Adaptive Signal Processing and Machine Intelligence

Course Introduction

Danilo Mandic

room 813, ext: 46271

Department of Electrical and Electronic Engineering

Imperial College London, UK

d.mandic@imperial.ac.uk, URL: www.commsp.ee.ic.ac.uk/∼mandic

Office hours: Tuesdays 14:00 - 15:00

Data analytics $\looparrowright$ current status

- Computers excel at algorithmic tasks (well-posed mathematical problems)

- Biological systems are superior to digital systems for ill-posed problems with noisy data

- Pigeon: $\sim 10^9$ neurons, cycle time $\sim$ 0.1 seconds. Each neuron sends 2 bits to $\sim$ 1,000 other neurons. This is equivalent to 2×10^{13} bit operations per second

- Old PC: $\sim 10^7$ gates, cycle time 10^{-7} seconds, connectivity $= 2$ $\looparrowright 10^{15}$ bit operations per second

- Both have similar raw processing capability, but pigeons are better at recognition tasks

- Is there a way to present large date streams to computers in a more physically meaningful manner $\looparrowright$ **to make sense from Big Data**?

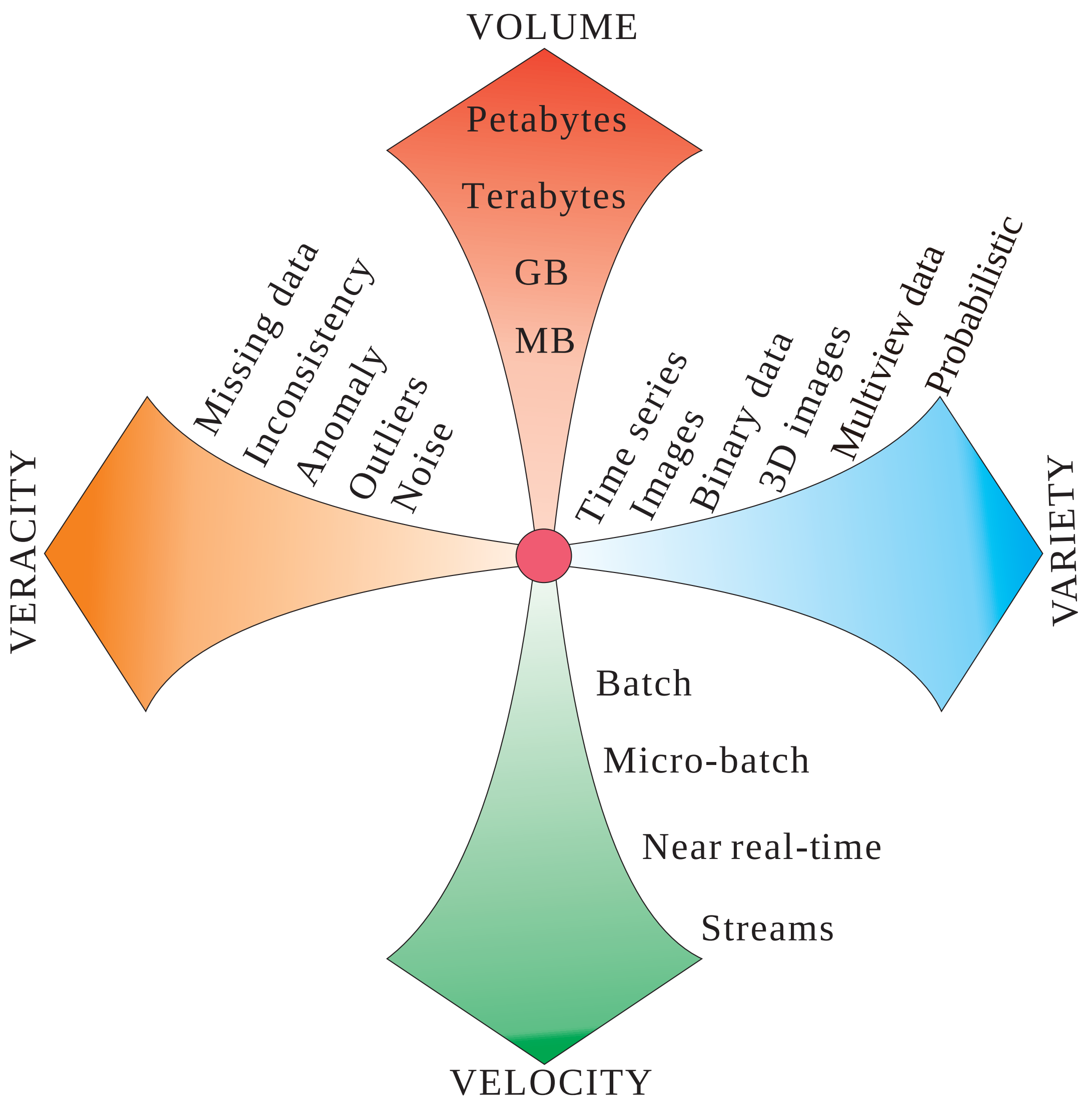
Some facts about Machine Intelligence opportunities

According to "Big Data: The next frontier for innovation, competition, and productivity", published by McKinsey Global Institute in May 2011:

- It would cost USD 600 to by a disk drive which can store all off the music in the world
- In 2010, there were 4 billion mobile phone users in the world
- There is more than 30 billion pieces of content shared on social networks every month
- There is a predicted 40 % growth in global data generated per year versus a 5 % growth in global IT spending
- This all tells us that there are big opportunities for us working in Adaptive Signal Processing and Machine Intelligence

The four V's of big data: Volume, Variety, Velocity, Veracity

Other V's may include Visualisation, Variability, Value (quality of data), ...



Adaptive Signal Processing & Machine Intelligence at Imperial

- **Dennis Gabor**, Nobel Prize laureate (1981 holography), theory of communications, nonlinear filters, Gabor kernels (1950s-1980s), setting the scene for neural networks, and kernel architectures

- **Colin Cherry**, 'cocktail party effect', 1950s - 1970s, setting the scene for 'Blind Source Separation' (BSS)

- **Anthony Constantinides**, 'digital filters', 'Lagrange Neural Networks', 'graph theory' for communications, 1970s - now, setting the scene for the digital processing of real world signals

- **Igor Aleksander**, 'binary neural networks', 1980s - now, early 'deep networks' with millions of neurons, artificial consciousness

- **Abe Mamdani**, a pioneer of fuzzy logic (in 1970s), his fuzzy inference has characteristics of human instincts

Our alumni are in leading technical positions around the world

Our “intellectual capital” at EEE

The experimental setup for Gabor’s Hologram
http://www.nobelprize.org/nobel_prizes/physics/articles/biedermann/

Spectral transformations for digital filters

A. G. Constantinides, B.Sc.(Eng.), Ph.D.

Indexing term: *Digital filters*

Abstract

The paper describes certain general transformations for digital filters in the frequency domain. The term digital filter is used to denote a processing unit operating on a sampled waveform, so that the input, output and intermediate signals are only defined at discrete intervals of time; the signals may be either p.a.m. or p.c.m. The transformations discussed operate on a lowpass-digital-filter prototype to give either another lowpass or a highpass, bandpass or band-elimination characteristic. The transformations are carried out by mapping the lowpass complex variable z^{-1} [where $z^{-1} = \exp(-j\omega T)$ and T is the time interval between samples] by functions of the form

$$e^{j\theta} \prod_{i=1}^{n} \frac{z^{-1} - \alpha_i}{1 - \alpha_i^* z^{-1}}$$

known as unit functions.

THEORY OF COMMUNICATION*

By D. GABOR, Dr. Ing., Associate Member.†

(*The paper was first received 25th November, 1944, and in revised form 24th September,* 1945

PREFACE

The purpose of these three studies is an inquiry into the essence of the “information” conveyed by channels of communication, and the application of the results of this inquiry to the practical problem of optimum utilization of frequency bands.

In Part 1, a new method of analysing signals is presented in which time and frequency play symmetrical parts, and which contains “time analysis” and “frequency analysis” as special cases. It is shown that the information conveyed by a frequency band in a given time-interval can be analysed in various ways into the same number of elementary “quanta of information,” each quantum conveying one numerical datum.

In Part 2, this method is applied to the analysis of hearing sensations. It is shown on the basis of existing experimental material that in the band between 60 and 1 000 c/s the human ear can discriminate very nearly every second datum of information, and that this efficiency of nearly 50% is independent of the duration of the signals in a remarkably wide interval. This fact, which cannot be explained by any mechanism in the inner ear, suggests a new phenomenon in nerve conduction. At frequencies above 1 000 c/s the efficiency of discrimination falls off sharply, proving that sound reproductions which are far from faithful may be perceived by the ear as perfect, and that “condensed” methods of transmission and reproduction with improved waveband economy are possible in principle.

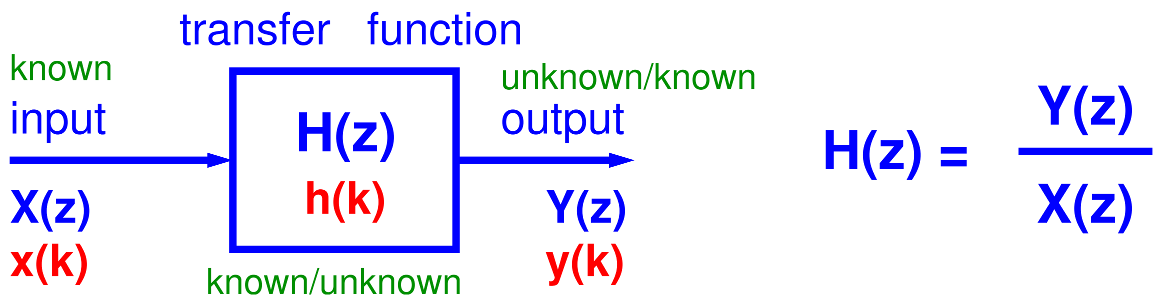
In Part 3, suggestions are discussed for compressed transmission and reproduction of speech or music, and the first experimental results obtained with one of these methods are described.

Learning from stochastic signals and adaptive systems

We will consider adaptive, multidimensional, spectral and neural estimators

So far, you are familiar with problems for which:

- We have a **well defined transfer function** in the form



  which is **parametric** (model based), e.g. the stochastic AR model
  $$\hat{x}(k) = \Phi\big(a_1(k)x(k-1) + \cdots + a_p(k)x(k-p)\big) + w(k), \qquad w \sim \mathcal{N}(0,1)$$
- **Data are mostly stationary**

In this course we will introduce models which are:

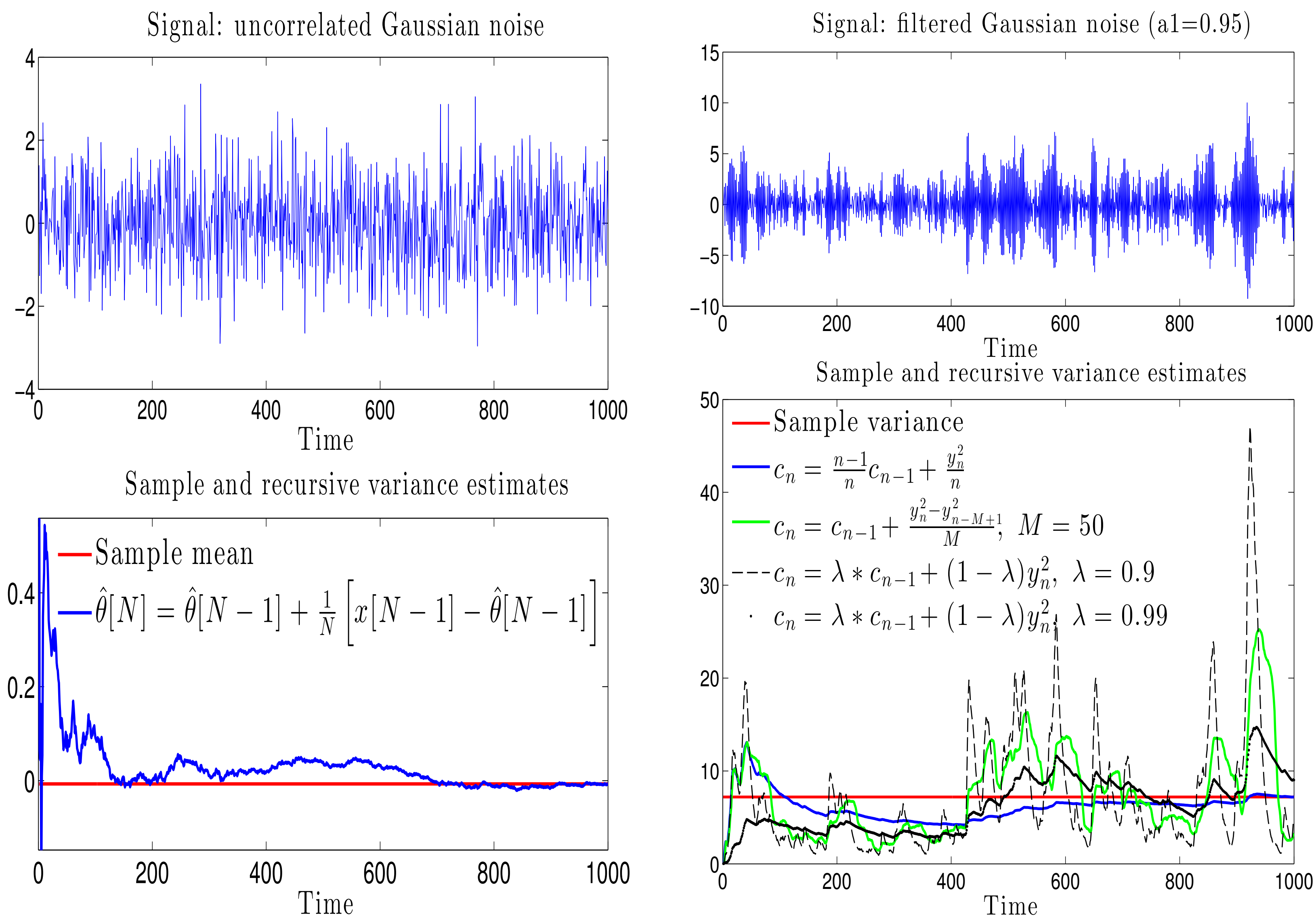
⊛ **Nonparametric**, that is, they do not assume any model *a priori*

⊛ **Adaptive**, capable of tracking the changes in the system/signal parameters in real time and in nonstationary environment

These are huge advantages that allow such models to operate in an online fashion and for real world data.

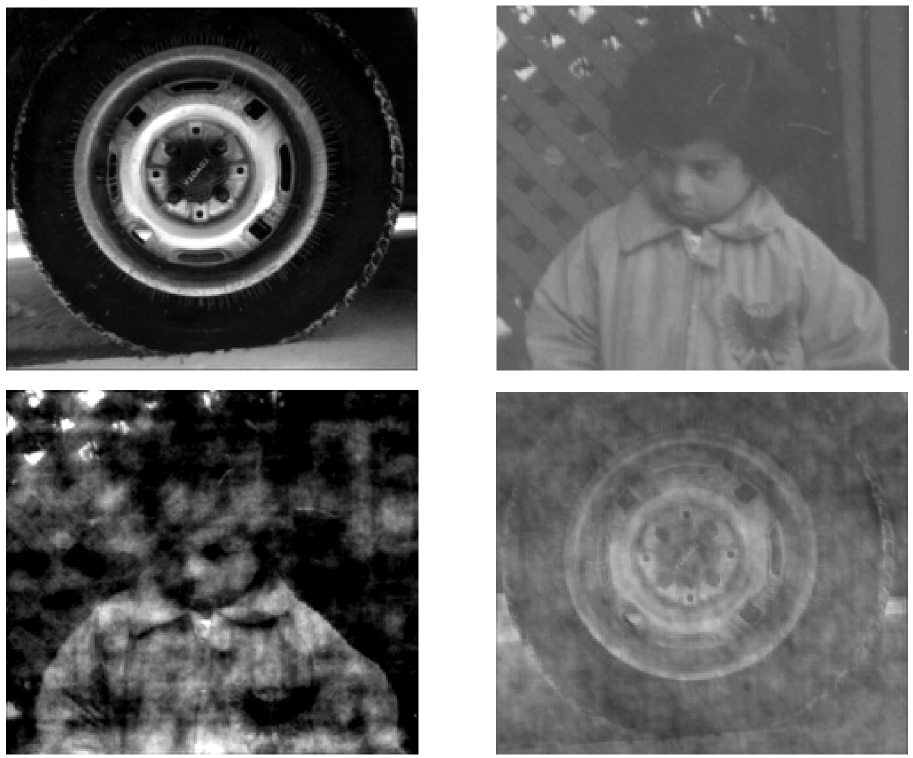
Example: From block to real-time estimators

Top left: uncorrelated Gaussian   Top right: correlated Gaussian   Bottom: corresponding variance estimates



Human perception & Machine Intelligence
Illustrative problem: “Quality of Experience” (QoE)

Statistical properties of conventional and model-based **spectral estimation** techniques. Can learning algorithms incorporate models of human perception $\rightsquigarrow$ the role of phase spectrum?

