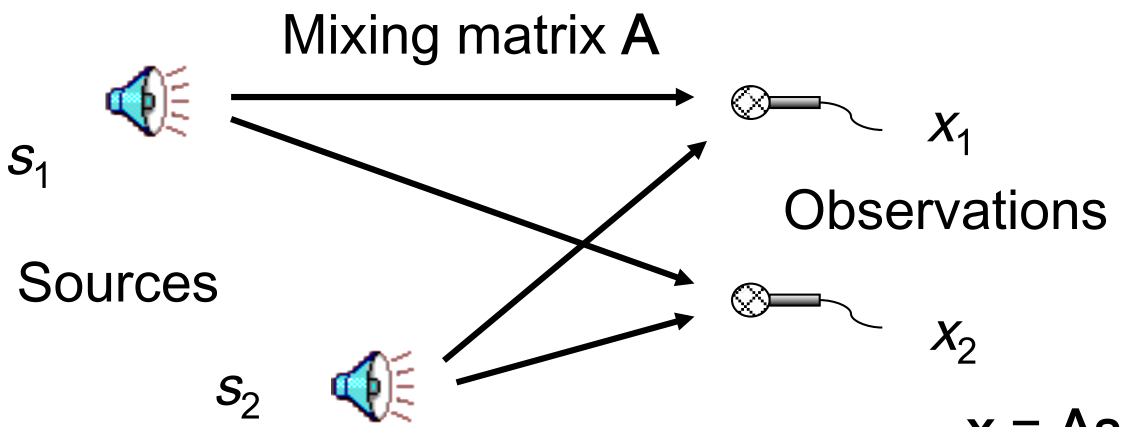


Principal & Independent Component Analysis

Decomposing signals into useful low dimensional subsets: Principal Component Analysis and Independent Component Analysis.

- For feature space selection in classification
- For redundancy reduction
- For blind signal separation (e.g. the “cocktail party problem”)

PCA eigenfaces



$\mathbf{x} = \mathbf{A}\mathbf{s}$

The cocktail party problem

Lecture timetable (approximate)

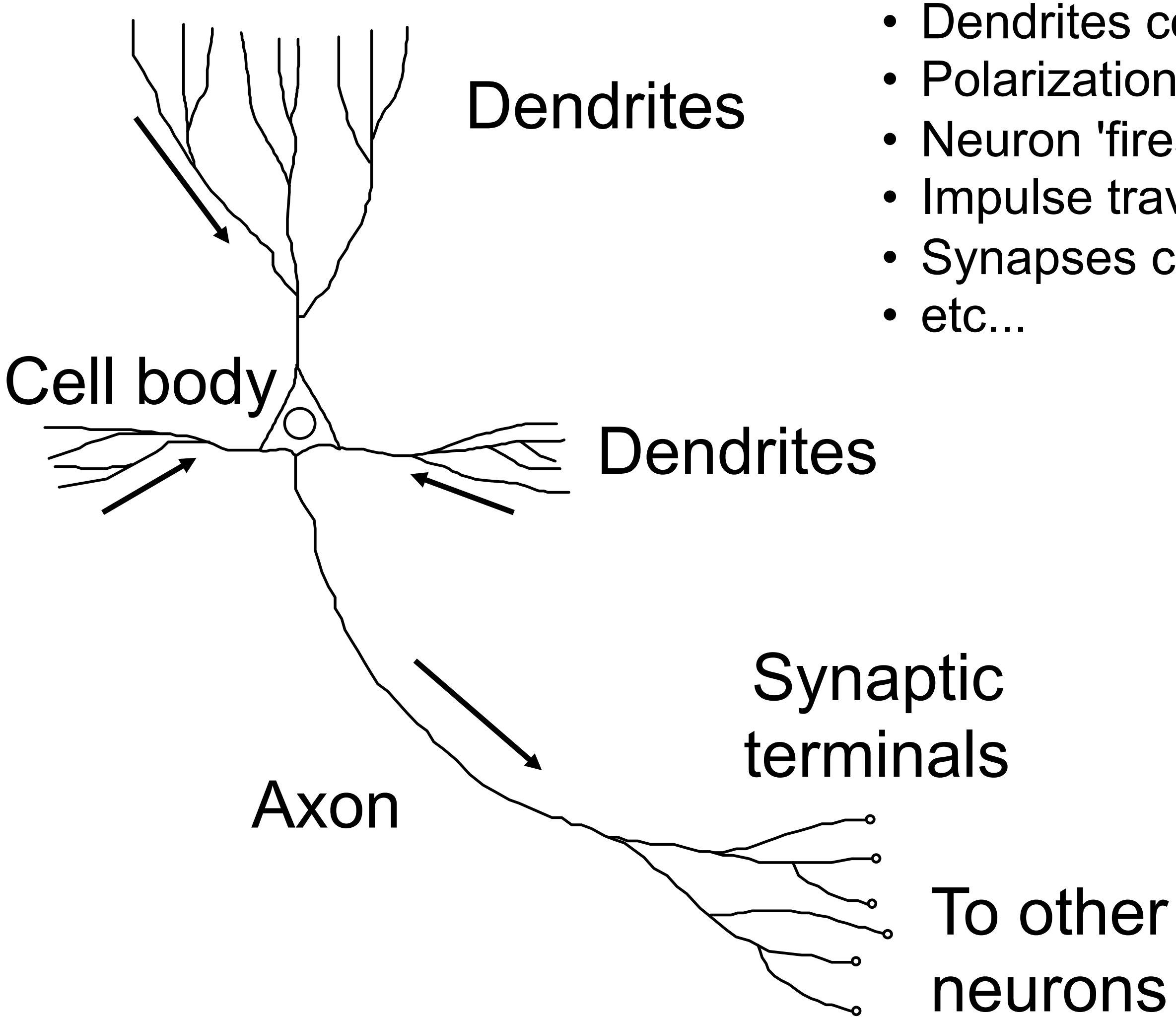
1. Introduction & Overview
2. Neural Networks
3. Linear Discriminant Functions
4. Linear Discriminant Functions
5. Linear Non-separable
6. Multi-layer Perceptrons
7. Multi-layer Perceptrons
8. Numerical Optimization
9. Numerical Optimization
10. Bayesian Decision Theory
11. Bayesian Decision Theory
12. Model Learning
13. Model Learning
14. Clustering
15. Clustering
16. Hidden Markov Models
17. Hidden Markov Models
18. PCA & ICA
19. PCA & ICA
20. Wrap up

