Machine Learning I and II

N. Bertschinger
M. Kaschube
V. Ramesh

Organization

- Lecture:

  Monday 10:00-12:00, SR 9

- Tutorial & problem session:

  Monday 12:00-14:00, SR 9

  Weekly or biweekly problem sets

  Tutorials on special topics

My email: kaschube@fias.uni-frankfurt.de

Schedule

WiSe 16/17

- 17. 10. Kaschube
- 24. 10. Bertschinger
- 31. 10. Ramesh
- 07. 11. Kaschube (Methods)
- 14. 11. Kaschube (Methods / applications)
- …
- Last part of course: Ramesh (Applications / methods)

SoSe 17:

Bertschinger (Methods / applications)

…

Ramesh (Applications / methods)

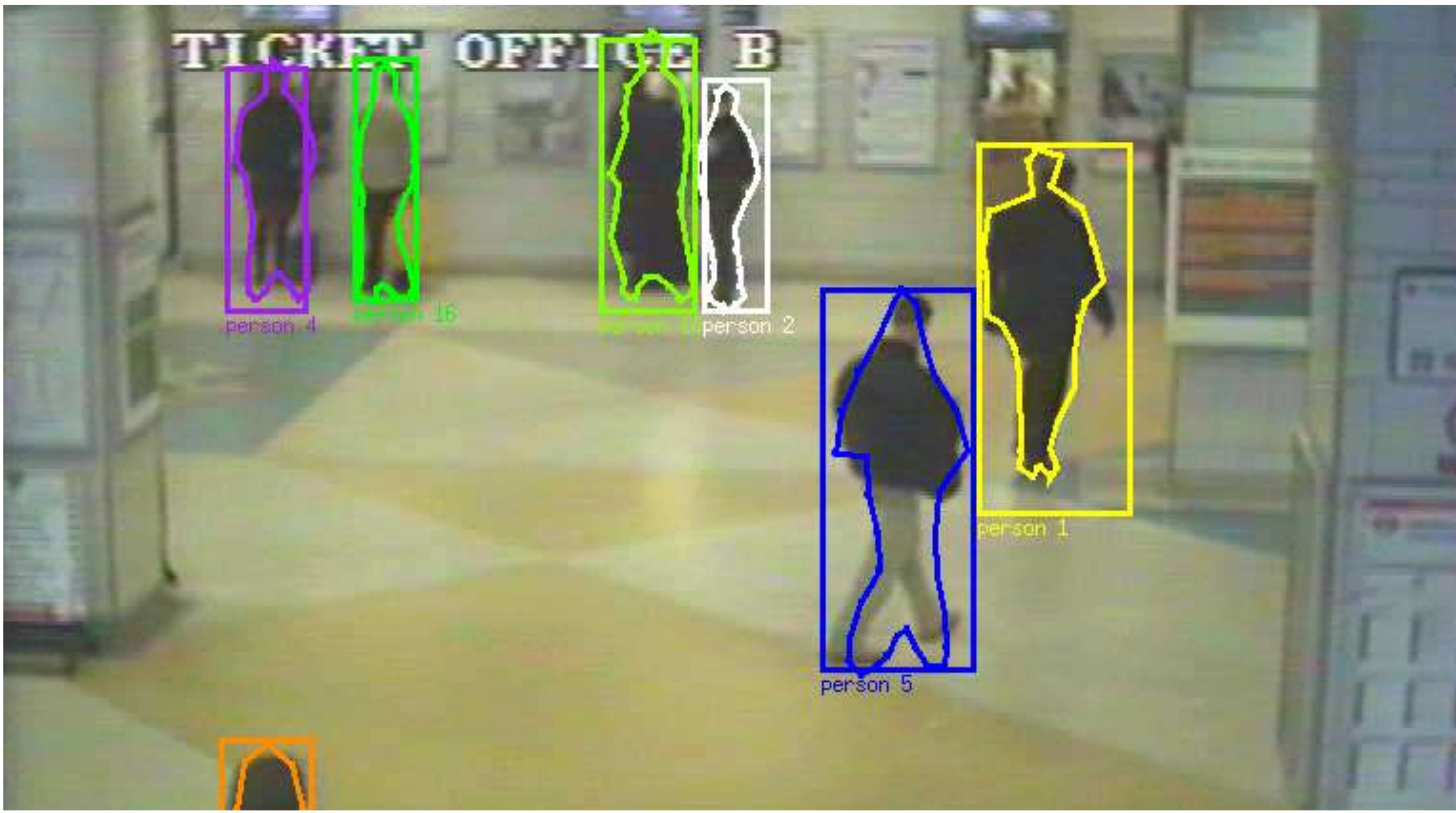
Machine learning and pattern recognition

Stock market

Surveillance

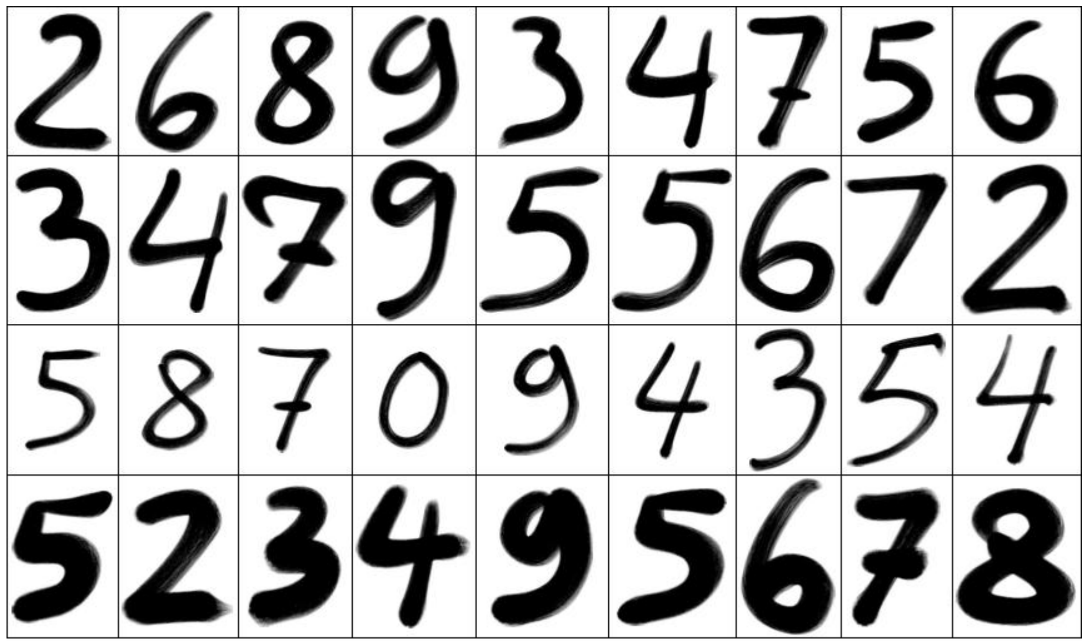
Surveillance

Surveillance



Identity authentication

Handwritten digit recognition



- MNIST database (Mixed National Institute of Standards and Technology database)

MNIST database

- It was created by "re-mixing" the samples from NIST's original datasets.
- The creators felt that since NIST's training dataset was taken from American Census Bureau employees, while the testing dataset was taken from American high school students, NIST's complete dataset was too hard.
- The database contains 60,000 training images and 10,000 testing images.