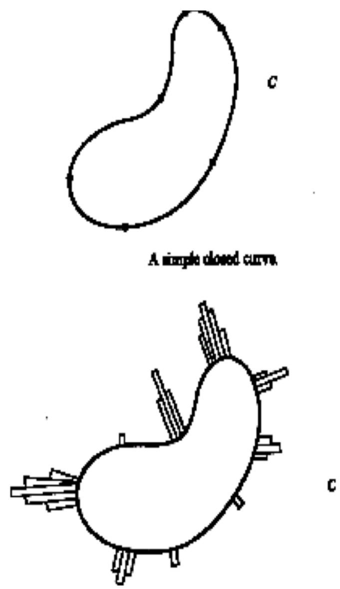


Surrogate images. *Top:* Original images I_1 and I_2; *Bottom:* Images $\hat{I}_1$ and $\hat{I}_2$ generated by exchanging the amplitude and phase spectra of the original images.

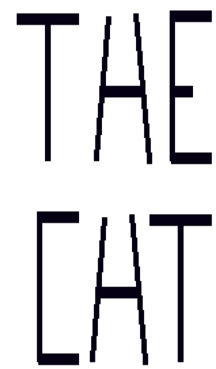
Sufficient information $\looparrowright$ humans vs machines

In the 50's phychologist Fred Attneave recorded eye dwellings on objects

A closed curve (top) and the histogram of eye dwellings (bottom)

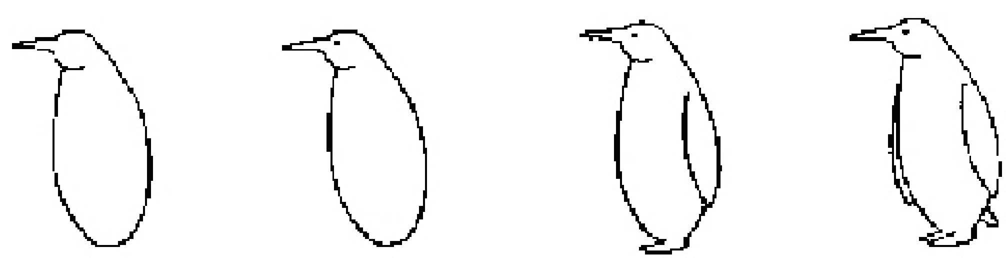


Beyond sparsity: What aspects of information can be compromised?



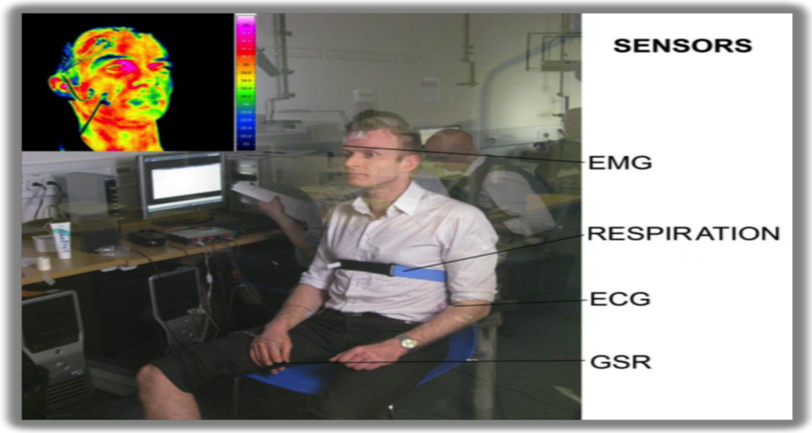
Marr & Hildreth (1980)

Is the drawing on the left still a penguin?



A challenge for ASP and MI in the 21st century
blurring the boundaries between ability and disability

Complexity in rehabilitation



Affective computing

Concierto for brain and chamber quartet

Spectral analysis: Towards instantaneous frequencies

Latin Specter means a ghostly apparation, English spectre

- The word *spectrum* introduced by Newton in relation to his studies of the decomposition of white light into a band of light colours, when passed through a glass prism

- Spectral analysis as an established and ever expanding discipline $\looparrowright$ we are currently working on time-frequency estimation on nonlinear and nonstationary data

- Beginning about one century ago with the work by Schuster on detecting cyclic behaviour in time series

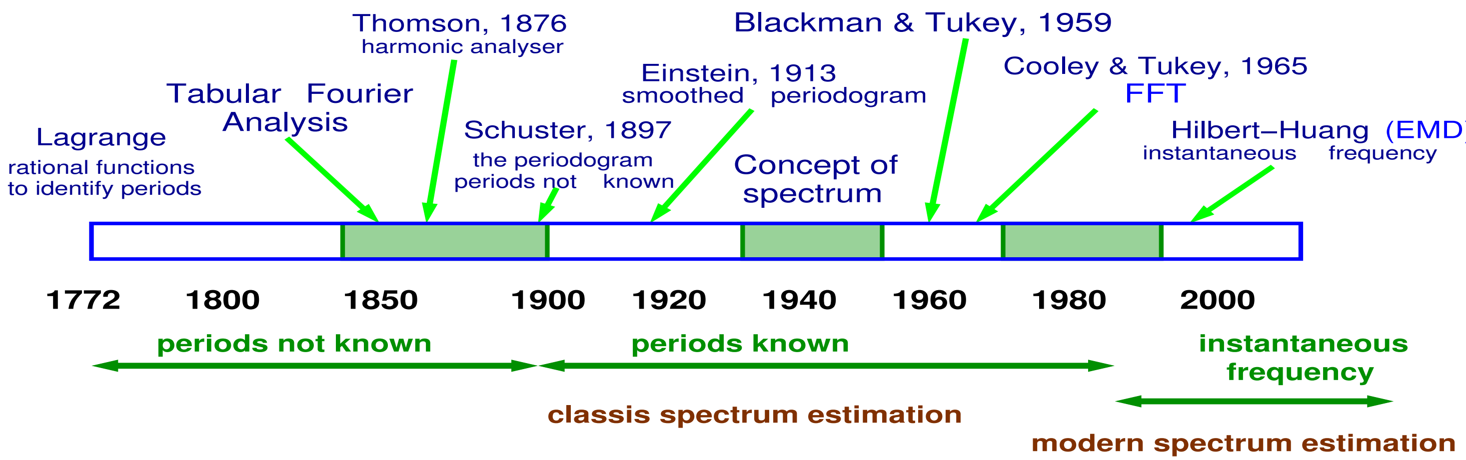
- Omni–present now (genomics, financial engineering, cognitive radio)

Despite the roots of the word spectrum, I hope the students will be a vivid presence in this course.

Spectrum estimation: The journey

From observing periodic and planetary motions

- Everywhere around us: the concept of frequency and sinewave



- Intimately related with the concept of complex numbers *(fundamental theorem of algebra)*
- Fourier's work from around ∼1800, FFT - mid 1960s, Lomb periodogram for irregularly sampled data (1972)

Recent advances in **Spectrum Estimation:** ⊛ irregularly sampled data, ⊛ very few data points (genomics and proteomics), ⊛concept of **instantaneous frequency**, ⊛ spectrum estimation of nonstationary data

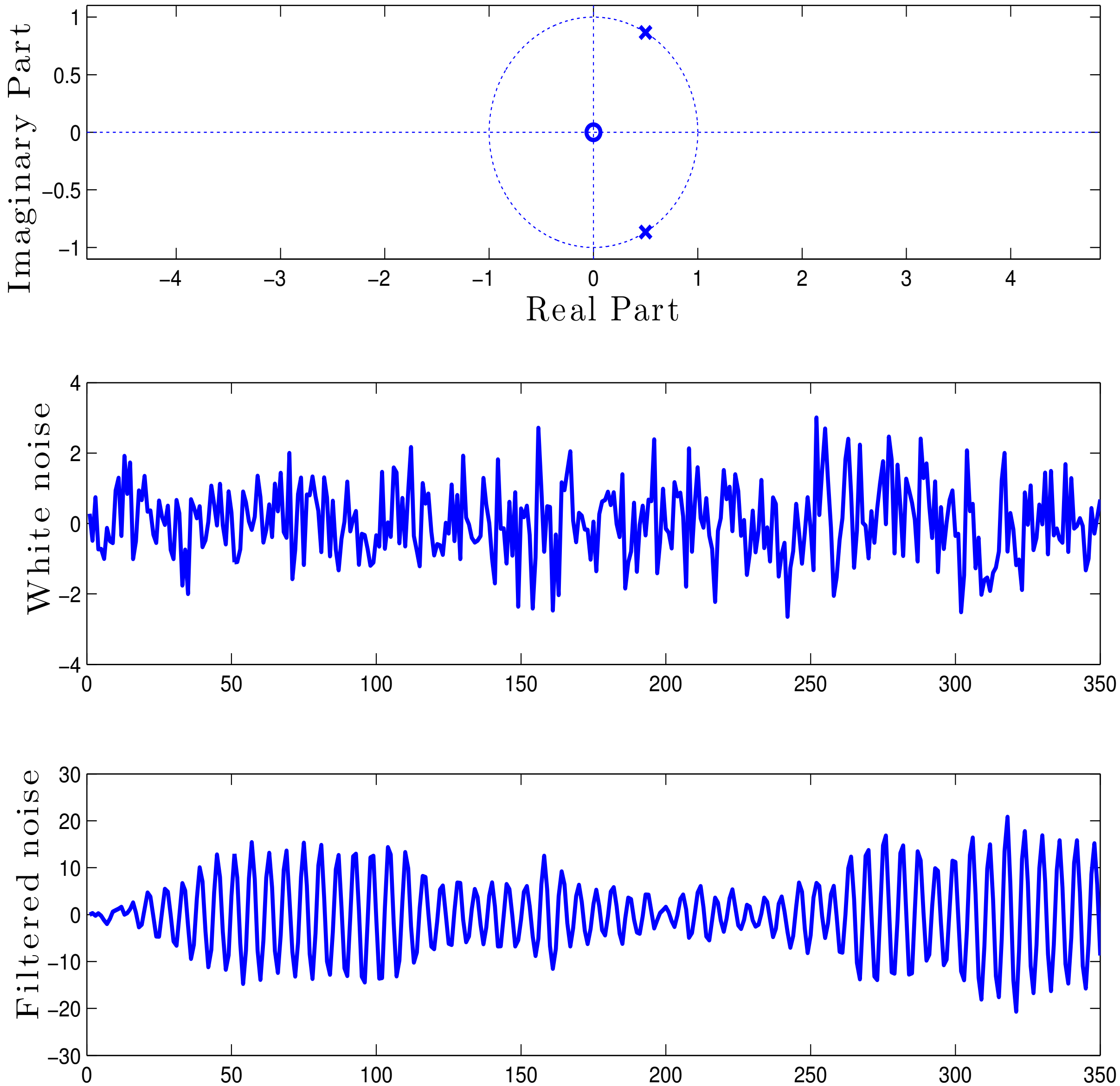
ASP & MI: The beginnings

The creation of the LMS algorithm (Widrow 1960)

- interference cancellation in telephone lines (in maths NLMS from 1927 Kaczmarz)
- complex least mean square (CLMS) 1975
- affine projection and proportionate NLMS (revolution in acoustics) - 1990s-2000s
- magnitude-only LMS, phase-only LMS (late 1990s onwards)
- widely linear (augmented) CLMS in the 2000s
- cooperative estimation over sensor networks, mid-2000s onwards
- quaternion LMS (QLMS) for 3D and 4D data (e.g. wind) - 2009
- Kalman filter: early 1960s until now (extended, unscented, particle)
- blind source separation early 1990s onwards
- neural networks: 1947, connectionsm 1986, reservoir computing "Echo State Networks" 2004 onwards, deep learning 2010-, Big Data 2012-

Spectral vs. adaptive processing: A real-world sinewave?

Shall we use spectral estimation or adaptive filtering to estimate it?



Matlab code:

```
z1=0;
p1=[0.5+0.866i,0.5-0.866i];
[num1,den1]=zp2tf(z1,p1,1);
zplane(num1,den1);
s=randn(1,1000);
s1=filter(num1,den1,s);
figure;
subplot(311),plot(s),
subplot(313),plot(s1),
subplot(312),;
zplane(num1,den1)
```

The AR model of a sinewave

```
x(k)=a1*x(k-1)+a2*x(k-2)+w(k)
a1=-1, a2=0.98, w~N(0,1)
```

Machine-learned Fourier transform

We can see FT as a convolution of a complex exponential and the data (under a mild assumption of a one-sided h sequence, ranging from 0 to ∞)

1) Continuous FT. For a continuous FT $\quad F(\omega) = \int_{-\infty}^{\infty} x(t) e^{-\jmath\omega t} dt$

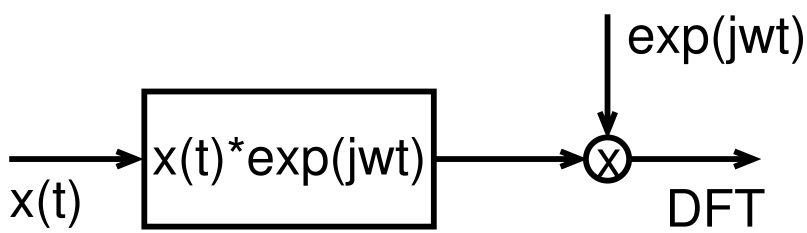
Let us now swap variables $t \to \tau$ and multiply by $e^{\jmath\omega t}$, to give

$$e^{\jmath\omega t} \int x(\tau) e^{-\jmath\omega\tau} d\tau = \int x(\tau) \underbrace{e^{\jmath\omega(t-\tau)}}_{h(t-\tau)} d\tau = x(t) * e^{\jmath\omega t} \qquad (= x(t) * h(t))$$