Neural Networks

Neural Networks

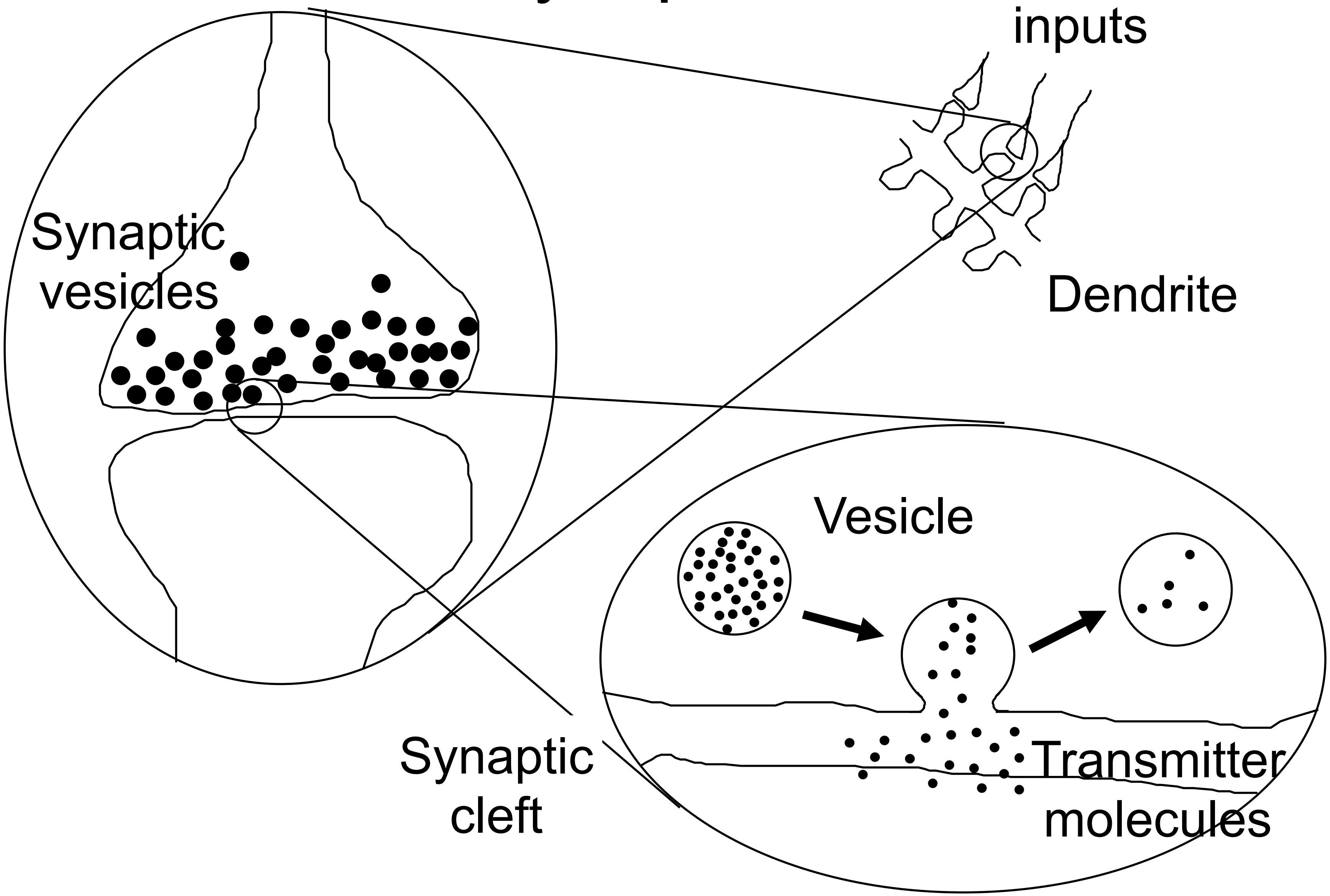
- Inspired by biological brains: *Parallel*, *distributed* processing
- Acquires knowledge through *learning.* Stores knowledge in connection strengths (*weights*) between *neurons*
- Applicable to data-driven problems
- Human brain is massively parallel:
  - 100 billion (10^{11}) neurons
  - 100 trillion (10^{14}) connections (synapses)
  - 100 (10^{2}) operations per second
- Very different from fast computers:
  - 1-1000 (1-10^{3}) processors
  - 1 trillion (10^{12}) operations per second