Performance

Type of classifier	Authors	Error rate (%)
Linear classifier	LeCun et. al, IEEE 1998	7.6
Non-Linear Classifier	LeCun et. al, IEEE 1998	3.3
Boosted Stumps	Kégl et. al, ICML 2009	0.87
Support vector machines	DeCoste & Schölkopf, MLJ 2002	0.56
K-Nearest Neighbors	Keysers et. al, IEEE PAMI 2007	0.52
Neural network	Ciresan et. al, Neural Comput 2010	0.35
Convolutional neural network	Ciresan et. al, CVPR 2012	0.23

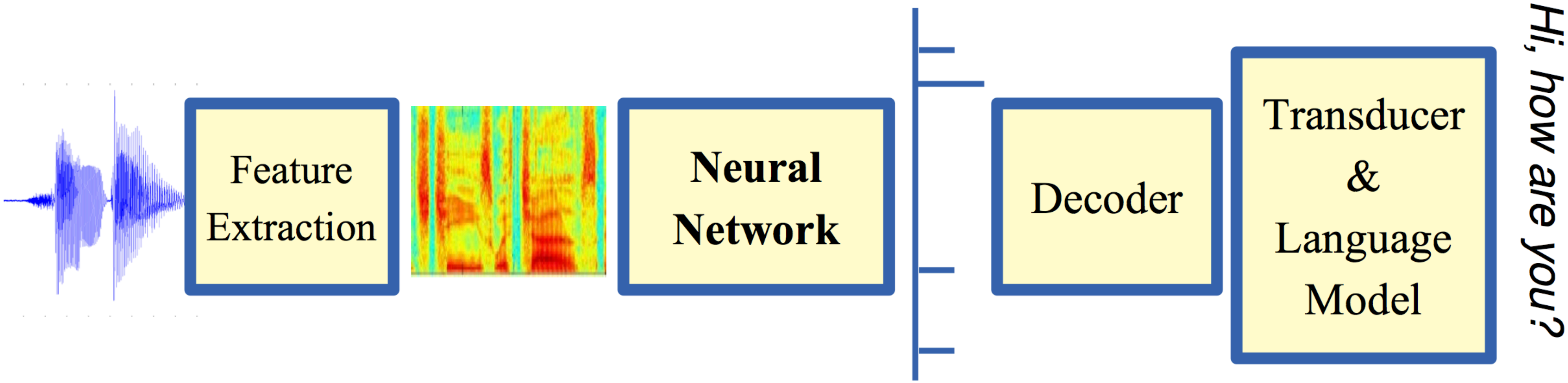
Speech recognition

Characteristics of the data

- US English:
  - Voice Search, Voice Typing, Read data
- Billions of training samples
- Input: log-energy filter bank outputs
  - 40 frequency bands
  - 26 input frames
- Output: 8000 phone states

(Zeiler et al. , ICASSP 2013; LeCunn & Ranzato , ICML 2013)

A typical speech recognition system



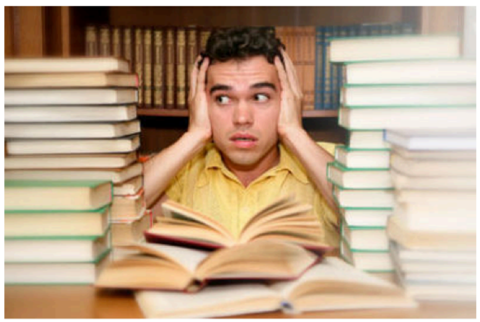
We focus only on the prediction of phone states from short time-windows of spectrogram.

(LeCunn & Ranzato , ICML 2013)

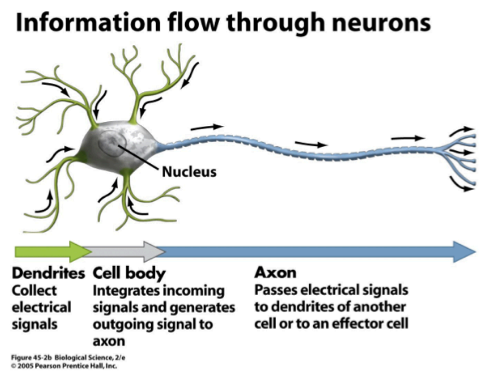
Note: The brain itself is a remarkable pattern recognition system -
and it has provided significant inspiration on artificial recognition systems

Krizhevsky et al., NIPS 2012

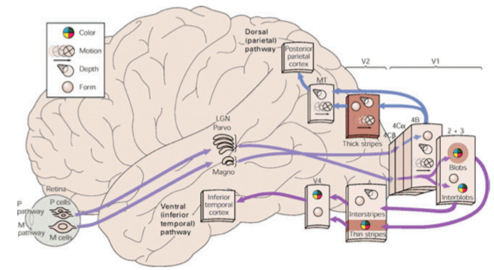
Deep Learning: Motivated from Human Learning



Learn massive data

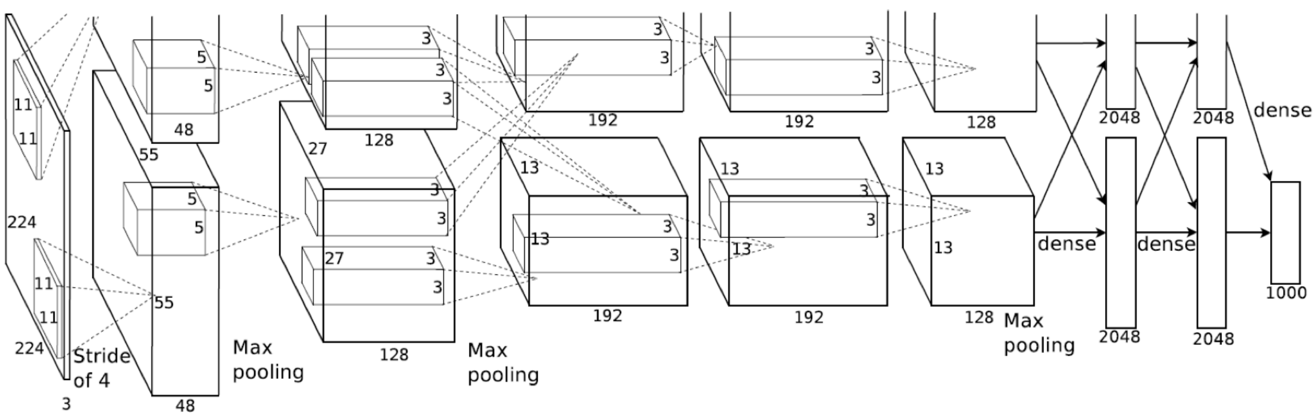


simple functions



(Van Essen&Gallant, 1994)

Multi-layered



(Krizhevsky et al., 2012)

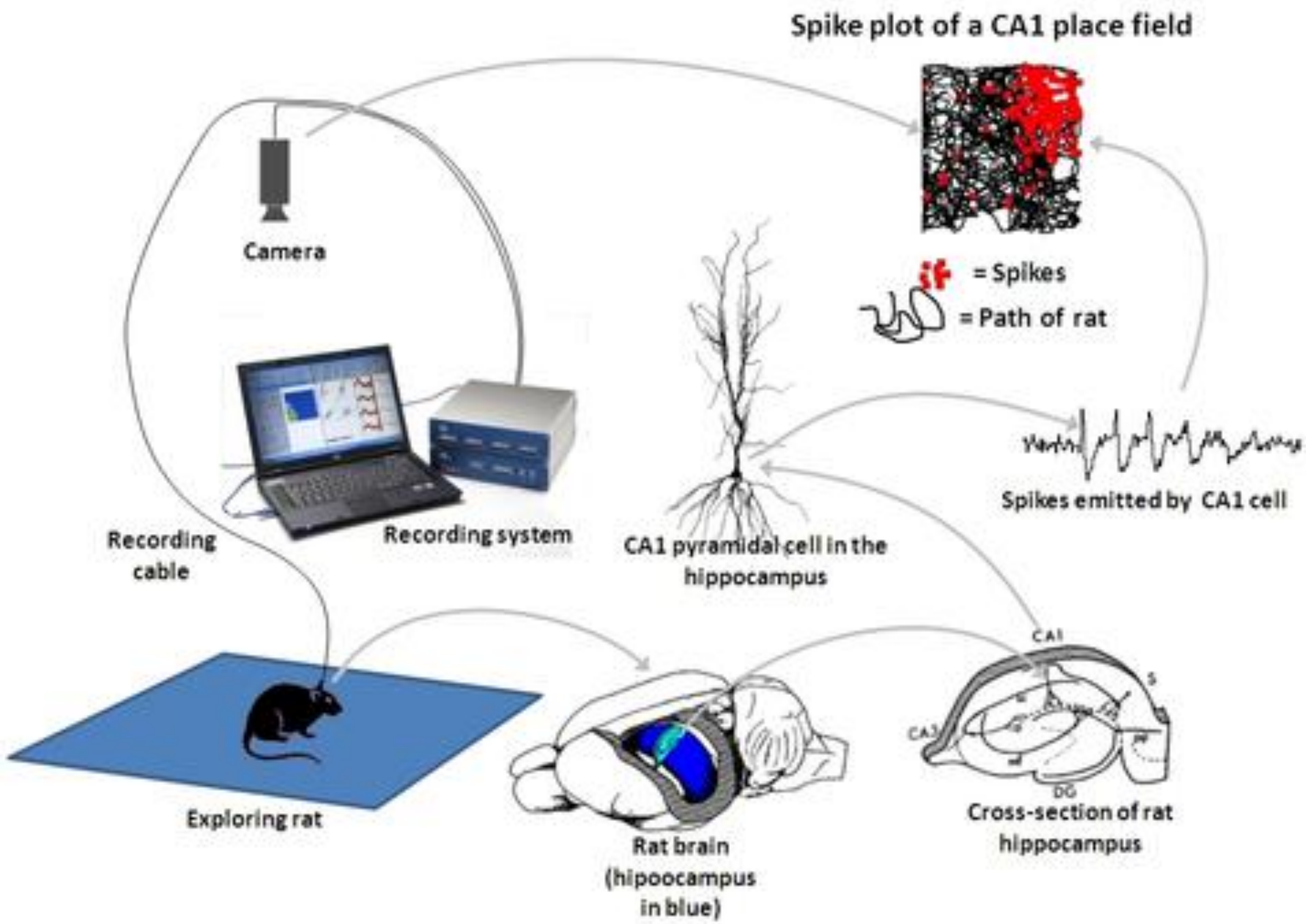
Try it out for yourself!