2) Discrete Fourier transform. For DFT, we have a filtering operation

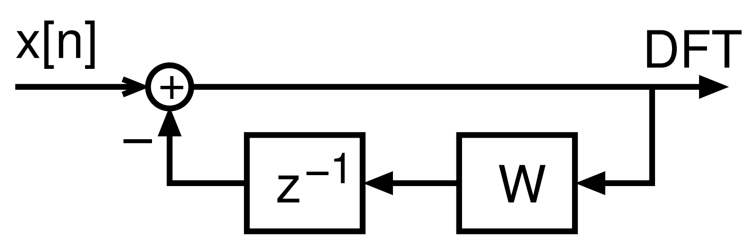
$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\jmath\frac{2\pi}{N}nk} = \underbrace{x(0) + W\Big[x(1) + W\big[x(2) + \cdots\Big]}_{\text{cumulative add and multiply}} \qquad W = e^{-\jmath\frac{2\pi}{N}n}$$

with the transfer function (large N) $\quad H(z) = \frac{1}{1-z^{-1}W} = \frac{1-z^{-1}W^*}{1-2\cos\theta_k z^{-1}+z^{-2}}$



continuous time case

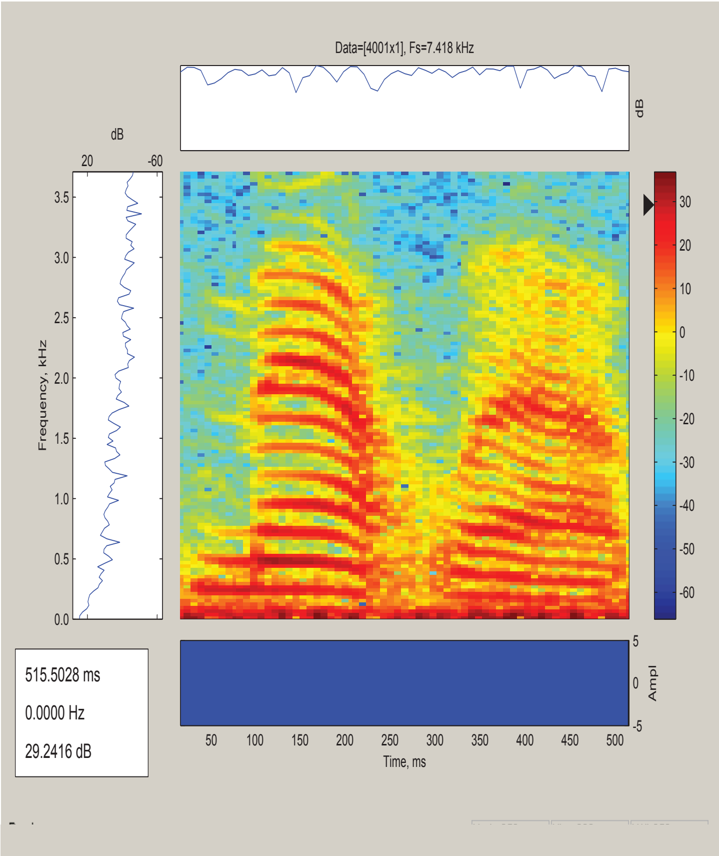
discrete time case

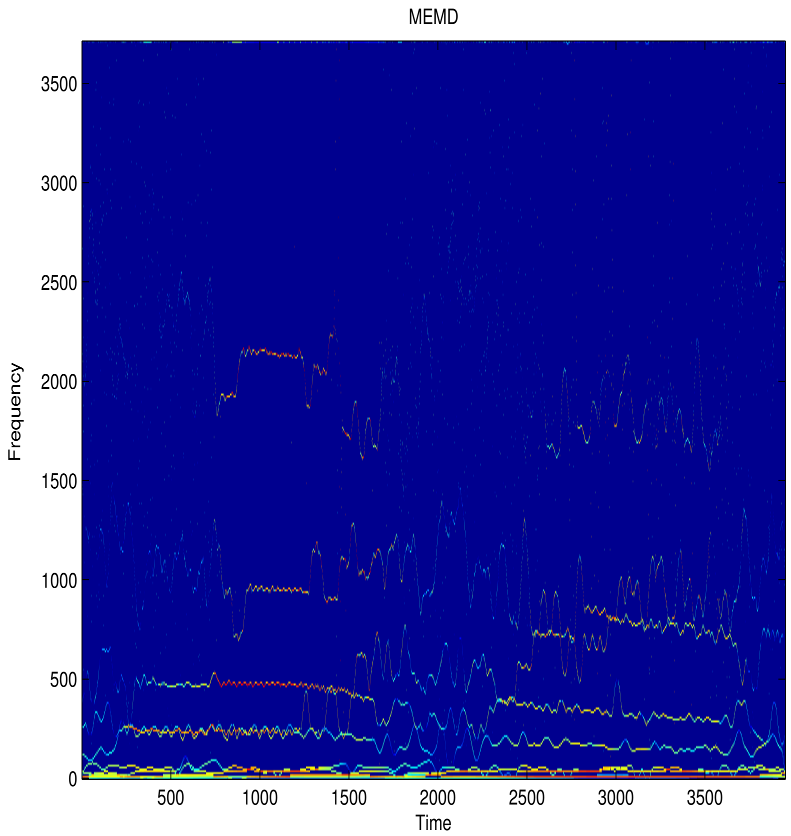


Seeing through mathematical artefacts $\looparrowright$ instantaneous frequency

A time-frequency representation of speech: "Matlab"

(STFFT spectrogram) **(Hilbert-Huang spectra)**



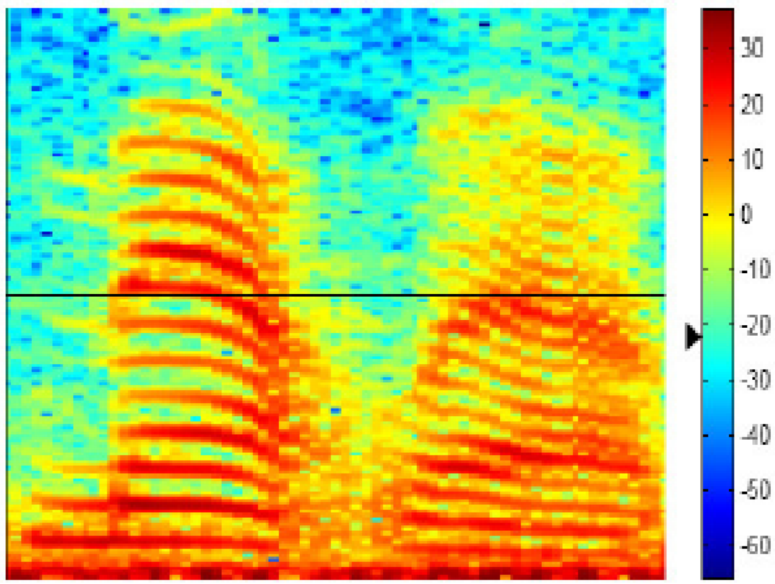


(win-len=256, overlap=200, ftt-len=256) **(no. of directions=64)**

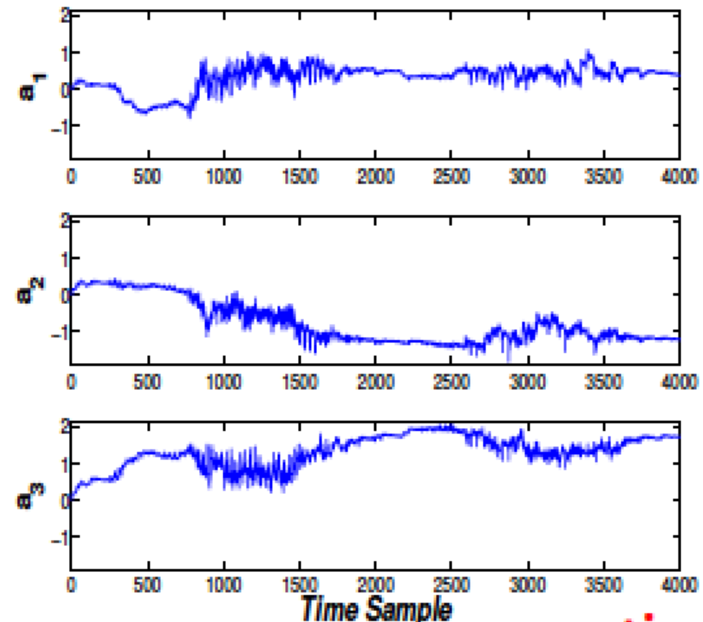
Real time estimation of signal statistics and parameters

In Matlab: S = SPECTROGRAM(X,WINDOW,NOVERLAP,NFFT,Fs)

Frequency **Filter coeff.** **mtlb_filtercoeffs.m**



time



time

M aaa t l aaa b **M aaa t l aaa b**

Time-frequency spectrogram: *Stack PSDs together, $\forall n$*

Darker areas: higher magnitude

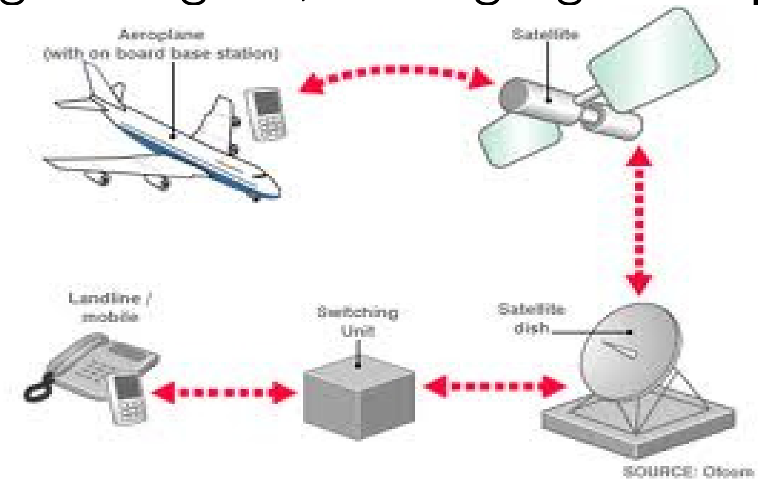
Evolution of AR coefficients: *They follow the signal statistics*

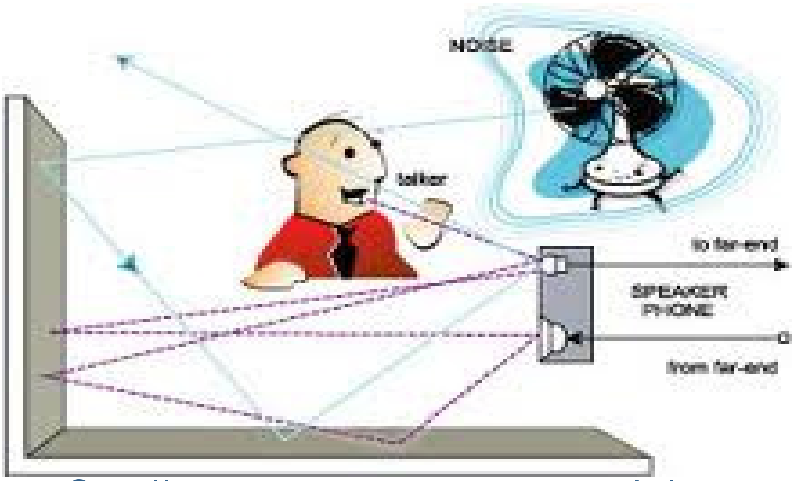
Vowels: more dynamics

Adaptive signal processing

(also adaptive learning systems and the concept of neural networks)

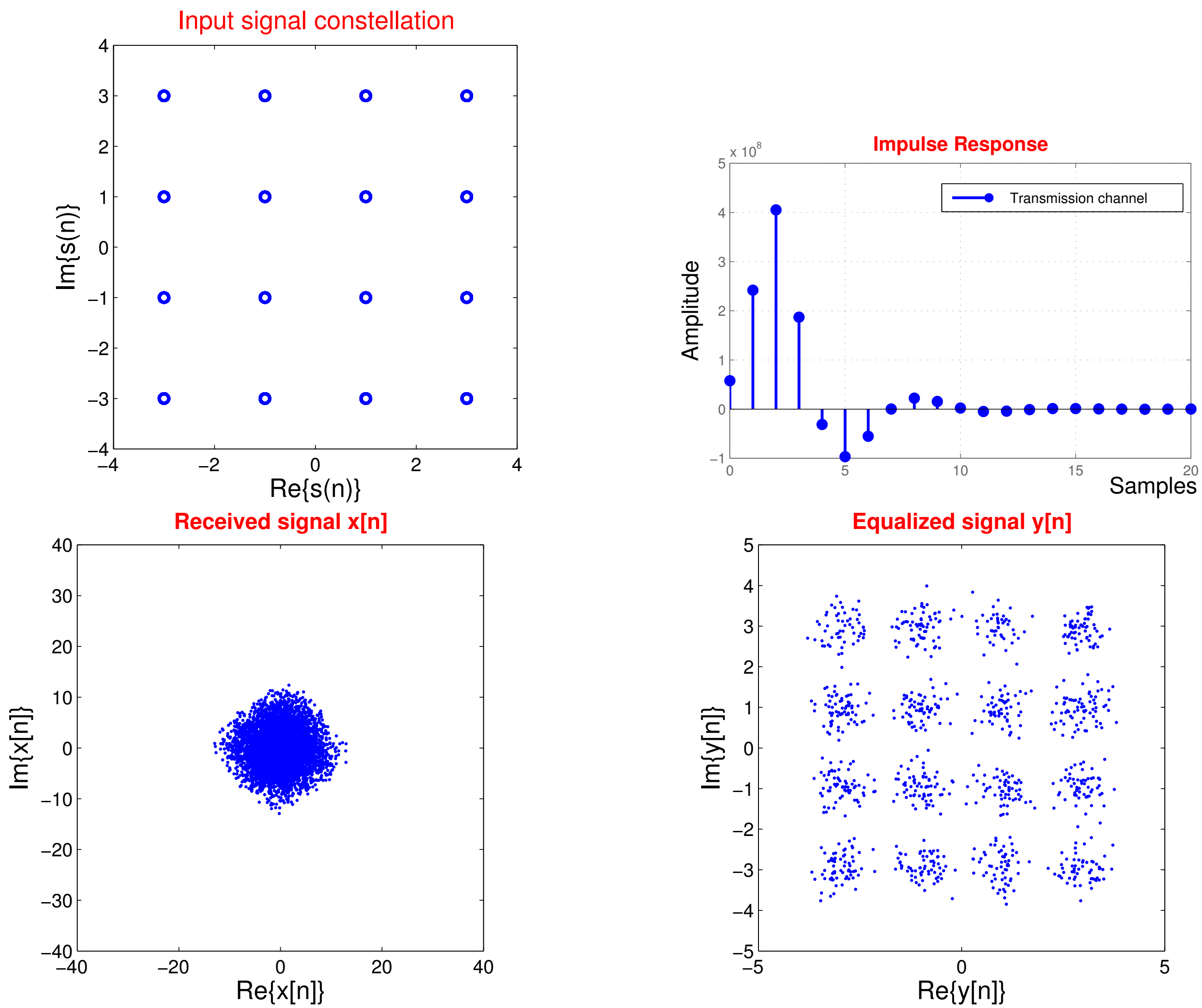
We will motivate the need for **adaptive signal processing** when processing real signals, and highlight its applications





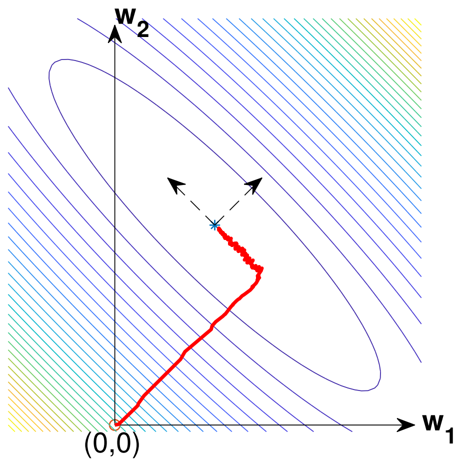
Applications. Satellite communications, mobile communications, finance, audio.