A Biological Neuron



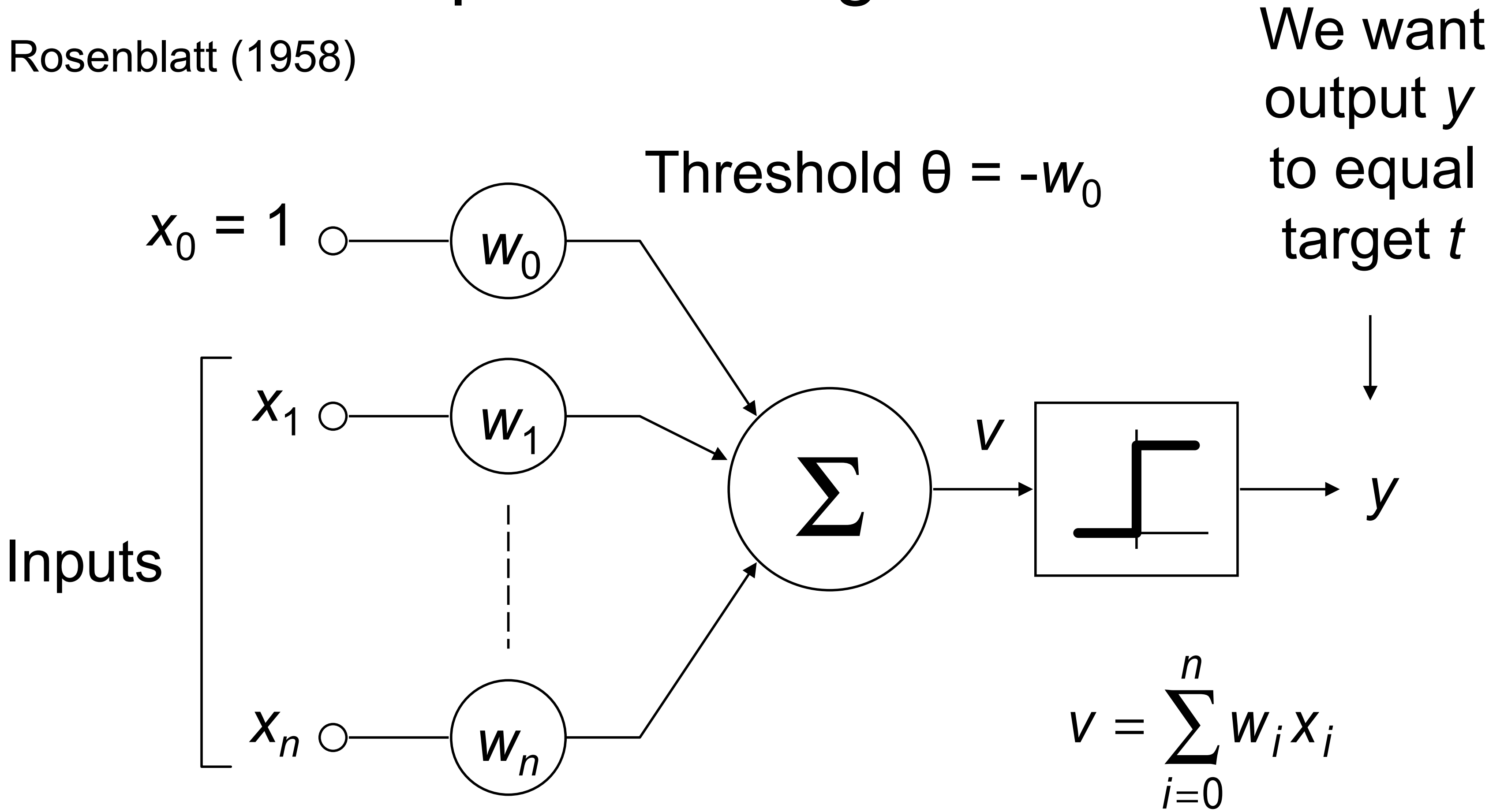
- Dendrites collect incoming impulse
- Polarization builds up in cell, until...
- Neuron 'fires'
- Impulse travels down Axon
- Synapses connected to other neurons
- etc...