http://www.image-net.org/
http://www.clarifai.com/
https://images.google.com/?gws_rd=ssl

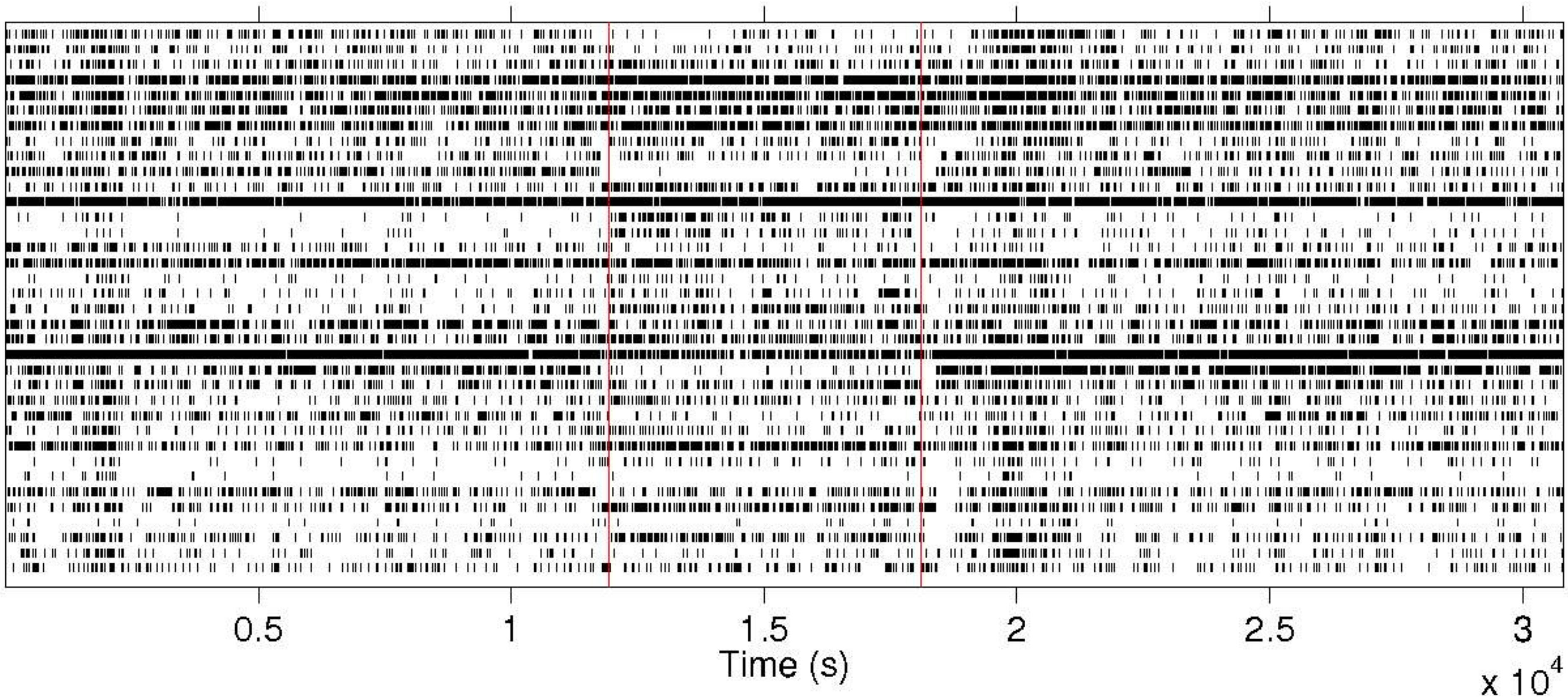
Upload your favorite image and let it recognize its content and/or find similar images!

Measuring brain activity

Example: representation of space in rat hippocampus (Nobel Prize for Medicine, 2014)



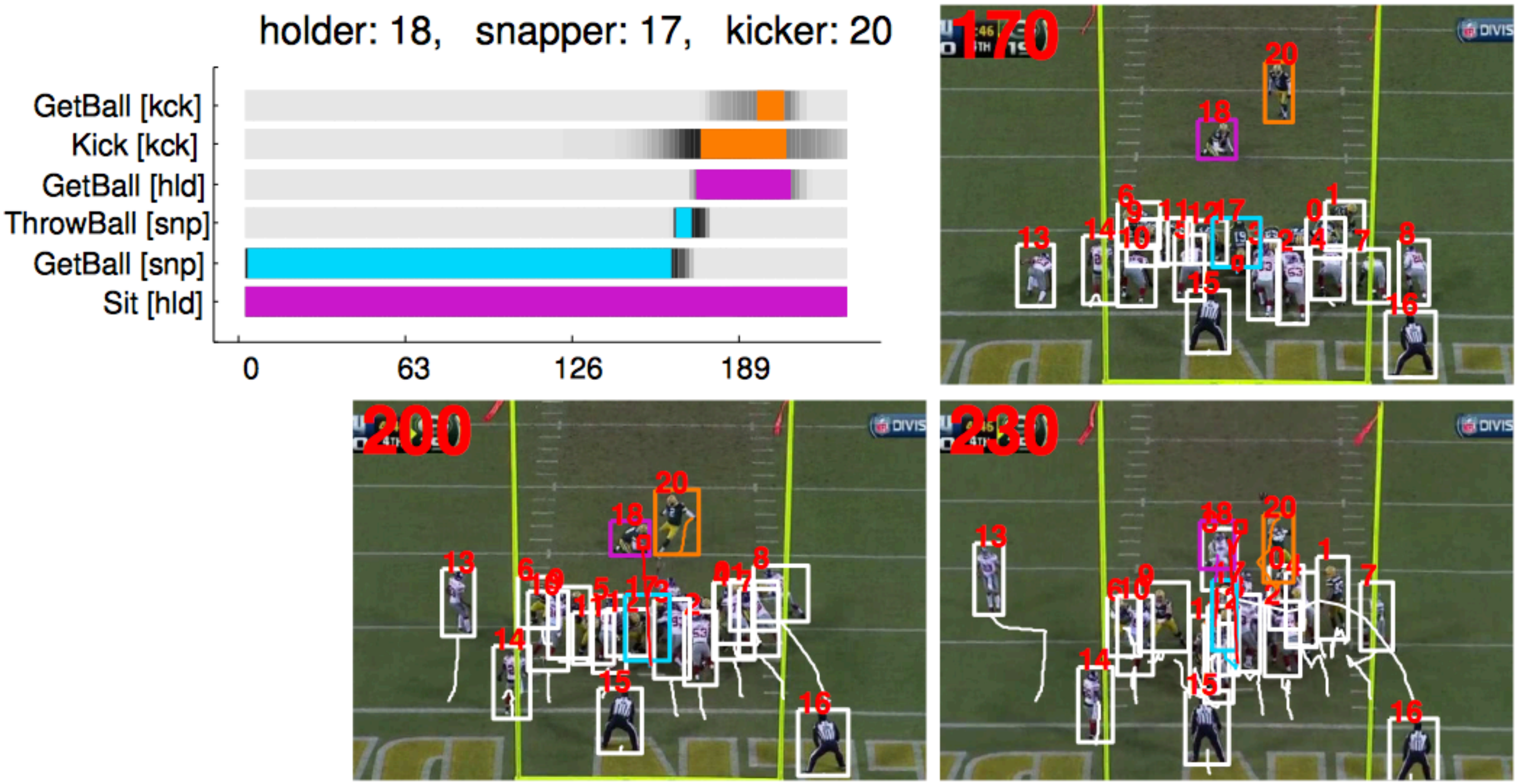
Brain activity



- Raster plot of our 36 spike trains
- Each row corresponds to a spike train, with small, black, vertical lines indicating the time of individual spikes.
- The vertical red lines indicate the start and end of the maze exploration period.