Example: Equalisation in digital communications

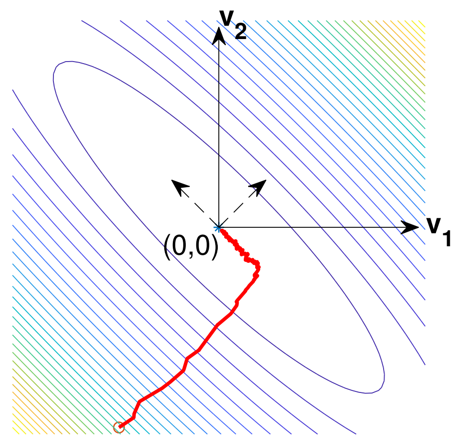


Convergence considerations: Speed and interpretability

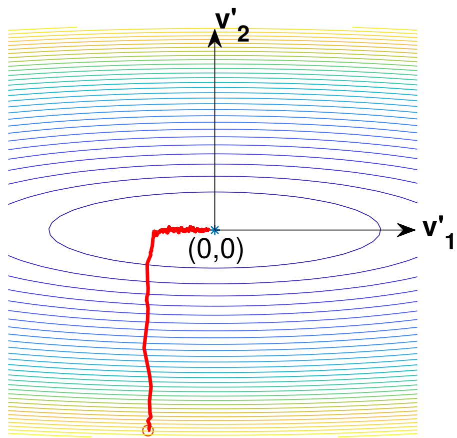
Original weight space, $\mathbf{w}$



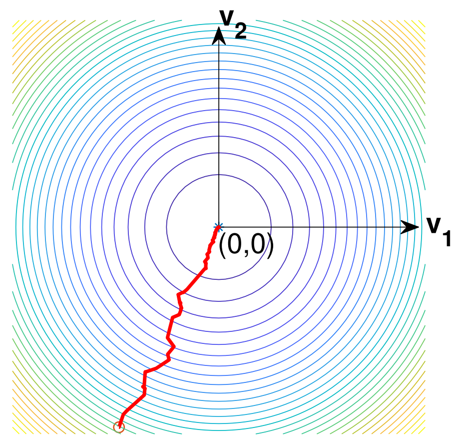
Weight error space, $\mathbf{v} = \mathbf{w} - \mathbf{w_{o}}_{pt}$



Weight errors in the rotated axes, $\mathbf{v}'$

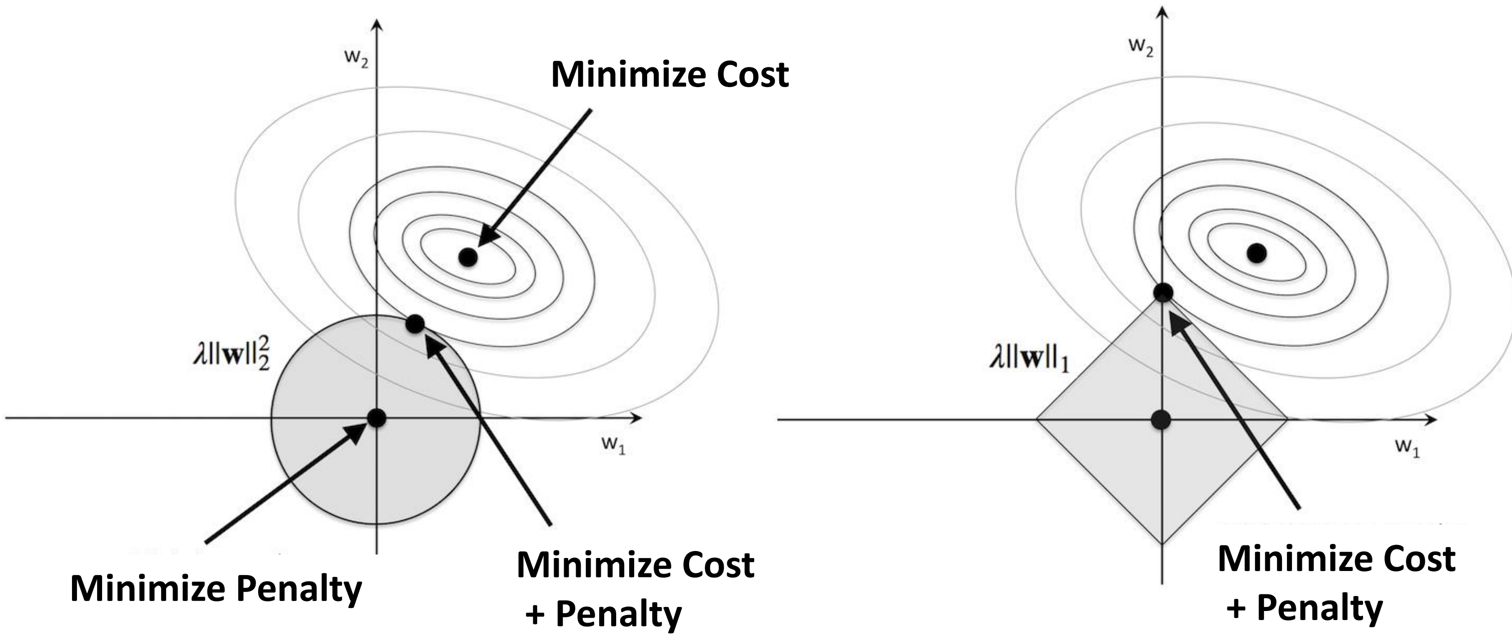


Weight errors for whitened data



Emphasis on efficiency vs expressivity

$\mathcal{J}_k(\mathbf{w}) = \underbrace{(d_k - \mathbf{w}^T\mathbf{x}_k)^2}_{\text{Cost}} + \underbrace{\lambda\|\mathbf{w}\|_p}_{\text{Penalty}}$ with Regularisation Parameter: λ.



Tikhonov: $\ell = 2$

Lasso: $\ell = 1$

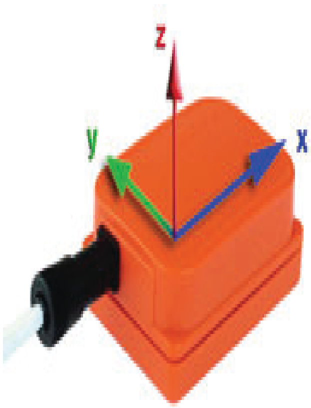
(Figure credit: Mlxtend, S. Raschka, 2016)

Sparsity promoting

For $\lambda \neq 0$ the **minima of the regularised cost functions do not correspond** to the squared error cost.

Applications: From environment to the “quantified self”

ASP and MI are benefitting enormously from the progress in sensor technology



Renewable Energy

2D and 3D anemometers

control of wind turbine

Body motion sensor

3D - position, gyroscope, speed

gait, biometrics

Wearable technologies

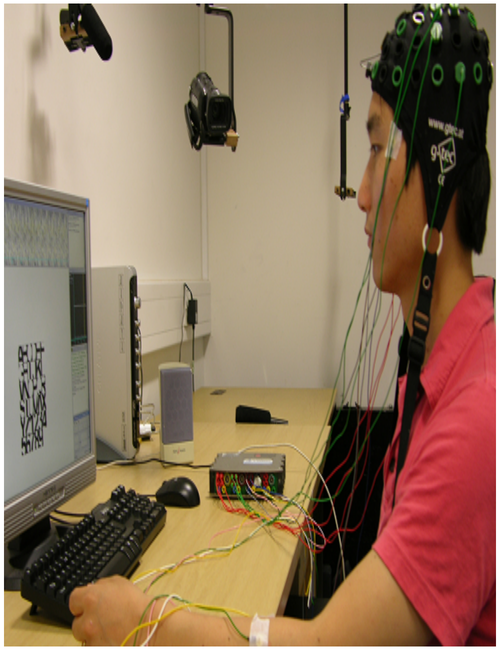
Biomechanics

virtual reality

Wind sensors - 3D anemometer

challenges for multidimensional DSP (e.g. quaternion division algebra)

Numerous other applications



Brain Computer Interface

Decoding brain activity

to control computers

Spect. Est., ASP

Medical Applications

3D time-space

2D and 3D electromagnetic field

ASP, MI,

Avionics

Trajectory tracking

Radar: Manoeuvre prediction

ASP, MI

The need for two- and multi-dimensional representations

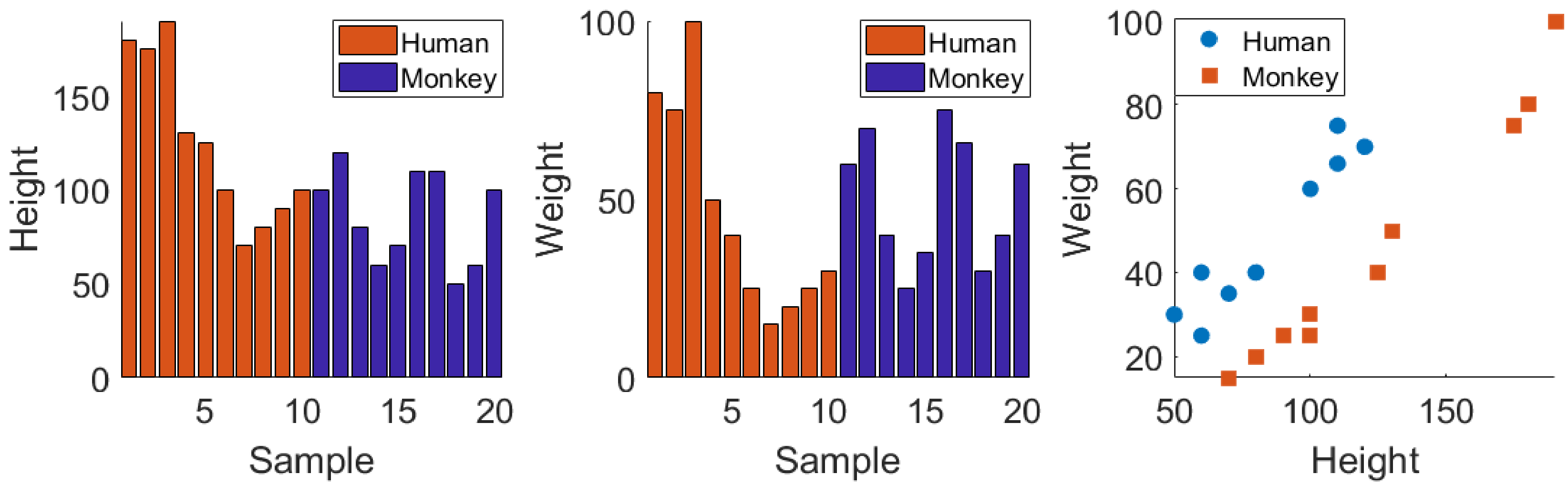
Q1: Can humans and other primates be distinguished by height?

Q1: Can humans and other primates be distinguished by weight?

A: There is a clear overall trend, and also the average heights and the average weights are different between classes.

☞ It is not possible to distinguish between these classes based on **univariate** data (height or weight)

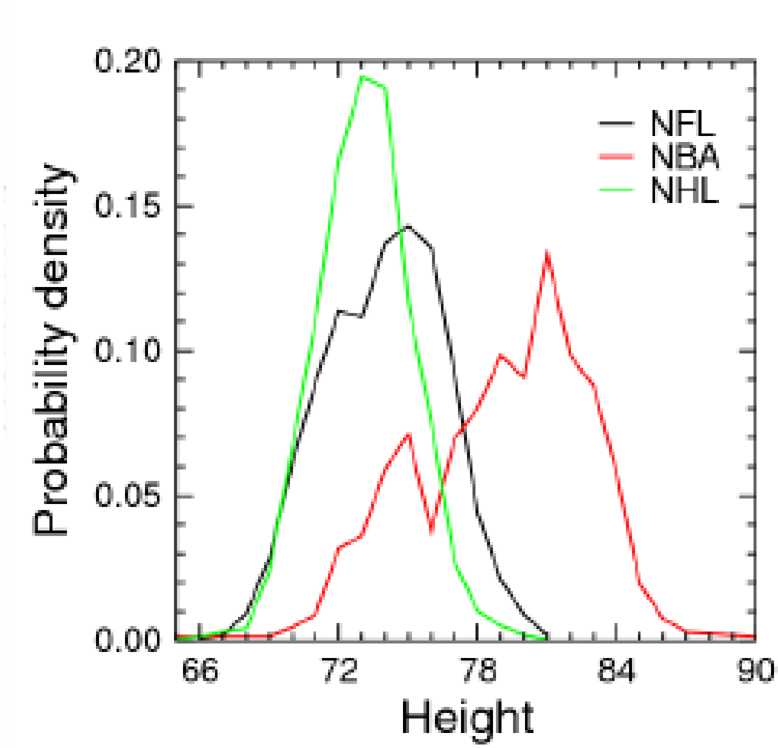
However, even a simplest **bivariate approach (scatter plot)** suffices



Enhanced discrimination ability owing to more degrees of freedom $\looparrowright$ how about a multivariate case?

Consider the histograms (empirical pdf's) of the height of athletes: American football, ice hockey, basketball

In this **trivariate case**, how likely it is that a linear discriminator would be appropriate?



Conditional pdf (can we e.g. estimate the weight given the height) "slice and normalise" the joint pdf $p(x, y)$

Formal definition

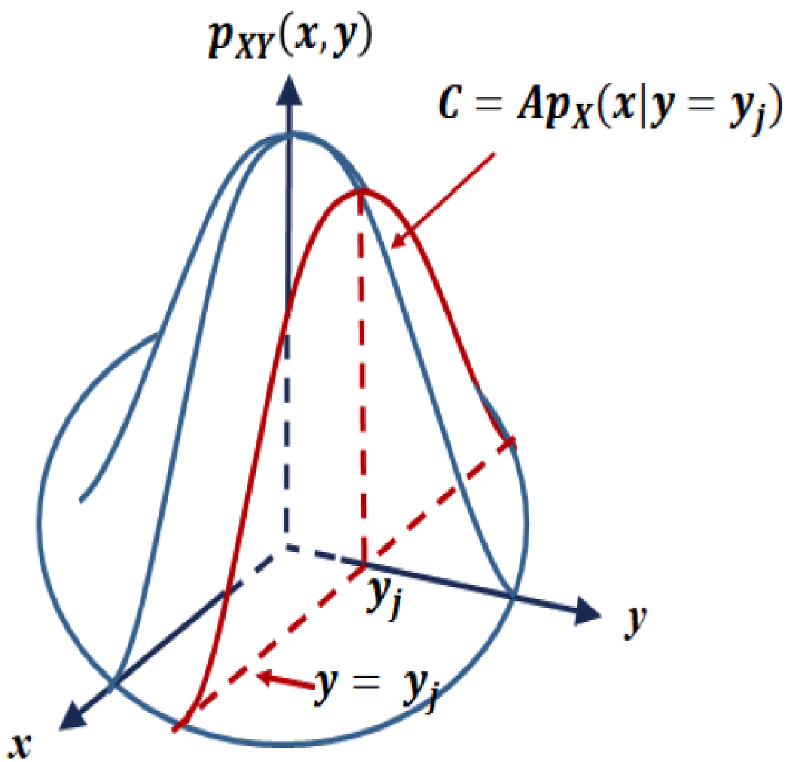
$$p_{Y|X}(y|x) = \begin{cases} \frac{p_{XY}(x,y)}{p_X(x)}, & p_X(x) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

x is held fixed

or more often

$$p(x|y) = \begin{cases} \frac{p(x,y)}{p(y)}, & p(y) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

y is held fixed



Conditional pdf $p(x|y)$

Depends on joint pdf $p(x, y)$ because there are two rand. variables, x and y.

Example: Length of holidays, X, conditioned on the salary $Y =$£60k?

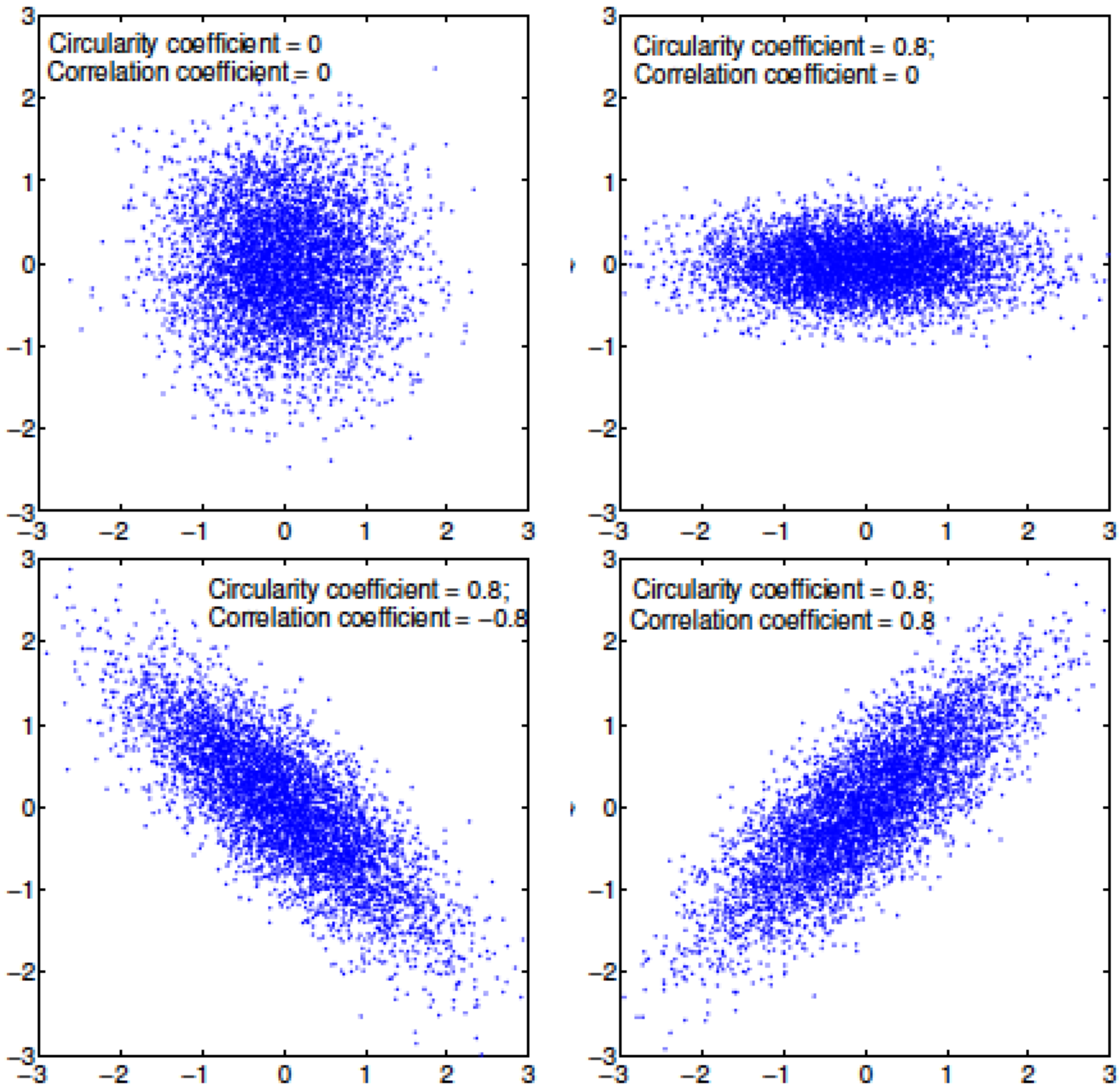
Ans: Find all people who make exactly £60k, how is holiday length distributed?