Synapses


Synaptic
inputs
Dendrite
Synaptic
vesicles
Vesicle
Synaptic
cleft
Transmitter
molecules

The Perceptron: A Simple Learning Neuron

Rosenblatt (1958)

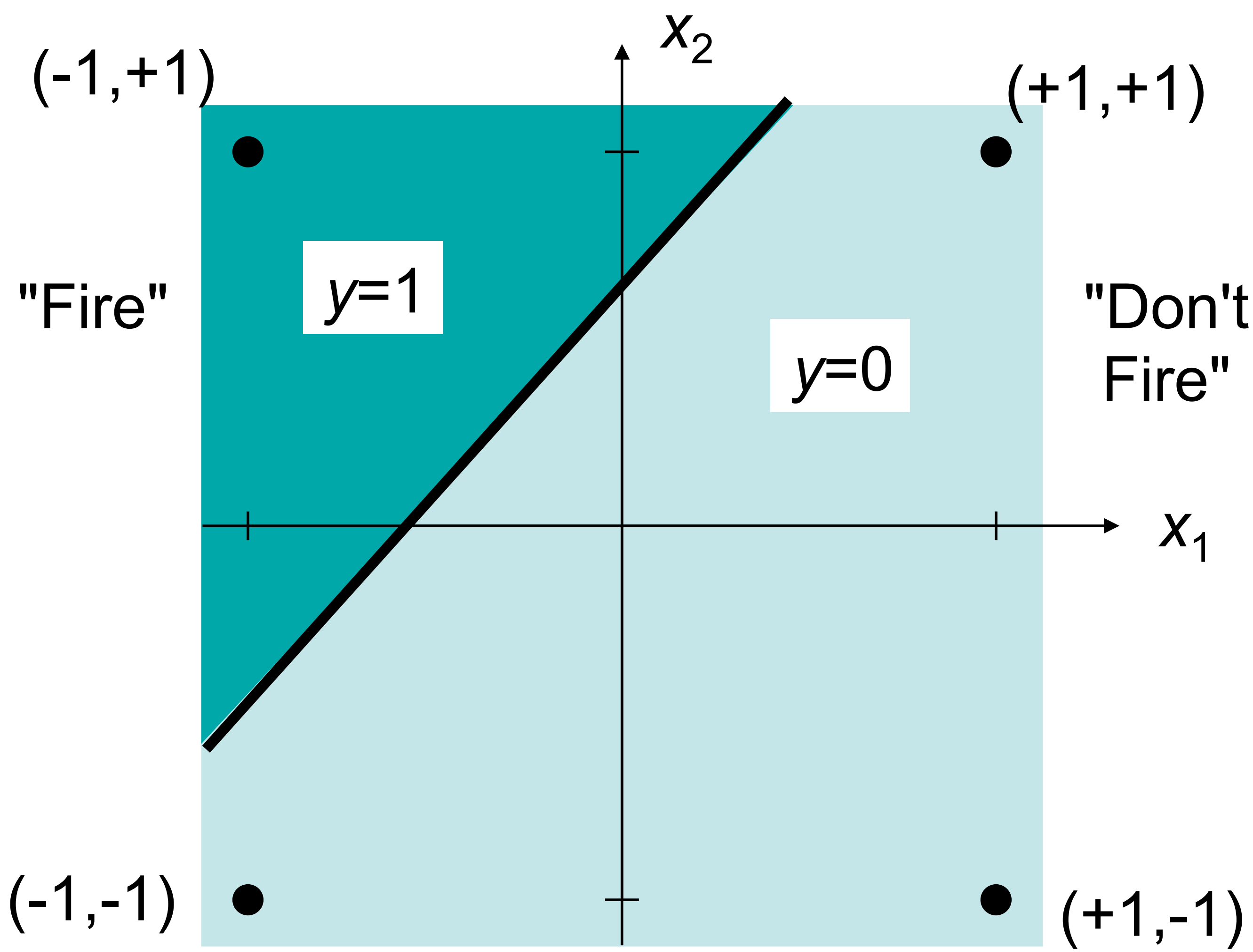


$$v = \sum_{i=0}^{n} w_i x_i$$

$$y = f(v)$$

Inputs may be from {-1, +1} or {0, +1}

Decision Boundary



Perceptron Learning Algorithm

One example of a learning algorithm (presents samples one at a time)

For all input vectors in training set:

1) Present input vector x
2) Calculate y=1 if $w^T x \geq 0$, y=0 if $w^T x < 0$
3) Compare y with target output t
    - a) If t=1 but y=0, set new w = old w + ηx [punish]
    - b) If t=0 but y=1, set new w = old w − ηx [punish]
    - c) Otherwise (If y=t), do nothing [reward]

Repeat until correct for all input vectors.

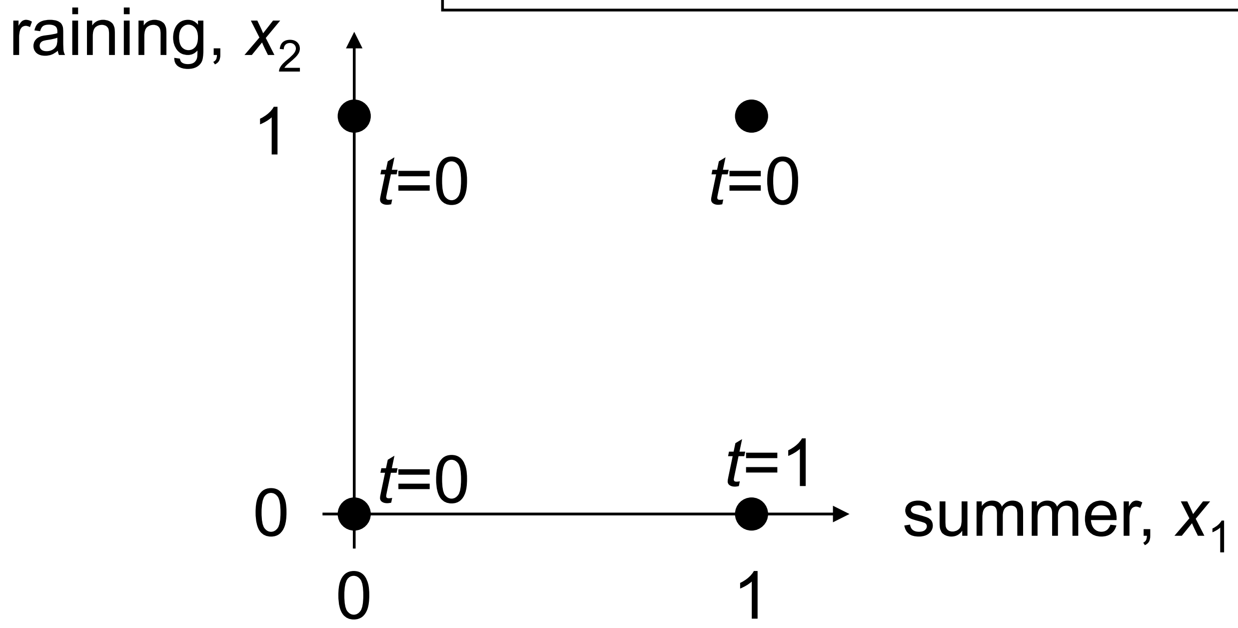
Factor η is called the *learning rate*

Simple Example

"If summer and not raining, play tennis"