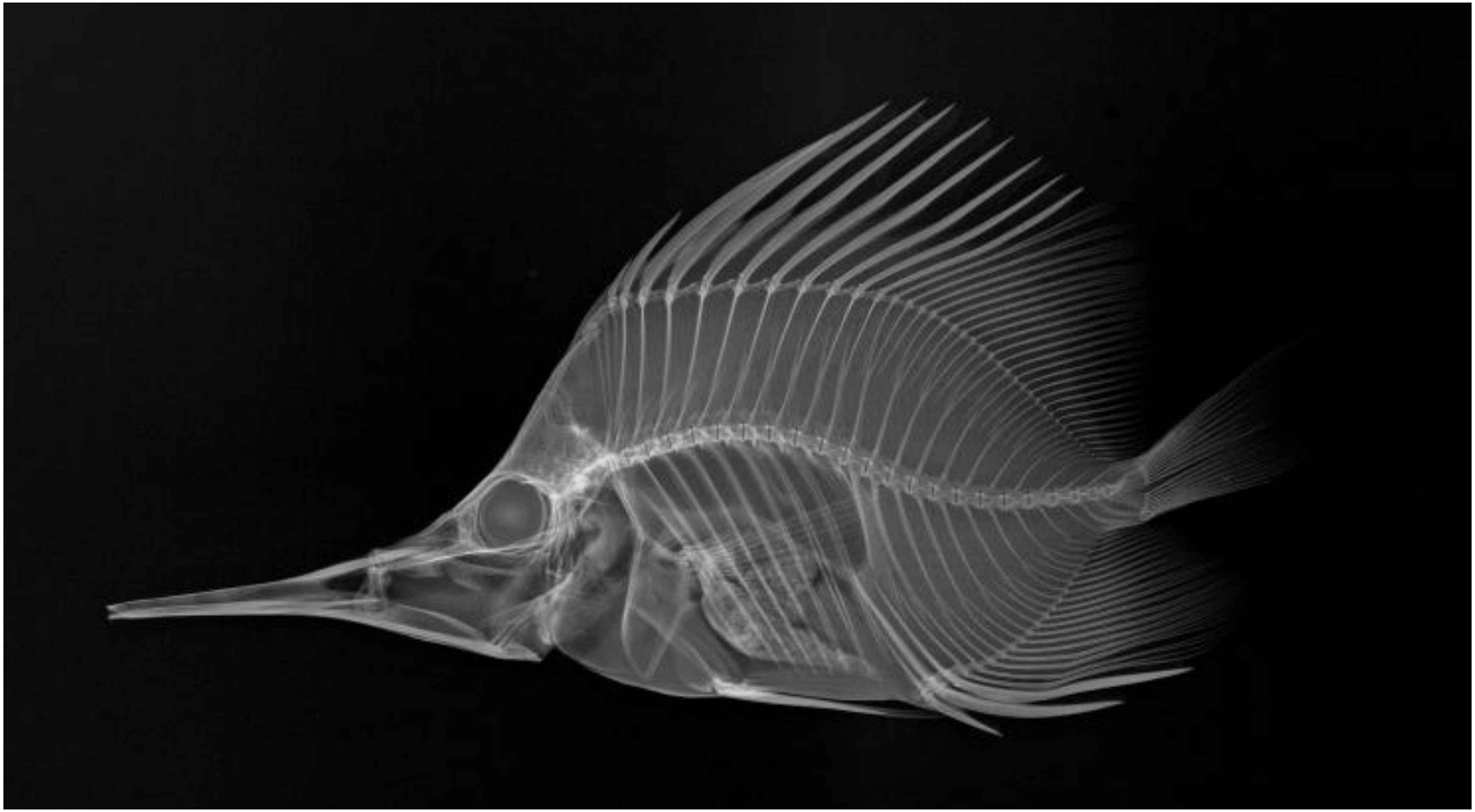
Object Recognition in Sports

Holder, snapper, and kicker are denoted by purple, cyan, and orange, respectively while outsiders are denoted by white boxes.

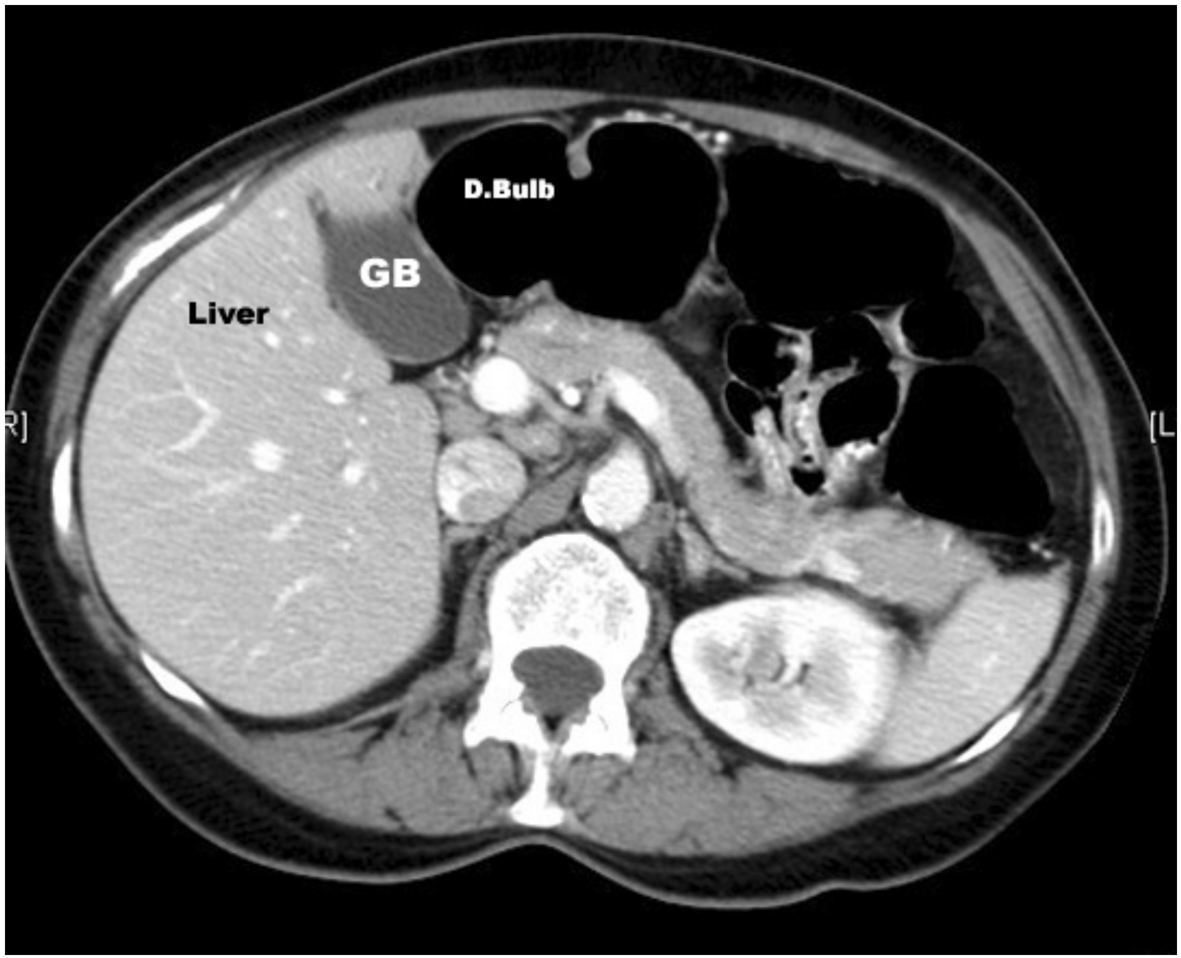


(Kwak et al., CVPR 2013)