We therefore:

- **slice** the joint $p(x, y)$ at $Y = £60k$
- **normalise** by $p_Y(60,000)$ so that $p(x|y){=}p(x, 60k)/p_Y(60k)$ is valid pdf

Bringing scatter diagrams into learning machines

Complex noncircularity $\looparrowright$ unique signature of a signal $\in \mathbb{C}$



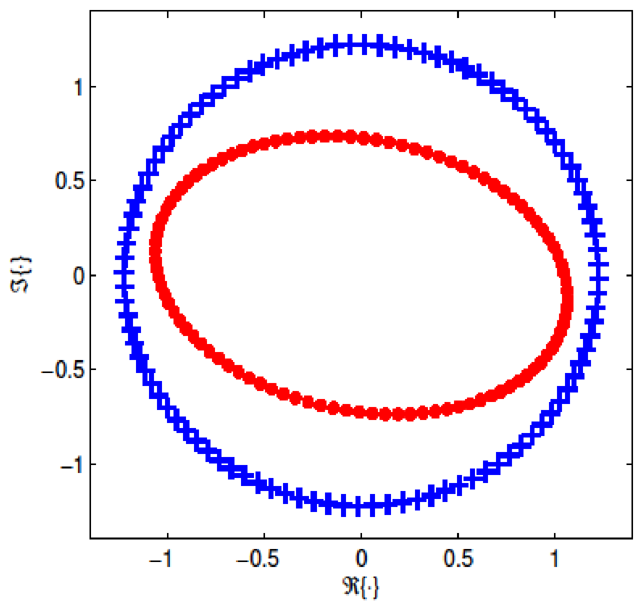
So, the degree of circularity can be used as a **fingerprint** of a signal, allowing us enormous additional freedom in estimation, compared with standard **strictly linear** systems.

For instance, we can now differentiate between different Gaussian signals!

Recall: Real valued ICA cannot separate two Gaussian signals.

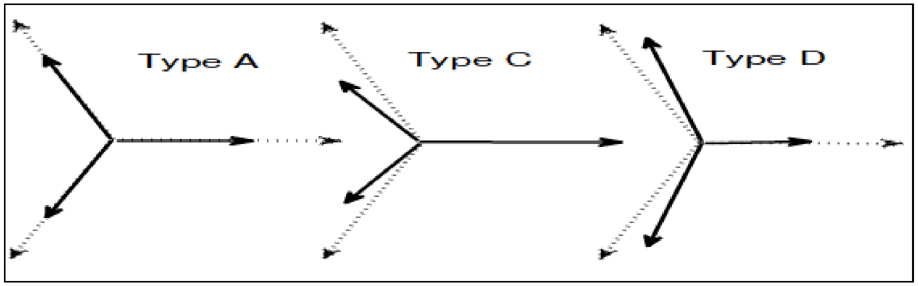
Application: Complex noncircularity and smart grid

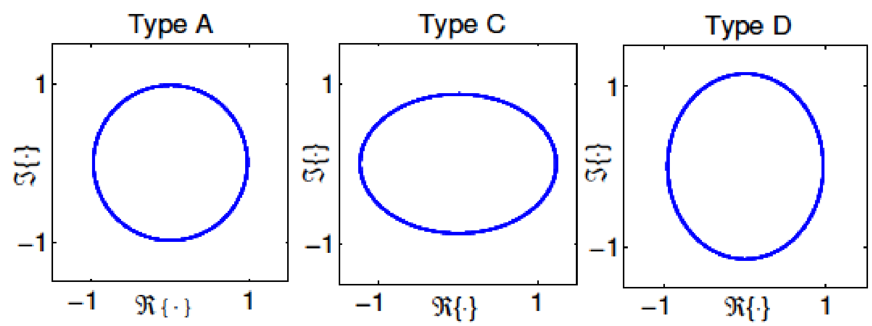
- **Balanced system:** $V_a(k) = V_b(k) = V_c(k)$, $A(k) = \text{const}$, $B(k) = 0$, and $v(k)$ is on a circle
- **Unbalanced system:** $V_a(k)$, $V_b(k)$, $V_c(k)$ are not identical
  - ⊛ $A(k)$ is no longer constant, $B(k) \neq 0$
  - ⊛ $v(k)$ is not on a circle $\rightarrow$ **a degree of noncircularity**



balanced and unbalanced system

Some types of voltage sags

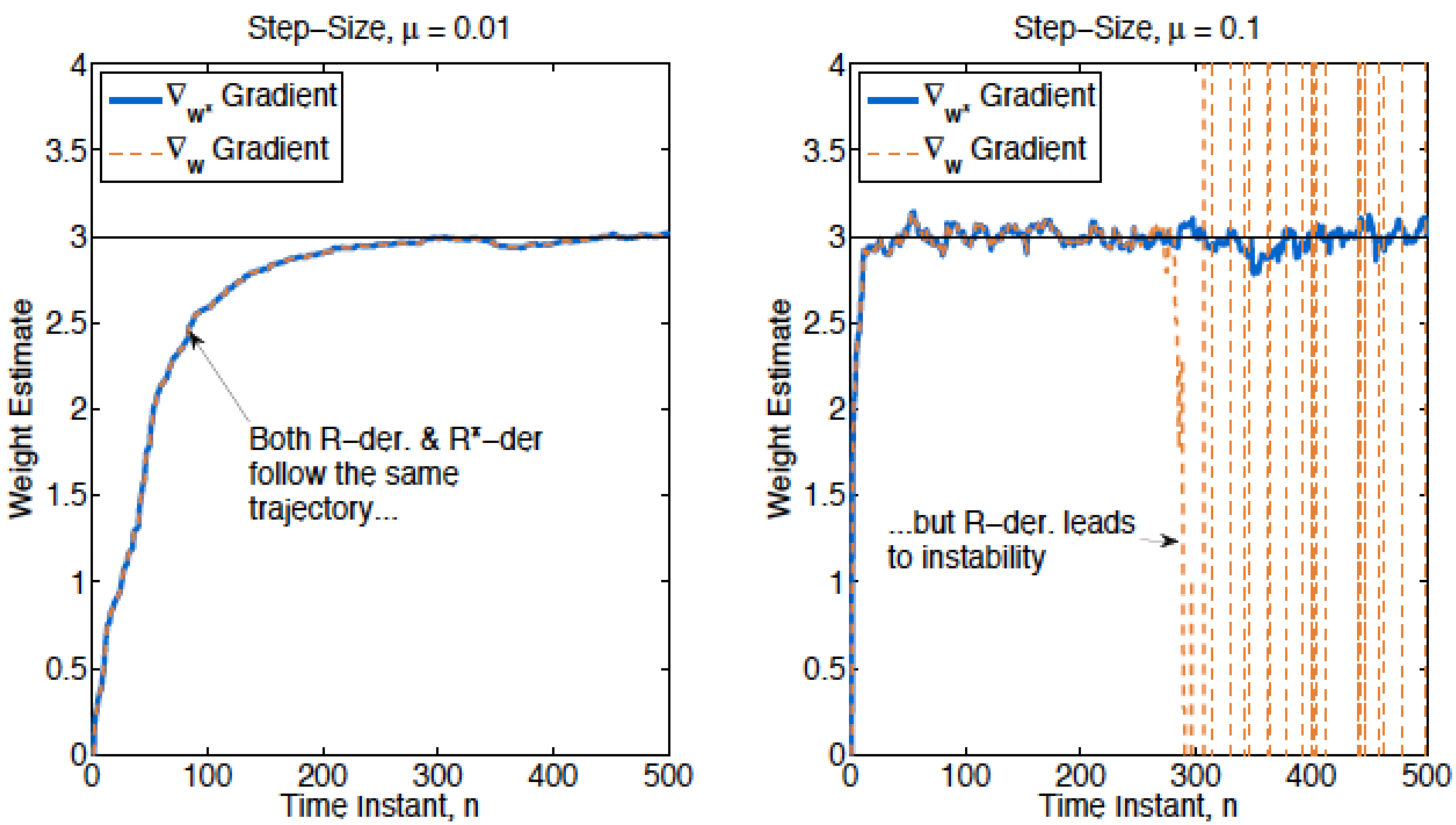




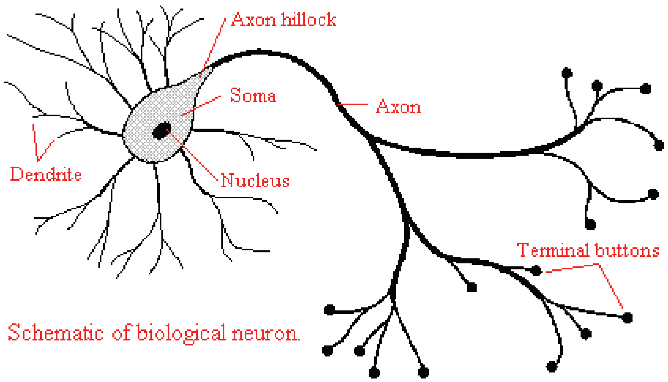
Their circularity diagrams

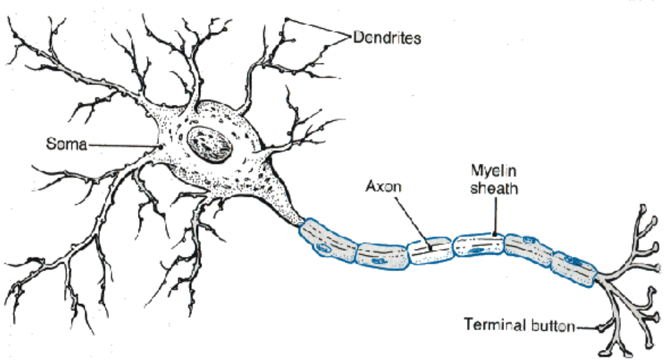
Adaptive learning systems $\looparrowright$ data + architecture + learning algorithm

Simulation for the CLMS derived using $\mathbb{R}$-der. and $\mathbb{R}^*$-der. $(\mathrm{w}_o = 3)$

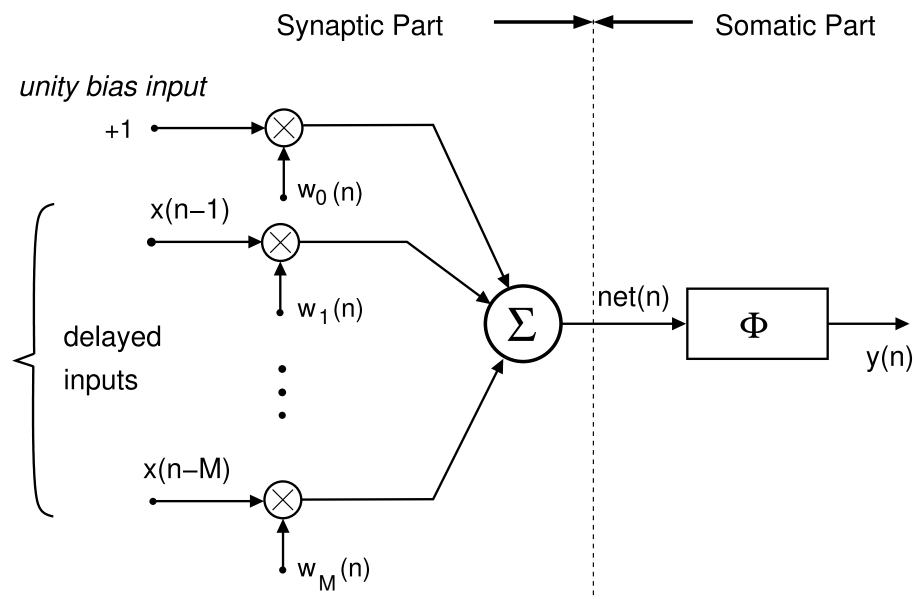


From a biological neuron to an artificial neuron model
Perceptron learning algorithm