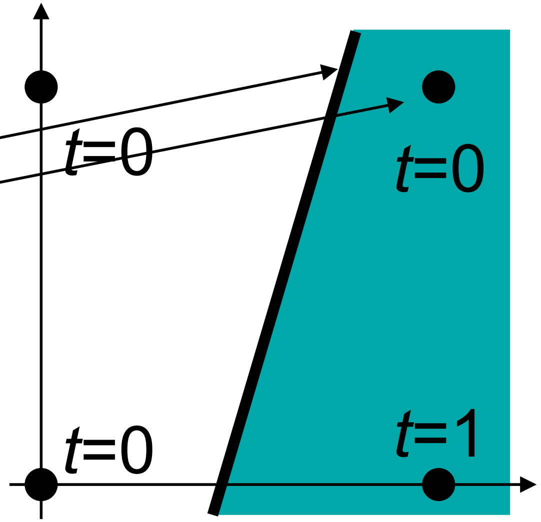
Training set.
Specifies target t
for different inputs

(threshold, x_0	1	1	1	1)
summer, x_1	0	0	1	1
raining, x_2	0	1	0	1
play tennis, t	0	0	1	0

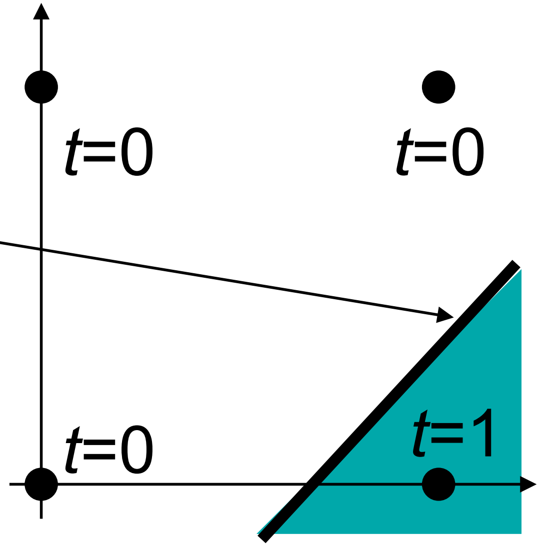


Simple Example (cont)

Suppose initially
$\mathbf{w} = (w_0, w_1, w_2) = (-0.5, +2.5, -1.5)$
Try input $\mathbf{x} = (1,1,1)$:
$\mathbf{w}^T\mathbf{x} = -0.5+2.5-1.5>0$
so y=1: <u>Wrong</u>



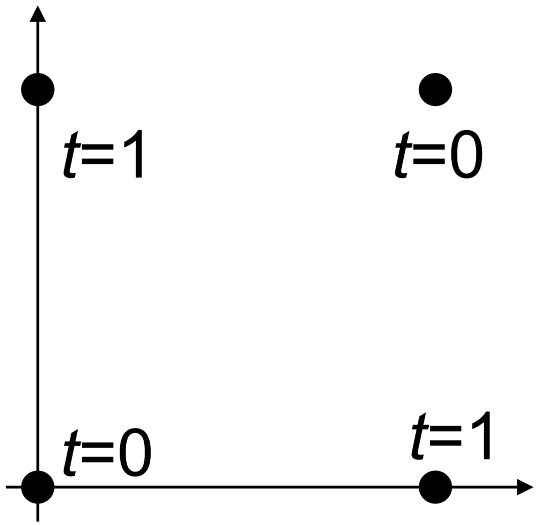
Using η=0.5,
subtract η**x** from **w** to give us
$\mathbf{w} = (-1.0, +2.0, -2.0)$



Perceptron decision boundary is now correct for all inputs.

Perceptron Limitations

- Problem must be *linearly separable*
- Classic non-linearly separable problem: *XOR problem*



- Minsky & Pappert (1969) - conjectured this limitation would *not* be overcome.
- But it was…