Medical applications



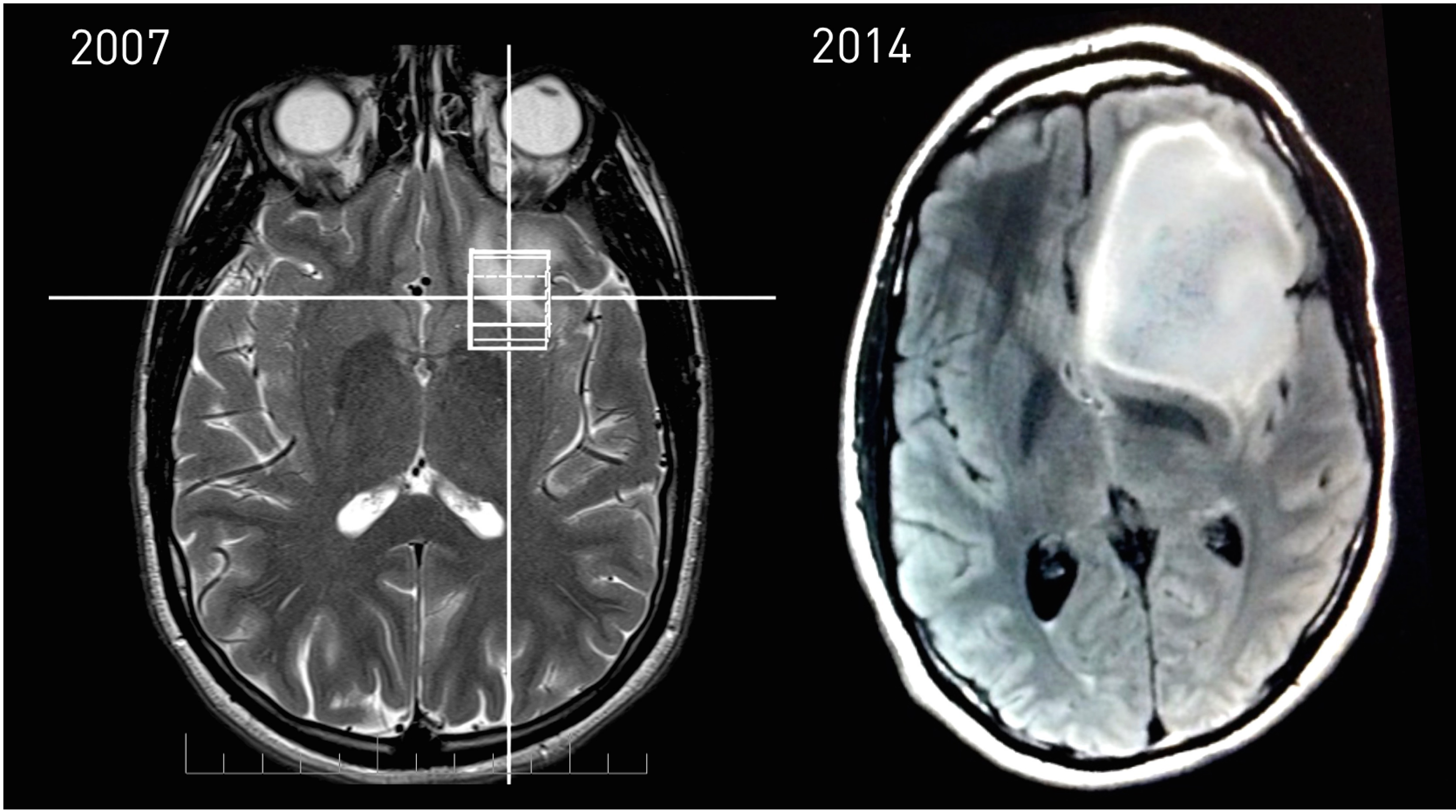
Computed Tomography (CT)



Computed Tomography (CT)

- A narrow beam of x-rays is aimed at a patient and quickly rotated around the body, producing cross-sectional images of the body.
- CT scans can be used to identify disease or injury within various regions of the body.
- CT image reconstruction is an ill-posed inverse problem. It may be solved using unsupervised learning techniques such as Filtered Back-Projection (FBP)

Magnetic Resonance Imaging (MRI)



(Keating, 2015)

Magnetic Resonance Imaging (MRI)

- MRI can detect tumors at early stages, revealing parts of the body that are not easily shown with other techniques.
- In the clinic, classification methods such as LDA, K-means, and SVMs are frequently used to detect tumors in early stages, providing an important support for radiologists' diagnosis.

Slightly more abstract…

Two approaches to pattern recognition

- Data-driven
- Model-driven

Frequentist vs. Bayesian*

*Thomas Bayes 1701–1761

Classification

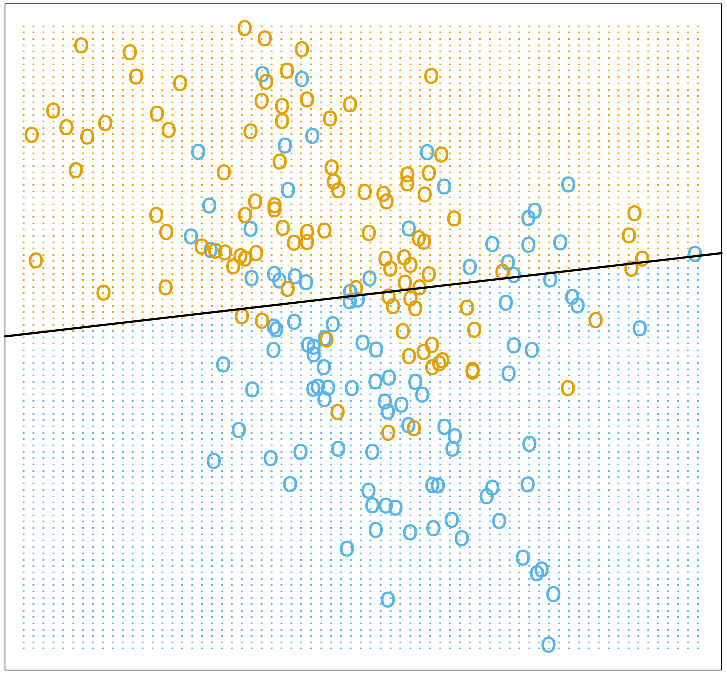


FIGURE 2.1. *A classification example in two dimensions. The classes are coded as a binary variable* (`BLUE` $= 0$, `ORANGE` $= 1$)*, and then fit by linear regression. The line is the decision boundary defined by* $x^T\hat{\beta} = 0.5$*. The orange shaded region denotes that part of input space classified as* `ORANGE`*, while the blue region is classified as* `BLUE`*.*

Classification

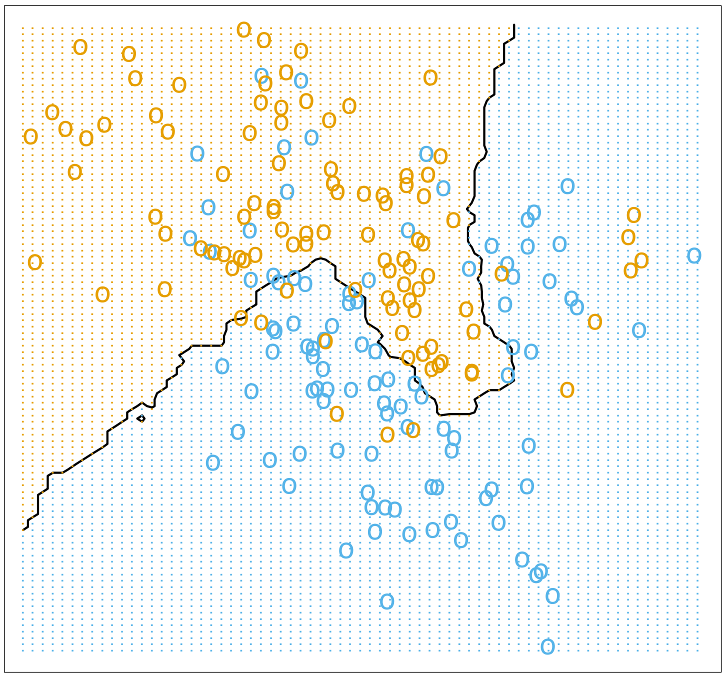


FIGURE 2.2. *The same classification example in two dimensions as in Figure 2.1. The classes are coded as a binary variable* (BLUE $= 0$, ORANGE $= 1$) *and then fit by* 15*-nearest-neighbor averaging as in (2.8). The predicted class is hence chosen by majority vote amongst the* 15*-nearest neighbors.*

Classification

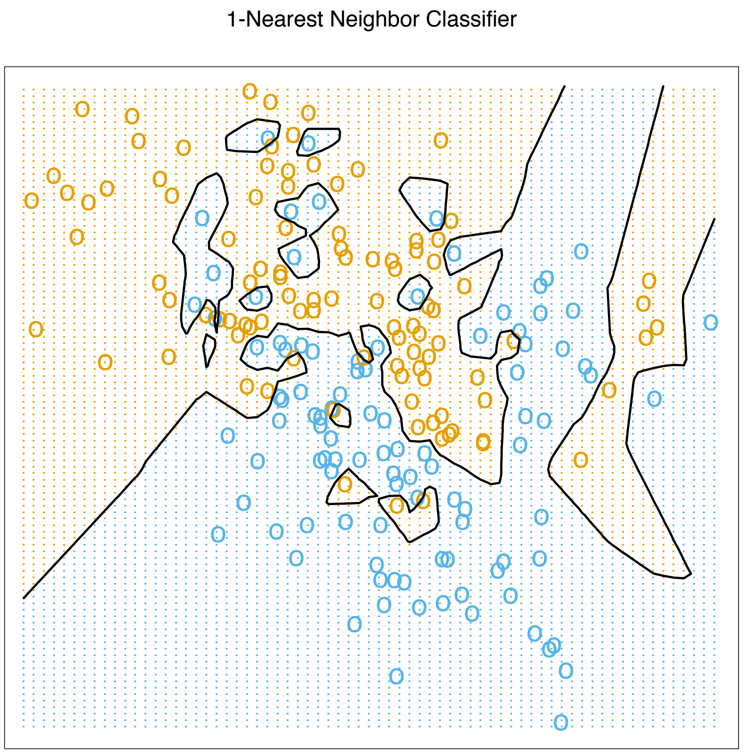


FIGURE 2.3. *The same classification example in two dimensions as in Figure 2.1. The classes are coded as a binary variable* (BLUE $= 0$, ORANGE $= 1$)*, and then predicted by* 1*-nearest-neighbor classification.*