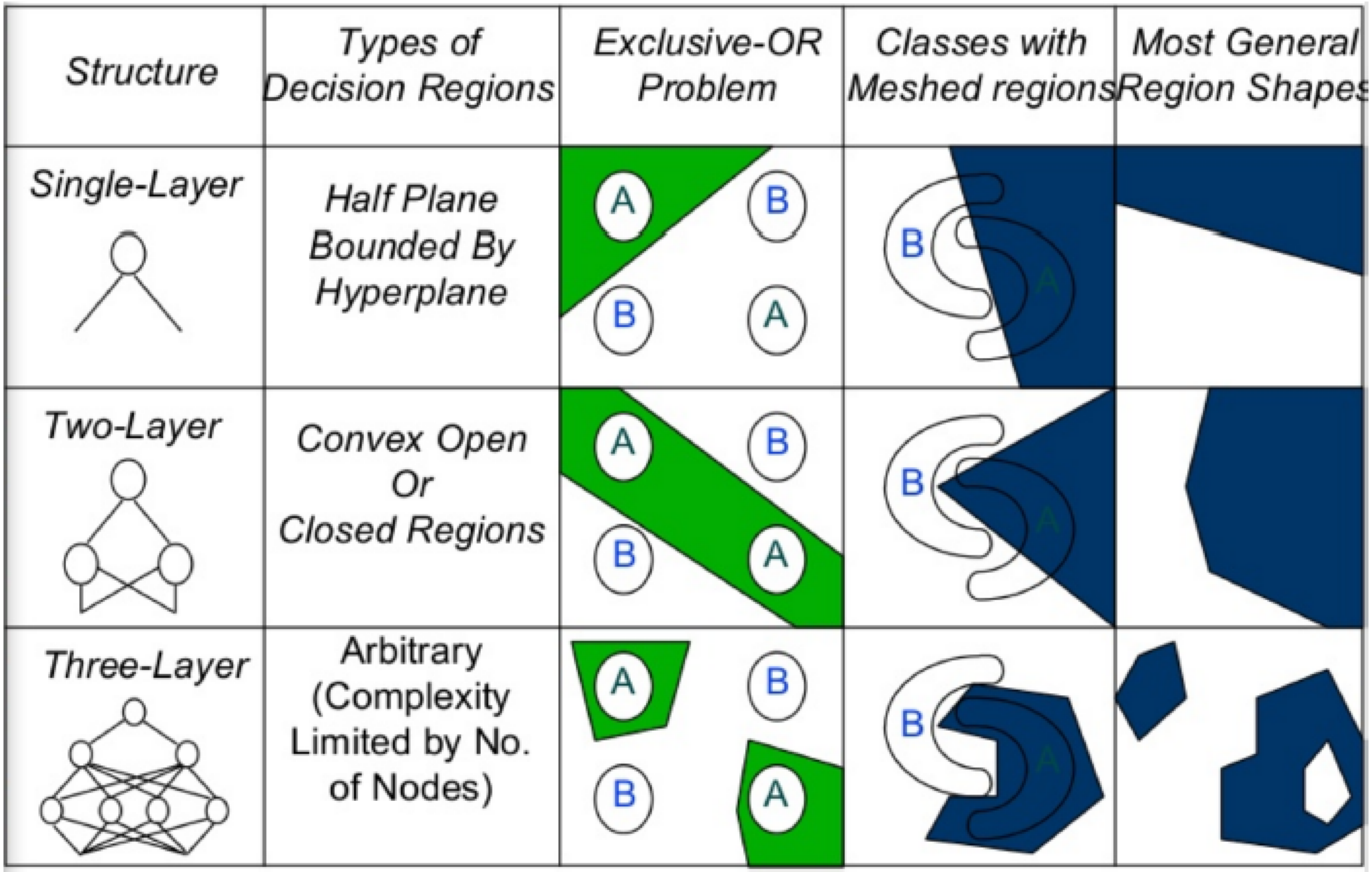
Biological neuron



Model of an artificial neuron

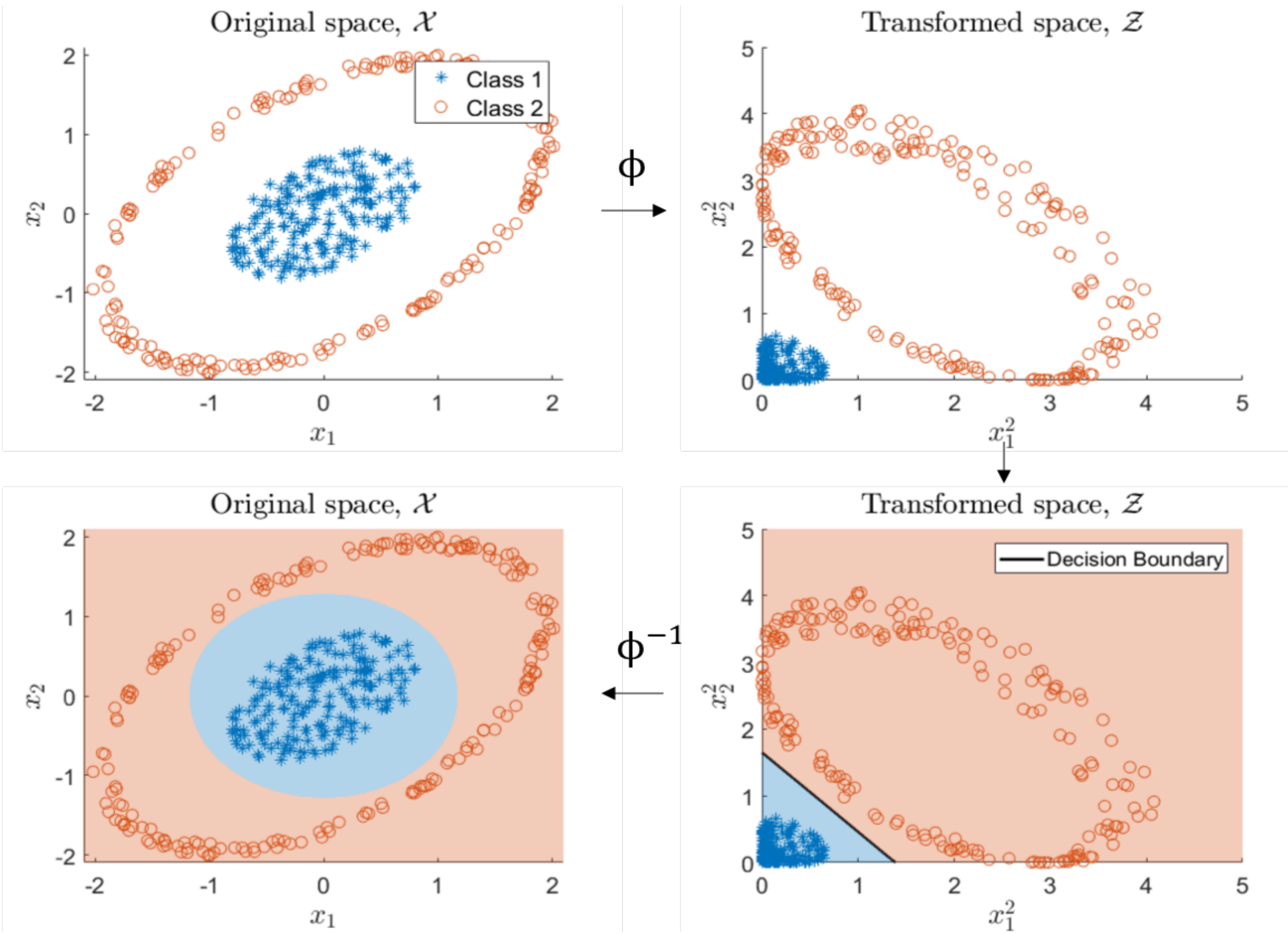
- delayed inputs $\mathbf{x}$
- bias input with unity value
- sumer and multipliers
- output nonlinearity

From a single neuron to artificial neural networks and nonlinear estimation



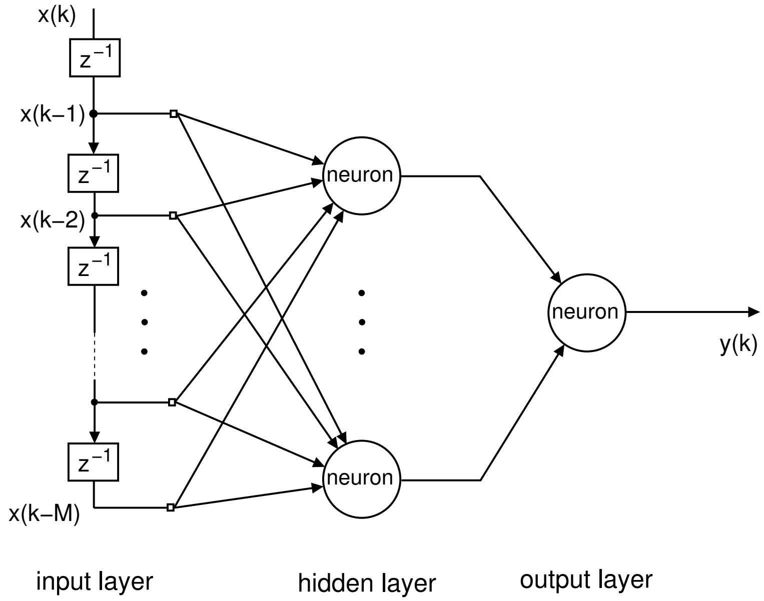
Data separation: Role of nonlinearity

Left: Original data. Right top: Separability. Right bottom: After nonlinear transf.

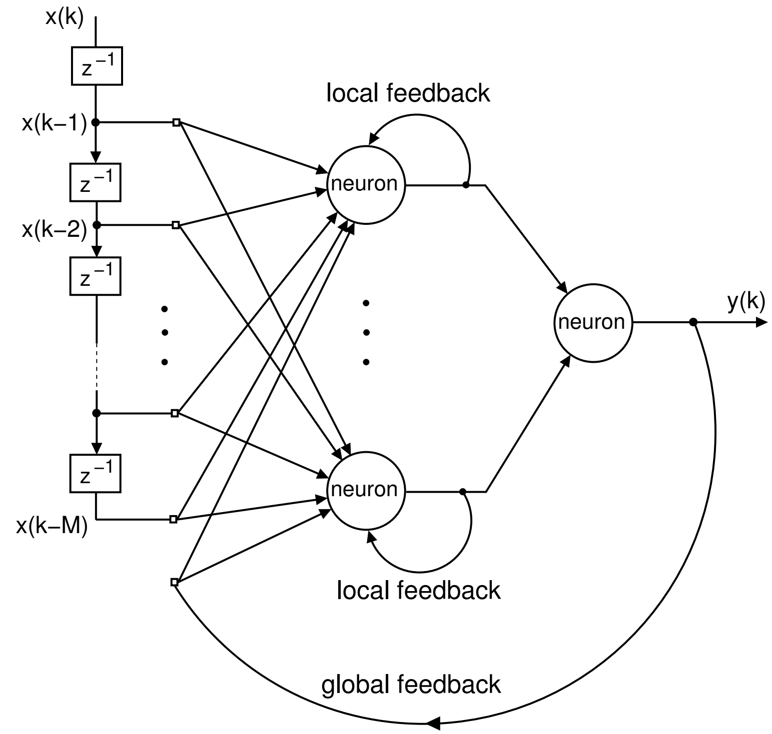


Feedforward and recurrent neural networks

Single artificial neuron models can be combined to form **neural networks.**



Feedforward network with hidden layer



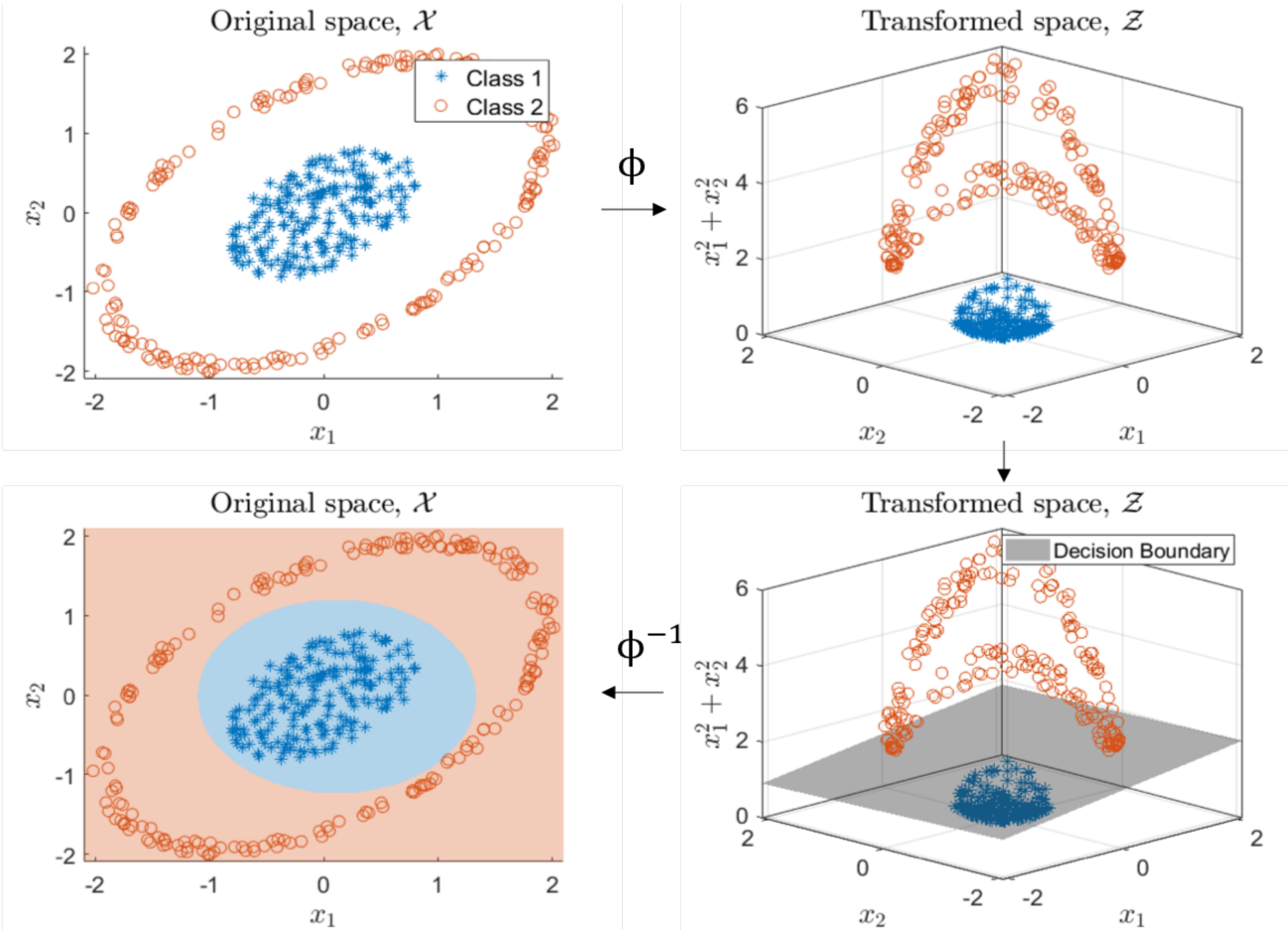
recurrent neural network

Direct gradient training of is difficult and we resort to "error backpropagation"

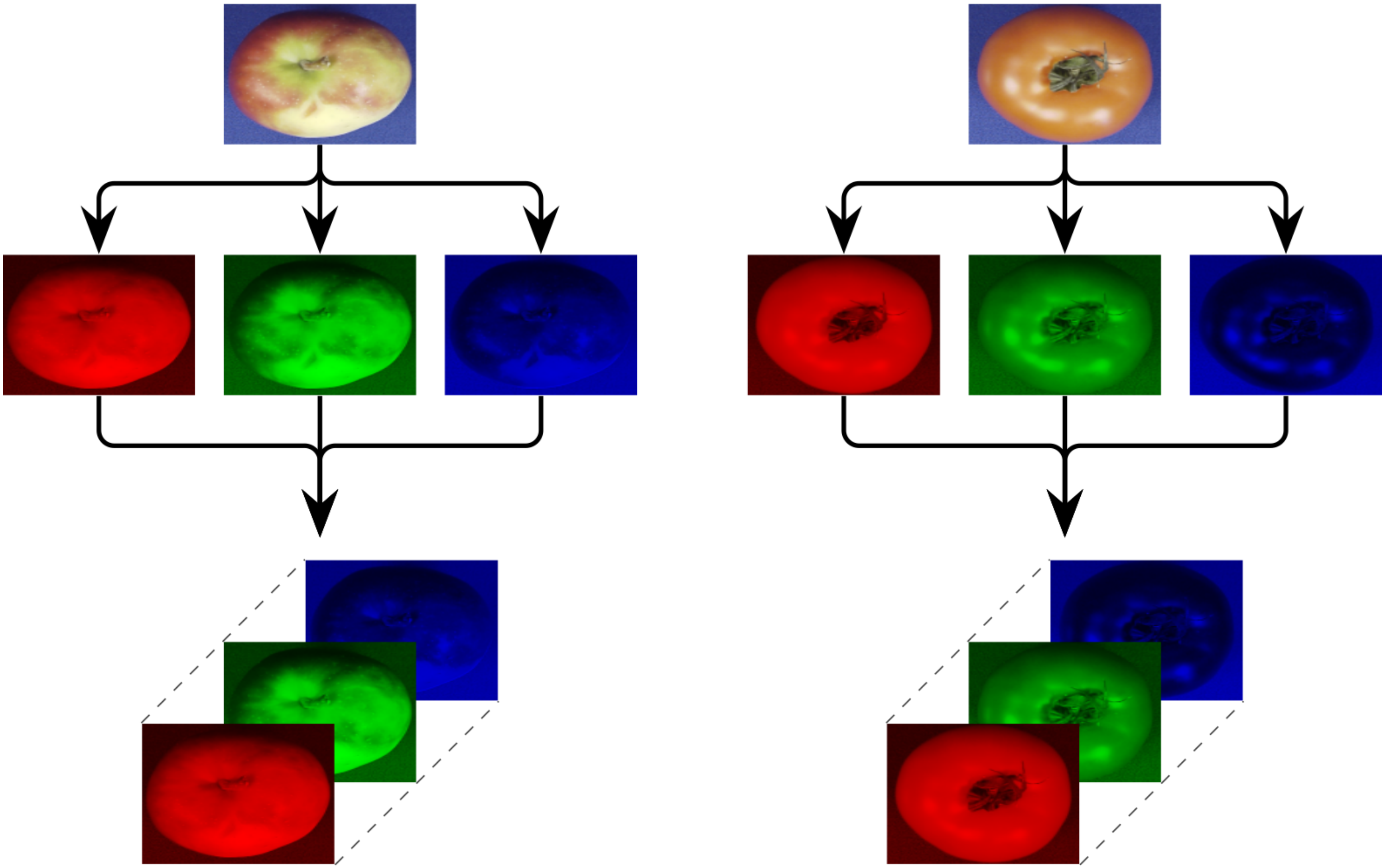
Recurrent 'feedback' connections can be local or global.