Bayesian classification

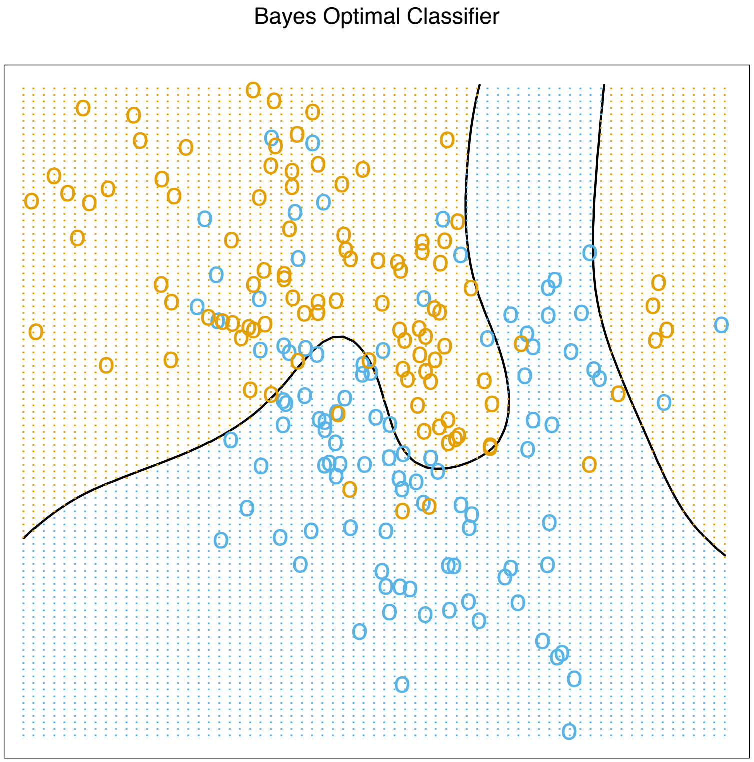
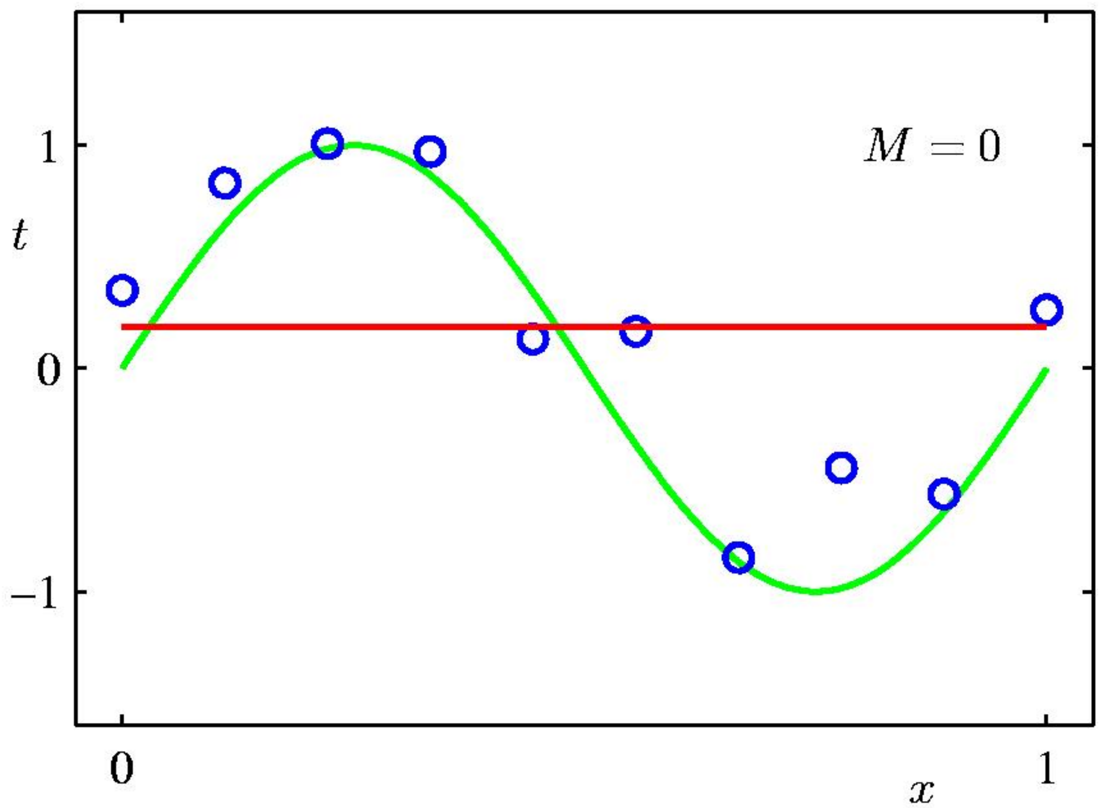
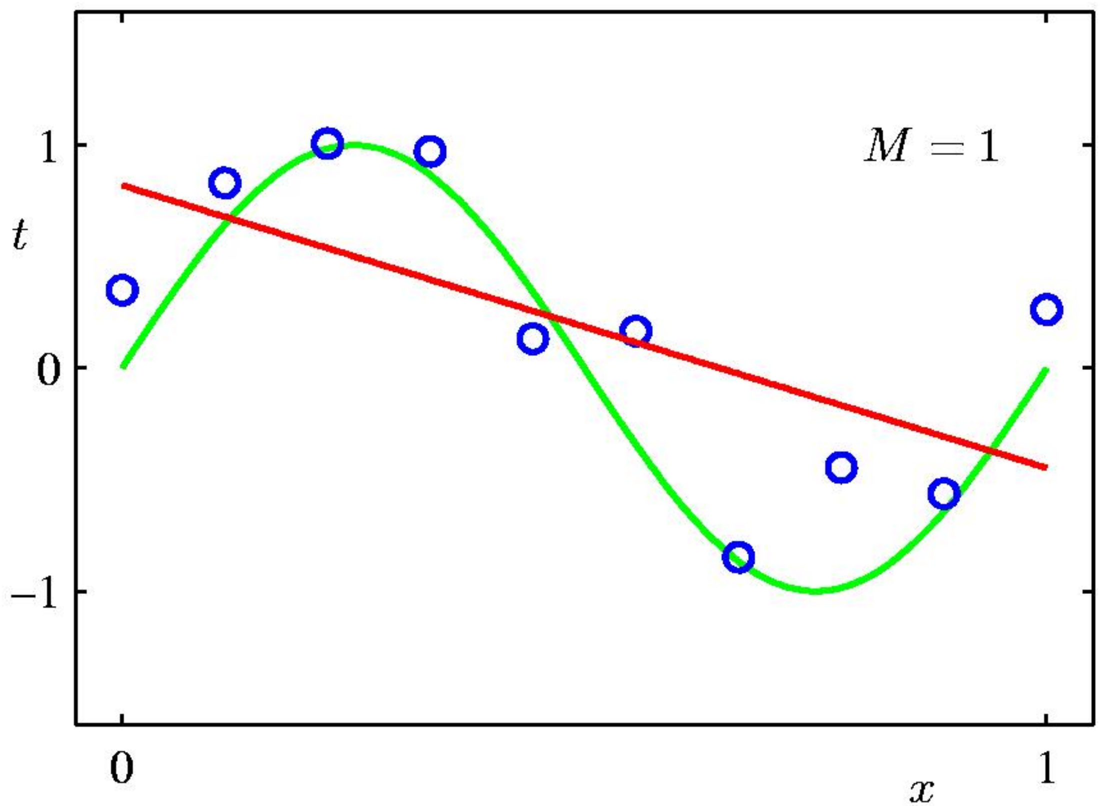