Data separation: Role of dimensionality

Left: Observed data. Right top: Original data. Bottom: Linear vs nonlinear separability

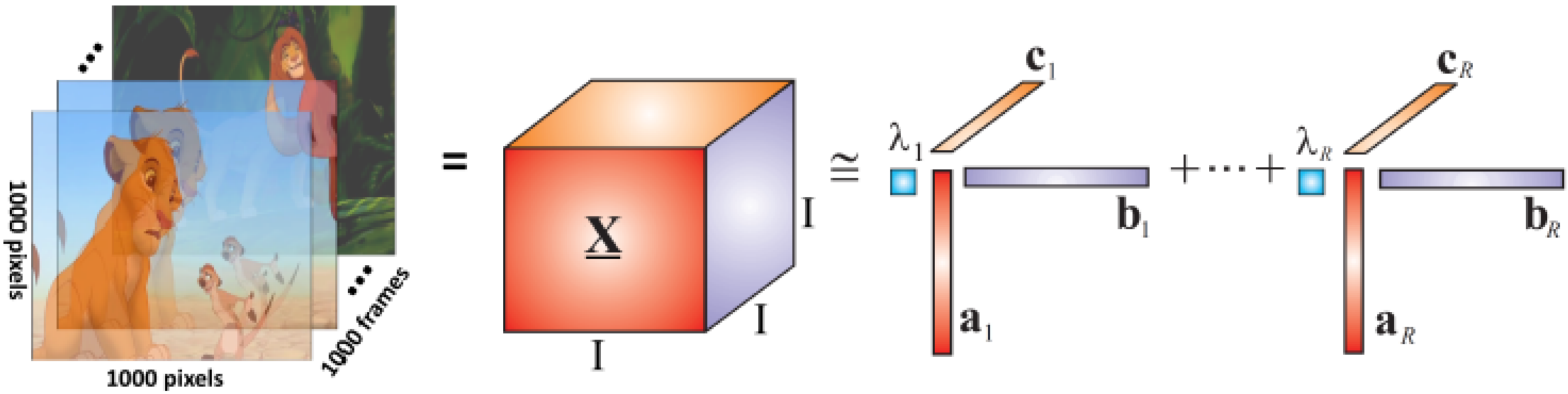


Towards Big Data: Tensor construction



Machine intelligence for Big Data

Tensors bypass the Curse of Dimensionality and 'flat-view' associated with matrices



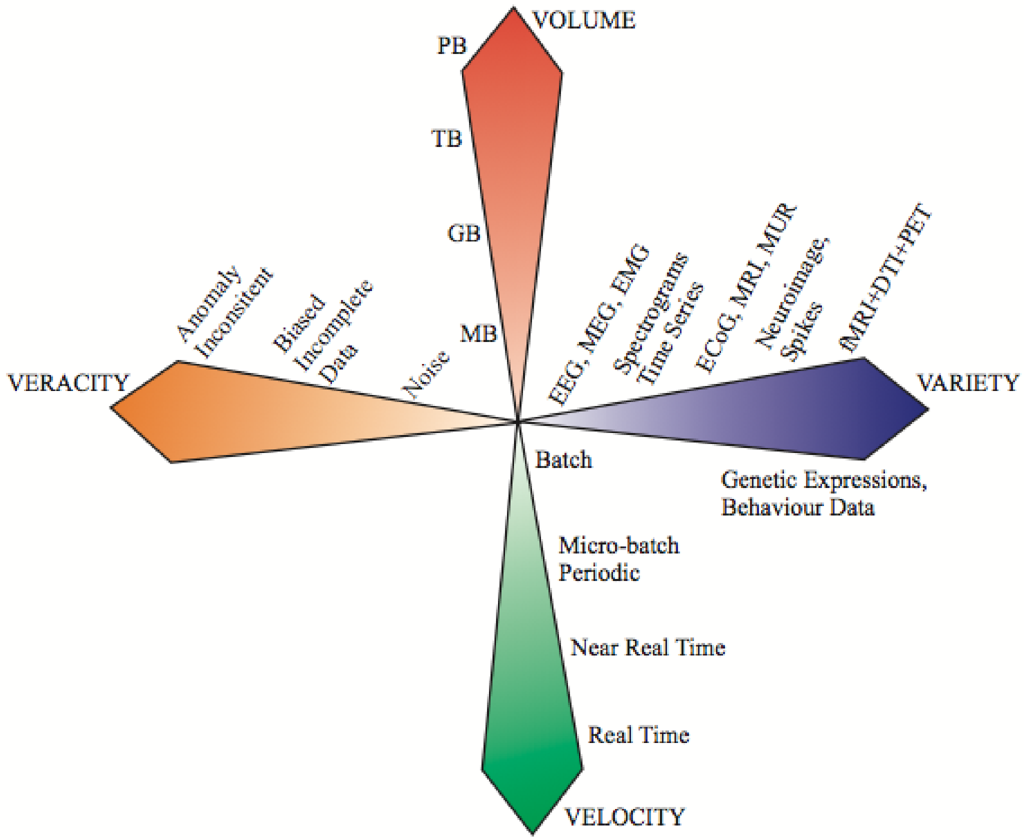
After tensorizing the video clip, tensor order $N = 3$, the dimension in every mode $I = 1000$, and the tensor rank is R. Typically $R \lll I$.

with $\texttt{length}(\mathbf{a}_i)$=1000, $\texttt{length}(\mathbf{b}_i)$=1000, $\texttt{length}(\mathbf{c}_i)$=1000, $\quad i = 1, 2, \ldots, R$

- **Raw data format** $\looparrowright I^N = 1000 \times 1000 \times 1000 = 10^9$ pixels $=$ 1 Giga-pixel
- **In the CPD format**, this becomes $N \times I \times R = 3 \times 1000 \times 10 = 30,000$ pixels (for R=10), that is, compression of almost 5 orders of magnitude
- In scientific computing, if we sample a cube at $I = 10,000$ points, then $I^N = 10^{12}$ raw samples become $N \times I \times R = 3 \times 10^5$ samples in CPD

  For N=4, I=10^4, R=10, the $I^N = 10^{16}$ raw samples $\rightsquigarrow 4 \times 10^5$ samples in CPD

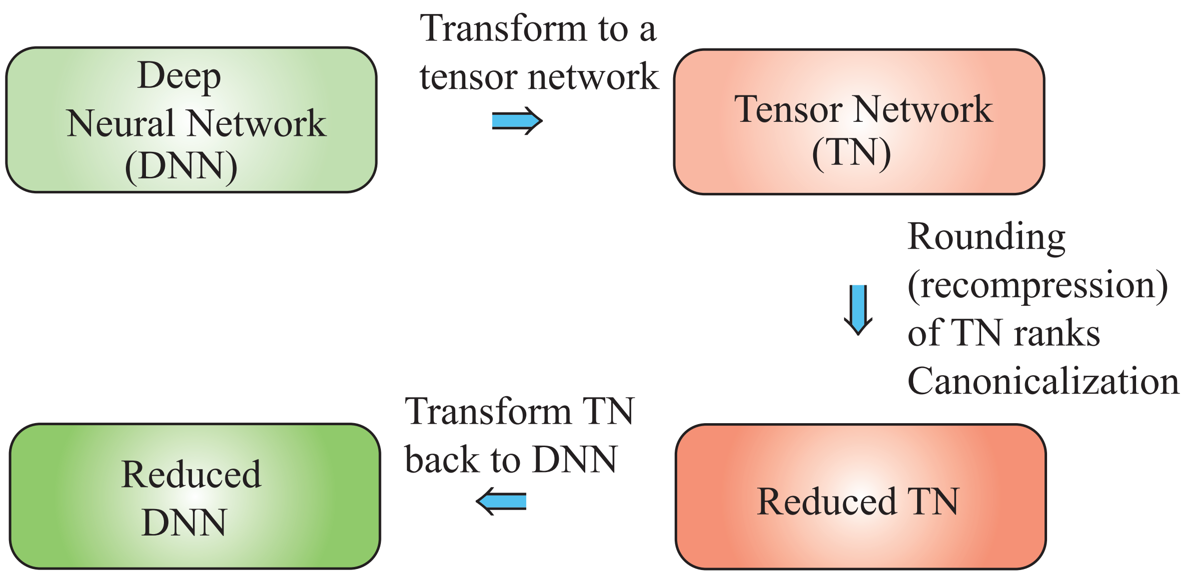
Current trends: The 'Four Vs' in big data analysis



Does this remind you of the Olympic motto:

Citius – Altius – Fortius (faster, higher, stronger)

Opening the DNN blackbox ↬ From neural networks to tensors and tensor networks



Depth efficiency ↬ DNNs can implement with polynomial size computations that would require super-polynomial size for shallow NNs.

☞ As a consequence, the deeper the network the better the performance

Problem: It is unclear to what extent convolutional neural networks leverage depth efficiency, what is the size of a deep network to perform computations not achievable by shallow networks?

Opportunities: Tensor ensemble learning

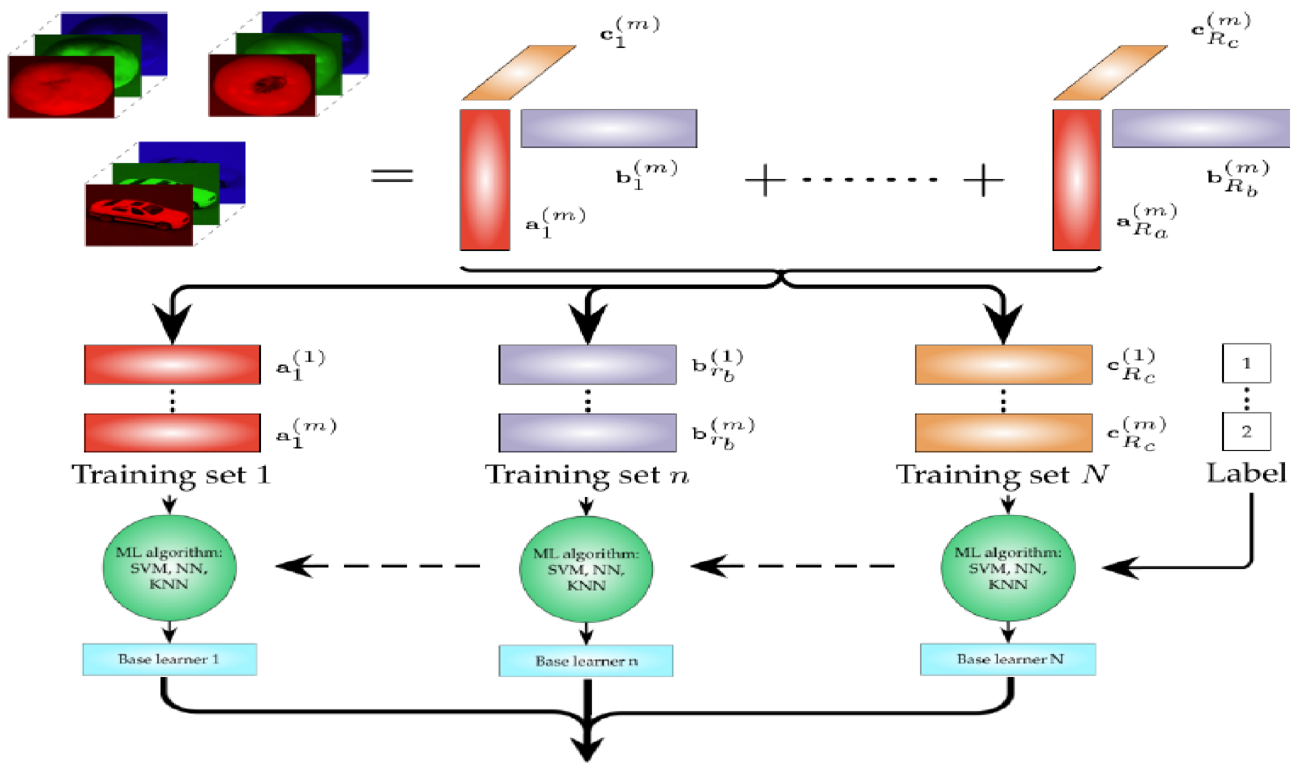


Fig. 11: The schematic diagram of Tensor Ensemble Learning – Vector Independent (TELVI) algorithm. Original tensor-valued samples are represented in Tucker form through the HOSVD with multi-linear rank (R_a, R_b, R_c). All factor vectors of training samples, $\underline{\mathbf{X}}^m$, are reorganised into separate datasets, thus only requiring to train $N = R_a + R_b + R_c$ classifiers. At the final stage, latent components $\mathbf{a}_1^{new}, \ldots, \mathbf{c}_{R_c}^{new}$ extract from a new sample, $\underline{\mathbf{X}}^{new}$, are utilised for majority voting to predict label of $\underline{\mathbf{X}}^{new}$.

Learning outcomes and course structure

1. **Dimensionality reduction.** Through "autoregressive" and "subspace" spectral estimation examples and tensor decompositions for Big Data
2. **Adaptive filtering.** Gradient based adaptive filters, complex-valued and multidimensional adaptive filters, recursive least squares
3. **Machine intelligence.** Links with state-space models, concept of an artificial neuron, neural networks, deep learning, links with tensors and Big Data
4. **Case studies.** Practical examples in communications, acoustics, biomedical engineering, renewable energy, finance
5. **Lecture course with problem sets**
   - *Modern spectral estimation: 4 lectures*
   - *Adaptive signal processing: 8 lectures*
   - *Machine intelligence: 8 lectures*
6. **Assessment:** 100% Coursework, Feedback after Assignment 1

Learning outcomes and course structure

Special emphasis will be on demonstrating the close links between:

- spectral estimation
- adaptive signal processing and
- machine intelligence

Throughout the course we will discuss Matlab implementation

- **There will always be signals**
- **They always need processing**
- **There will always be new mathematics and analysis frameworks for processing them**

$\rightsquigarrow$ **Guaranteed job security**

A note on our data sources: Portable data acquisition

Comparison of features of bio-amplifier

Imperial Amplifier (I-Amp)

Dimensions: ~40 x 15 x 15 mm
Weight: ~ 30 g
Power supply: one coin baterry
Recording time: 48+ hrs

g.USBamp

Avatar EEG

Dimensions: 197 x 155 x 40 mm
Weight: 1000 g
Channel no: 16
Power supply: mains
Price: 13,000 Euro (~10,000 GBP)

Dimensions: 76 x 53 x 38 mm
Weight: 60 g
Channel no: 8
Power supply: 2 AA baterries
Recording time: 24 Hrs
Price: >3,000 GBP

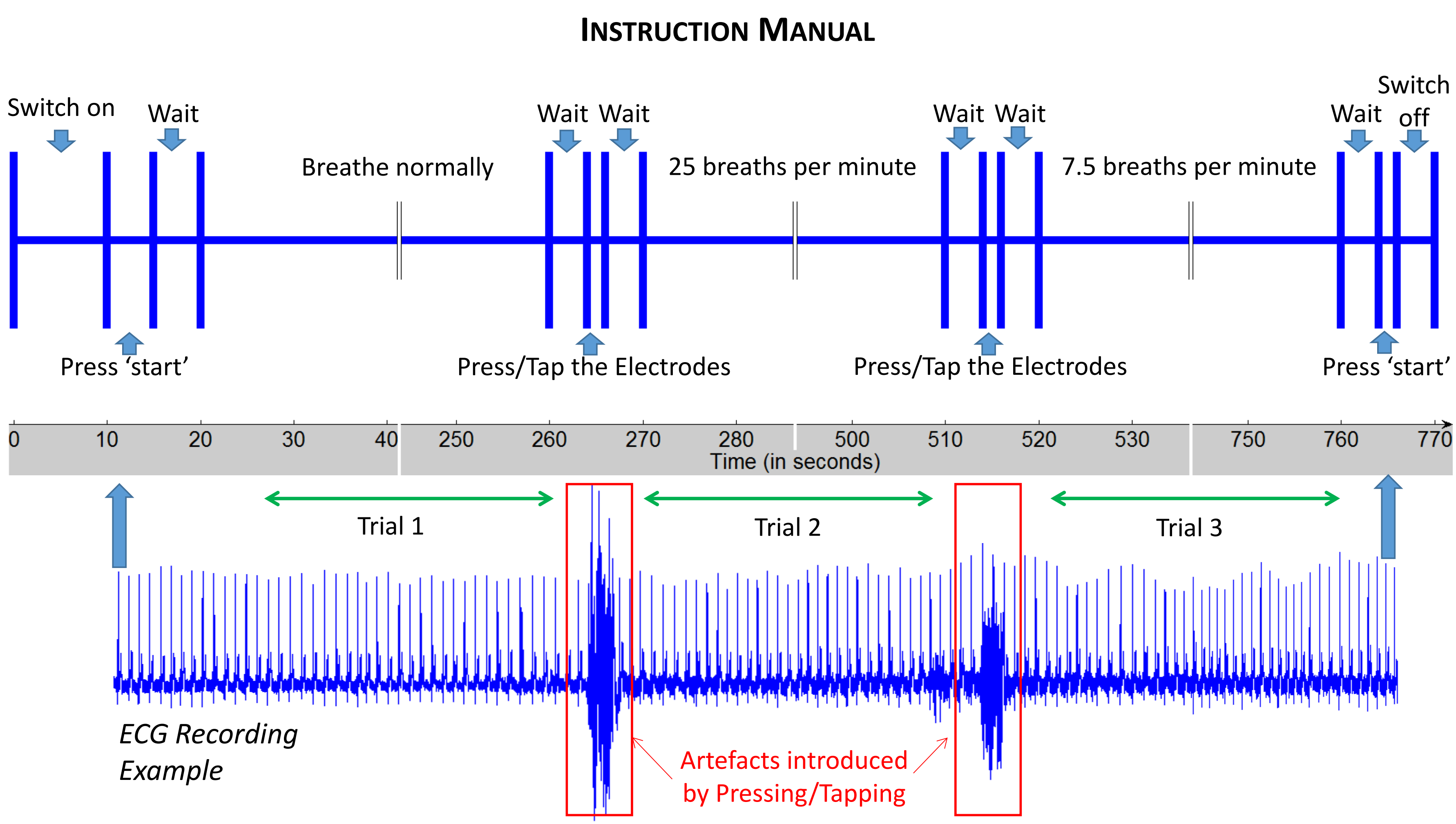
Coursework $\looparrowright$ partly based on students' own data

Experiment summary: Students will record their ECG under varying respiration patterns: (1) unconstrained breathing; (2) controlled breathing at a fast pace (50 breaths/min) and slow pace (15 breaths/min).

Easy to follow on-screen instructions.

Student placing electrodes on a team-mate.

ECG recording instructions



Prerequisites and reference material

- The course is largely self-contained, there are no prerequisites
- Having attended Digital Signal Processing, Advanced Signal Processing, and basic Probability Theory and Statistics would be useful
- Knowledge of some form of programming language is essential for the understanding of the implementation issues of the algorithms covered
- My preference is Matlab, the *de facto* language of scientific computing (at least in technology), especially for non-hardware implementation

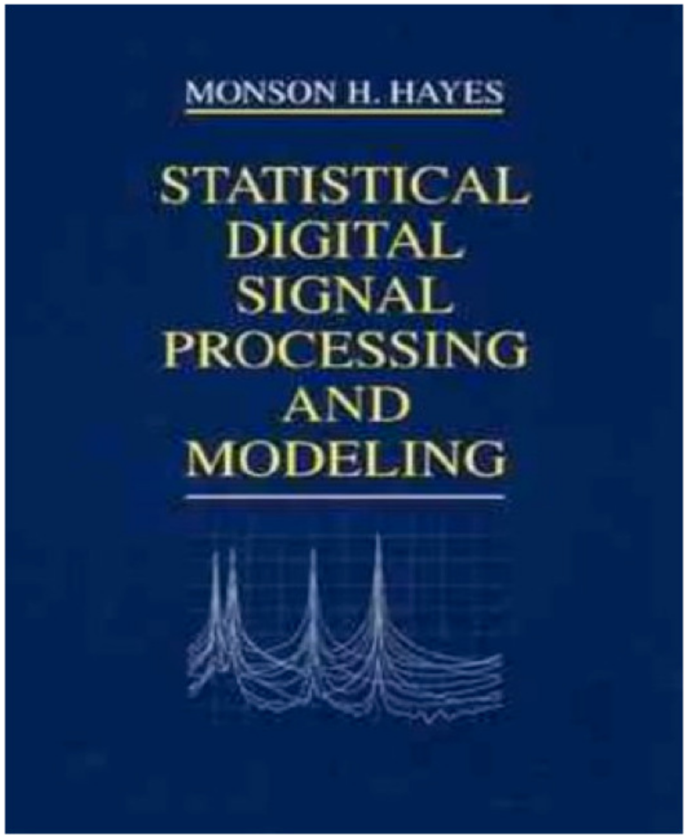
Literature:

1. **Course notes and problem/answer sets: by Dr Mandic**
2. **The book by M. Hayes for Spectral Estimation**
3. **Hayes' book and D. Mandic's book for Adaptive Filtering and Neural Networks**

   Any other bits and pieces will be in course notes

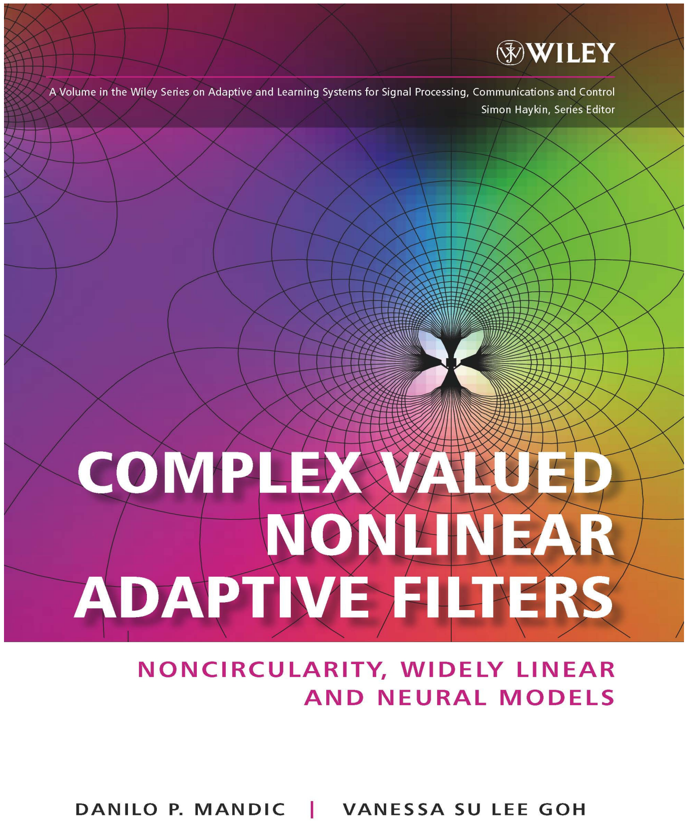
Textbooks: Recommended

M. Hayes (*Statistical Signal Processing*, several editions)



spectral estimation and part of adaptive filters

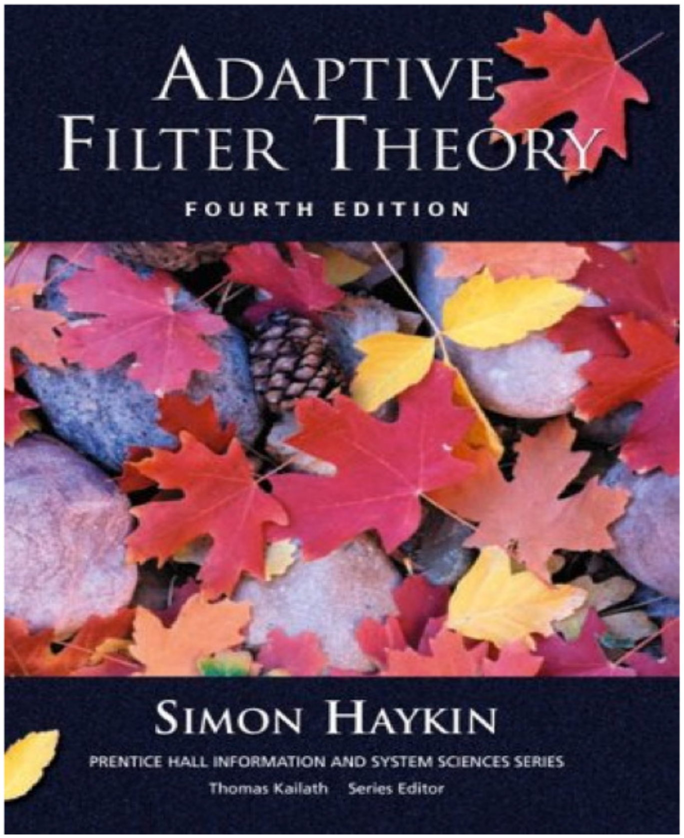
D. Mandic and S. Goh (*Complex Adaptive Filters*, Wiley 2009).



real, complex, and neural adaptive filters

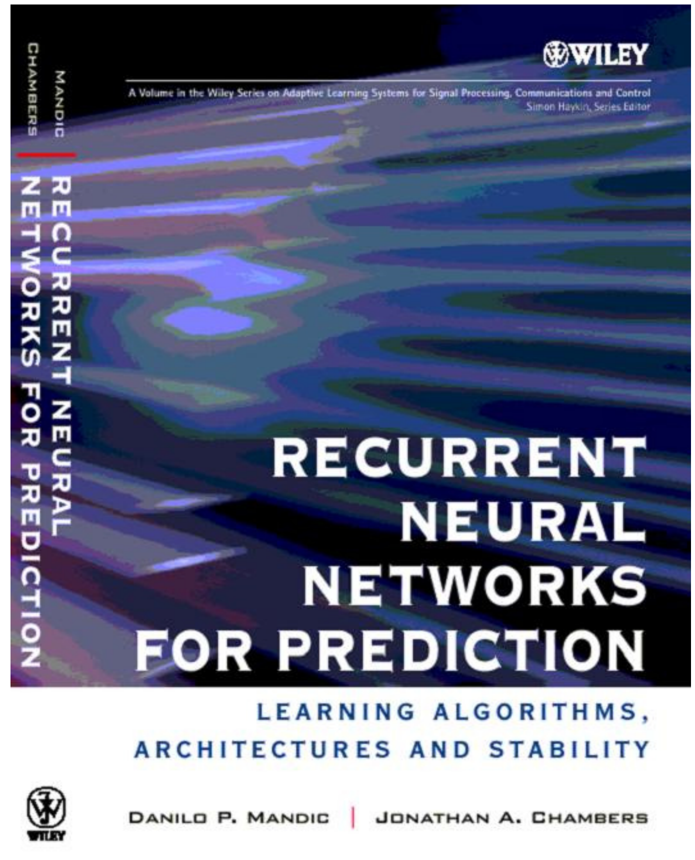
Textbooks: Additional reading

S. Haykin (*Adaptive Filtering Theory*, several editions)



a comprehensive account of adaptive filters

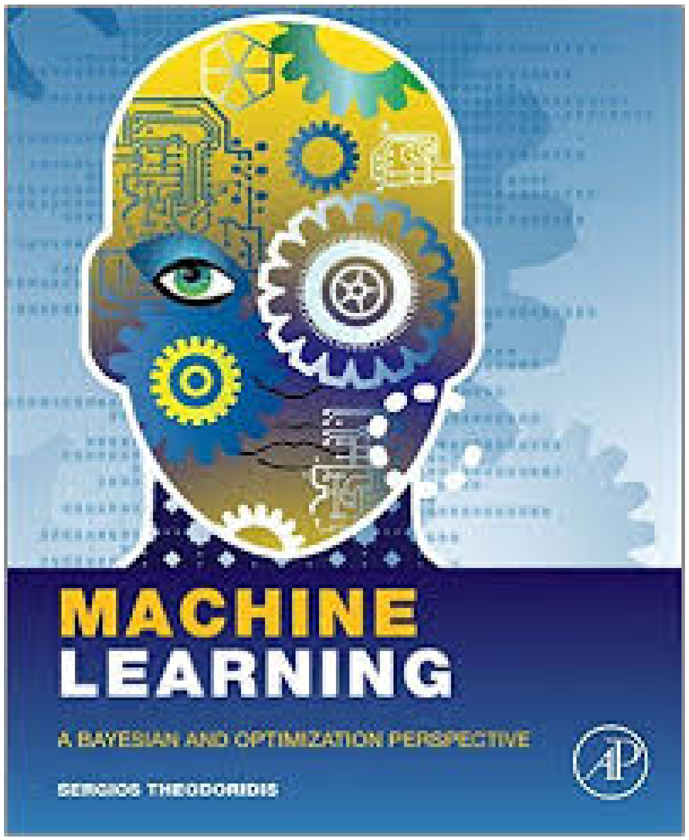
D. Mandic & J. Chambers (*RNNs for Prediction*, Wiley 2001).



feedback and neural network architectures

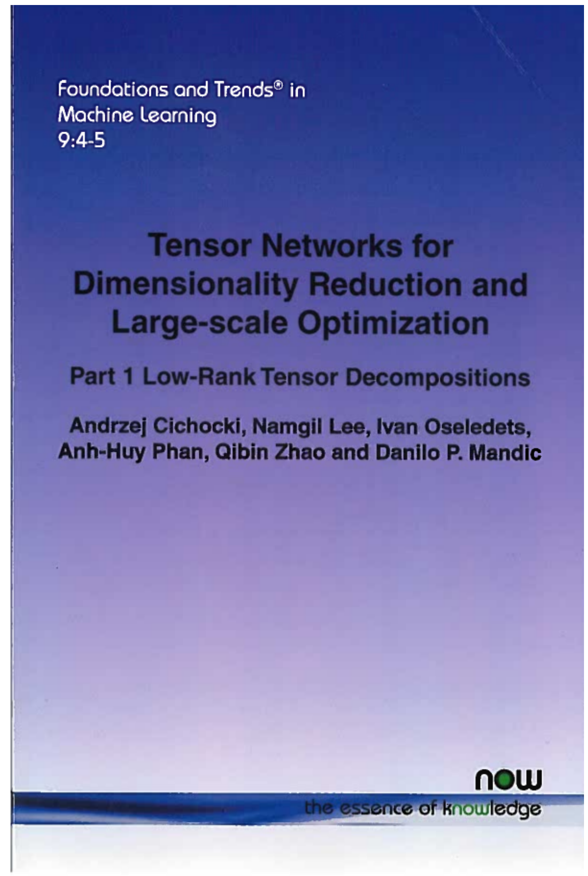
Useful supporting material

S. Theodoridis (*Machine Learning*, 2015)



a Bayesian and optimisation perspective to machine learning

A. Cichocki, D. Mandic, *et al.* (*Tensor Networks*, 2016).



Big Data, tensors, dimensionality reduction

Course plan

1 Lect: Week 2, Course introduction and motivation

3 Lect: Week 2-3, Spectral estimation, classical and subspace–based

4 Lect: Week 3-4, Adaptive learning systems, adaptive filters

4 Lect: Week 4-6, Complex adaptive filters and applications

4 Lect: Week 6-8, Nonlinear and neural filters, neural networks for temporal data, recurrent neural networks

4 Lect: Week 9-10, Dimensionality reduction, tensor methods for Big Data

Course web page: www.commsp.ee.ic.ac.uk/∼mandic/Teaching

Lectures, additional reading, homework, problem sets, and other material will be put on course webpage

To conclude

ASP & MI is a combination of two very important areas in modern data analytics

- Finding latent components in data, through various factorisations, from block to recursive PCA, with application in spectral estimation
- Adaptive filters are an enabling technology for many real world applications in nonstationary environments (acoustics, speech, communications, teleconferencing, biomedicine, renewable energy, genomics and proteomics)
- Machine intelligence algorithms (neural networks) are nonparametric (black box), data-adaptive, with no model imposed on the data
- Dimensionality reduction through tensor decompositions, links between tensor networks and deep neural networks

The material in this course - statistical learning systems - is generic and is applicable across the areas of engineering and computing

Notes:

◦

Notes:

○