FIGURE 2.5. *The optimal Bayes decision boundary for the simulation example of Figures 2.1, 2.2 and 2.3. Since the generating density is known for each class, this boundary can be calculated exactly (Exercise 2.2).*

Curve fitting (regression)

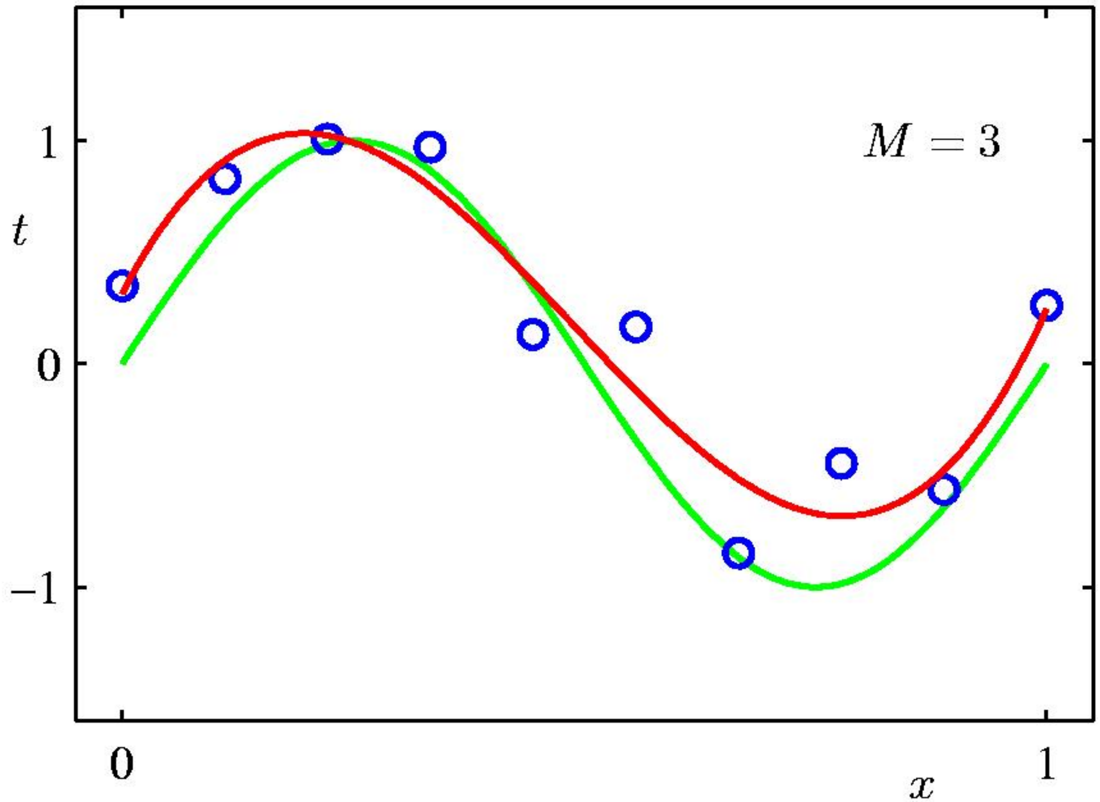


Curve fitting (regression)



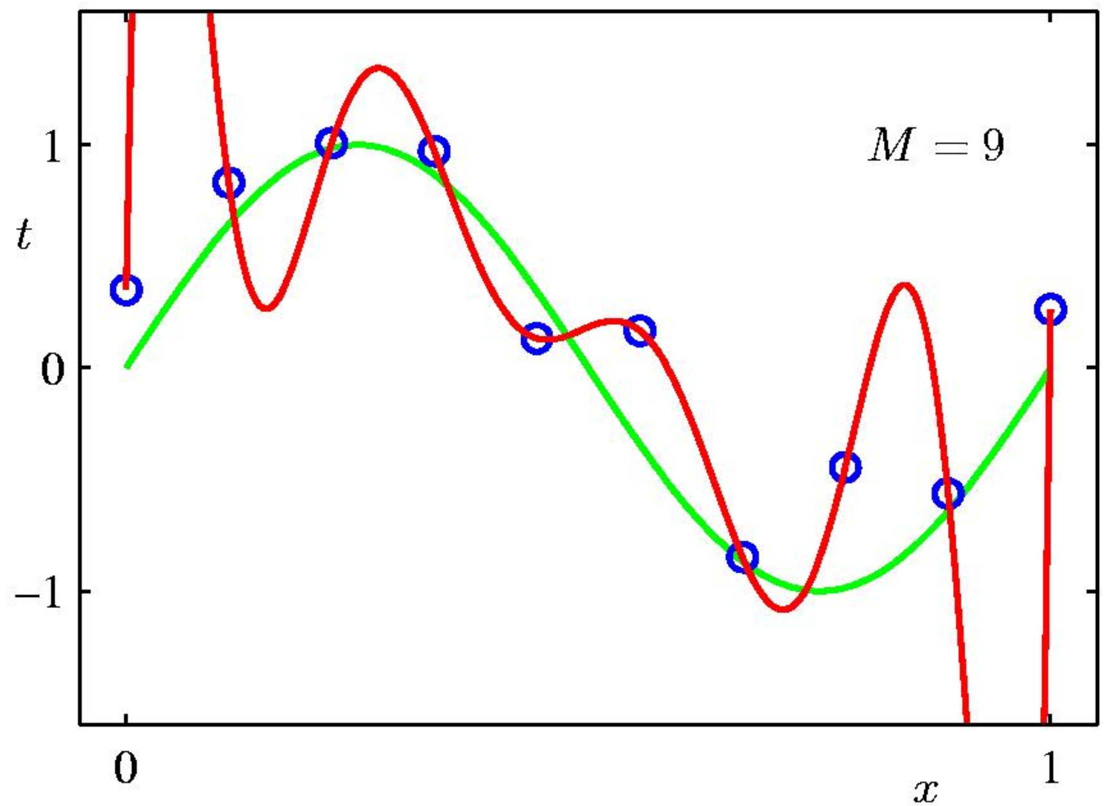
1^{st} Order Polynomial

Curve fitting (regression)



3rd Order Polynomial

Curve fitting (regression)

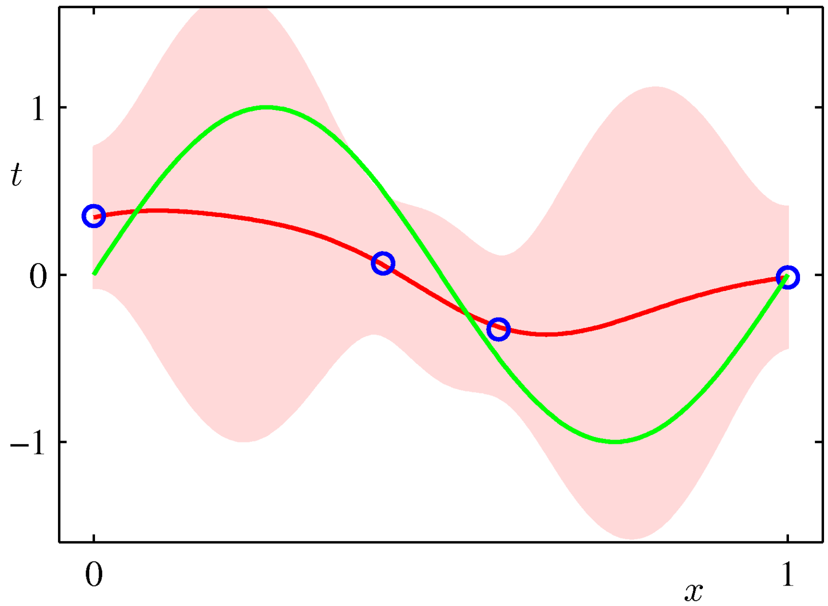


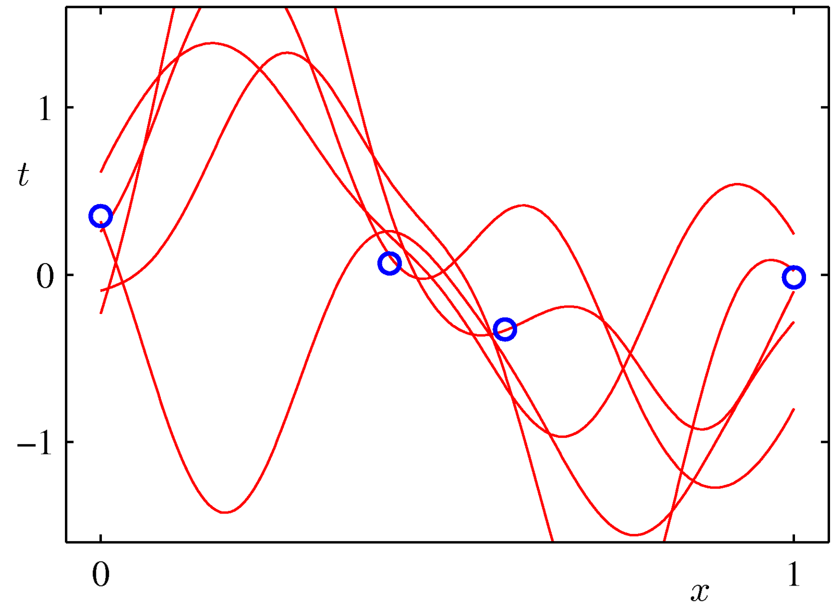
9^{th} Order Polynomial

Overfitting, a major issue!

Bayesian Curve Fitting

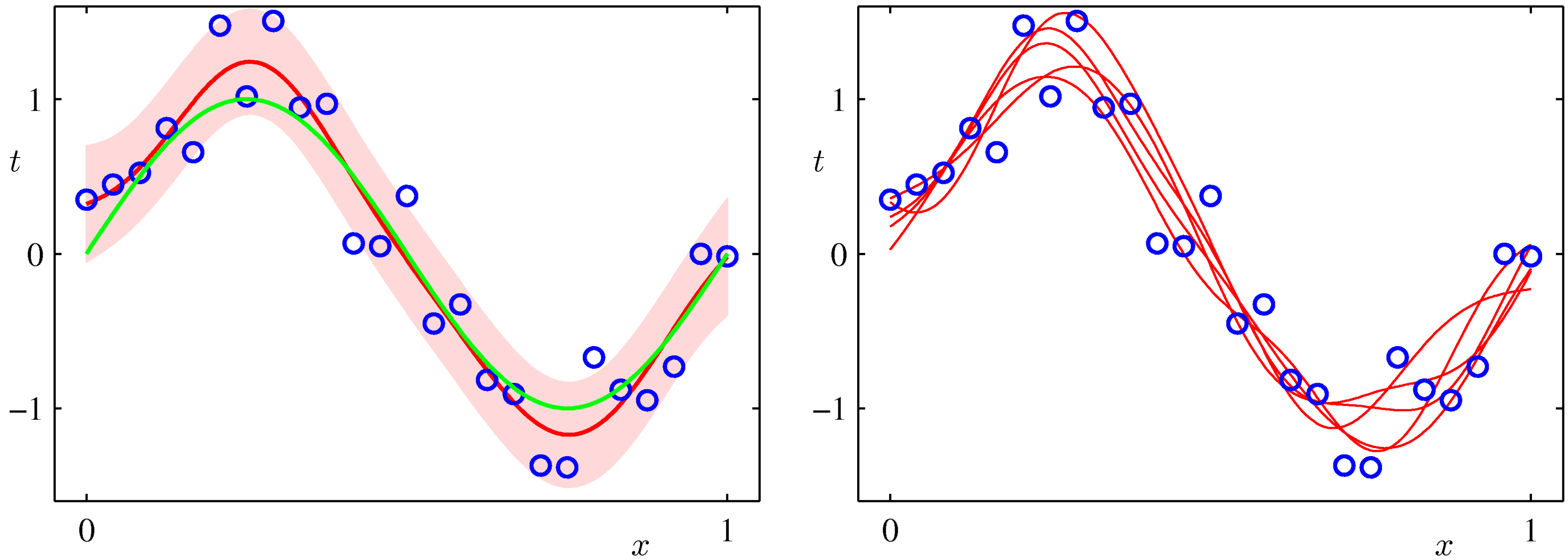
- Example: Sinusoidal data, 9 Gaussian basis functions, 4 data points





Bayesian Curve Fitting

- Example: Sinusoidal data, 9 Gaussian basis functions, 25 data points



Supervised vs. unsupervised

- Supervised – unsupervised pattern recognition:
  The two major directions
  - **Supervised**: Patterns whose class is known a-priori are used for training.
  - **Unsupervised**: The number of classes is (in general) unknown and no training patterns are available.

Topics ML1
(WiSe16/17; Kaschube and Ramesh)

Topics:
- Math Recap (linear algebra, probability)
- Probability distributions
- Linear Regression, Bayesian linear Regression
- Classification
- Neural Networks, deep learning
- Latent variables
- Decision trees, randomized forests
- Clustering

Applications:
- Neuroscience and bio-imaging data
- Vision Applications for detection, tracking, recognition, segmentation, registration, geometry estimation, etc.

Topics ML2
(SoSe16/17; Bertschinger and Ramesh)

- Motivation: Bayesian vs frequentist statistics
- Decision theory: handling uncertainty, loss functions
- Probability theory: Conjugate priors
- Modeling: Latent variables, hierarchical models, Bayesian nonparametrics
- Model selection: Marginal likelihood, sparsity priors
- Algorithms: Variational Bayes (ELBO), sampling
- Mixture models , hmm's, EM algorithm
- Graphical model based inference, MCMC based sampling

Applications:

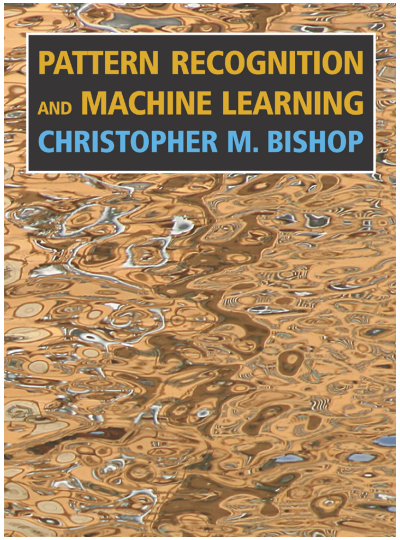
- Social data: Voting results, network models
- Economic data: GDP forecasting, volatility modeling
- Vision Applications for detection, tracking, recognition, segmentation, registration, geometry estimation, etc.
- Other applications in language and/or speech processing will be covered if time permits.

Books 1

Pattern Recognition and Machine Learning

Christopher M. Bishop

Springer

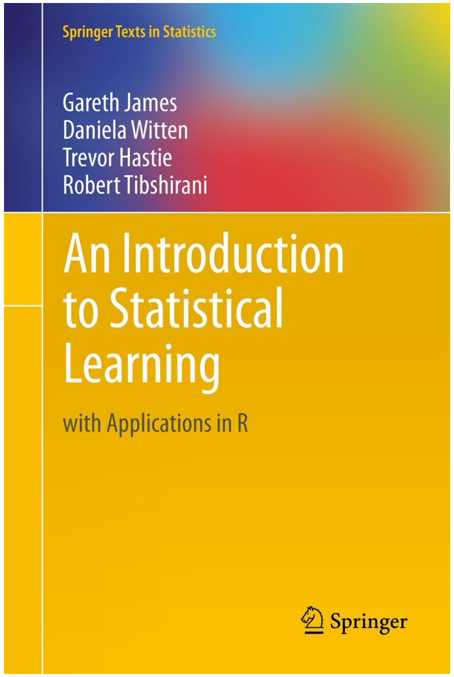


Books 2

An Introduction to Statistical Learning with Applications in R

Authors: Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

Springer

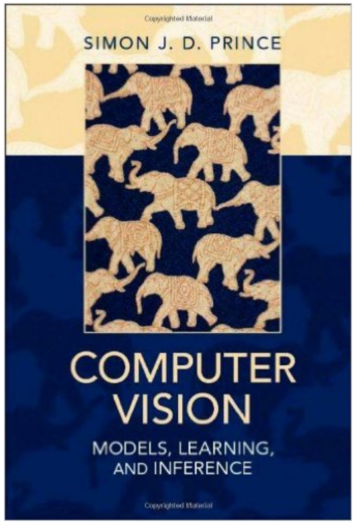


Books 3

Computer vision: models, learning and inference,

Simon J.D. Prince,

Cambridge University



Video lectures

E.g. online course by Andrew Ng (Stanford):

https://www.youtube.com/view_play_list?p=A89DCFA6ADACE599

https://www.youtube.com/watch?v=qeHZOdmJvFU&index=1&list=PLZ9qNFMHZ-A4rycgrgOYma6zxF4BZGGPW

Real time survey

http://www.eduvote.de/en/

Books 4

The Elements of Statistical Learning - Data Mining, Inference, and Prediction
Authors: Trevor Hastie, Robert Tibshirani, Jerome Friedman
Springer