# Machine Learning II

M. Kaschube

V. Ramesh

N. Bertschinger

# Real time survey


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

# Machine learning and pattern recognition

# Organization

- Lecture:

  Wednesday 16:00-18:00
- Tutorial & problem session:

  Wednesday 14:00-16:00

  Weekly or biweekly problem sets

  Tutorials

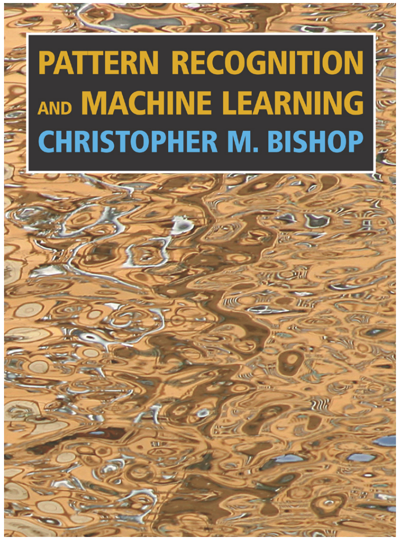
My email: kaschube@fias.uni-frankfurt.de

# Books

Pattern Recognition and Machine Learning

Christopher M. Bishop

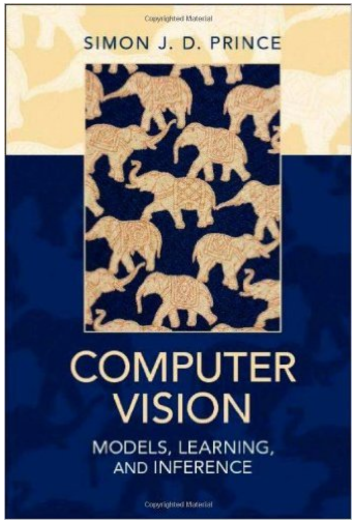
Springer




Computer vision: models, learning and inference,

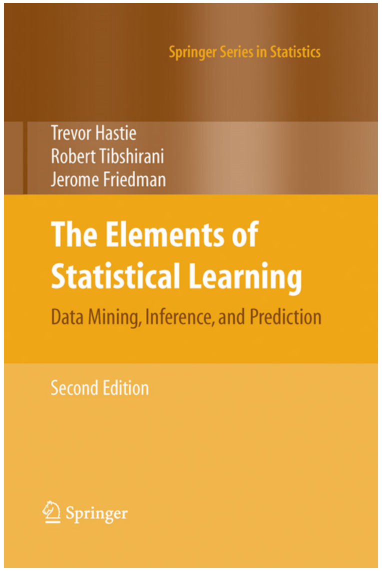
Simon J.D. Prince,

Cambridge University



# Books

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



# Video lectures

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

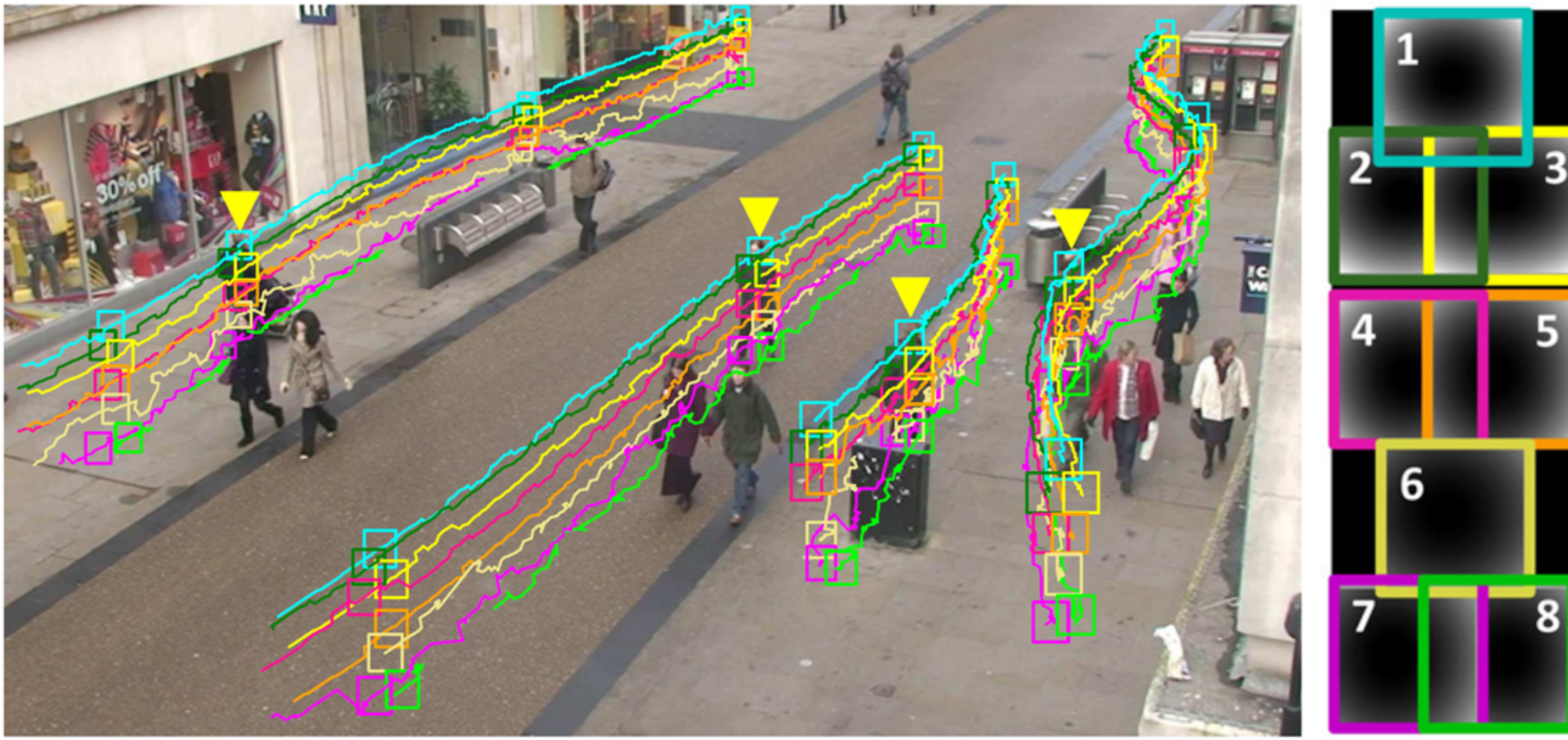
https://www.youtube.com/view_play_list?p=A89DCFA6ADACE599

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

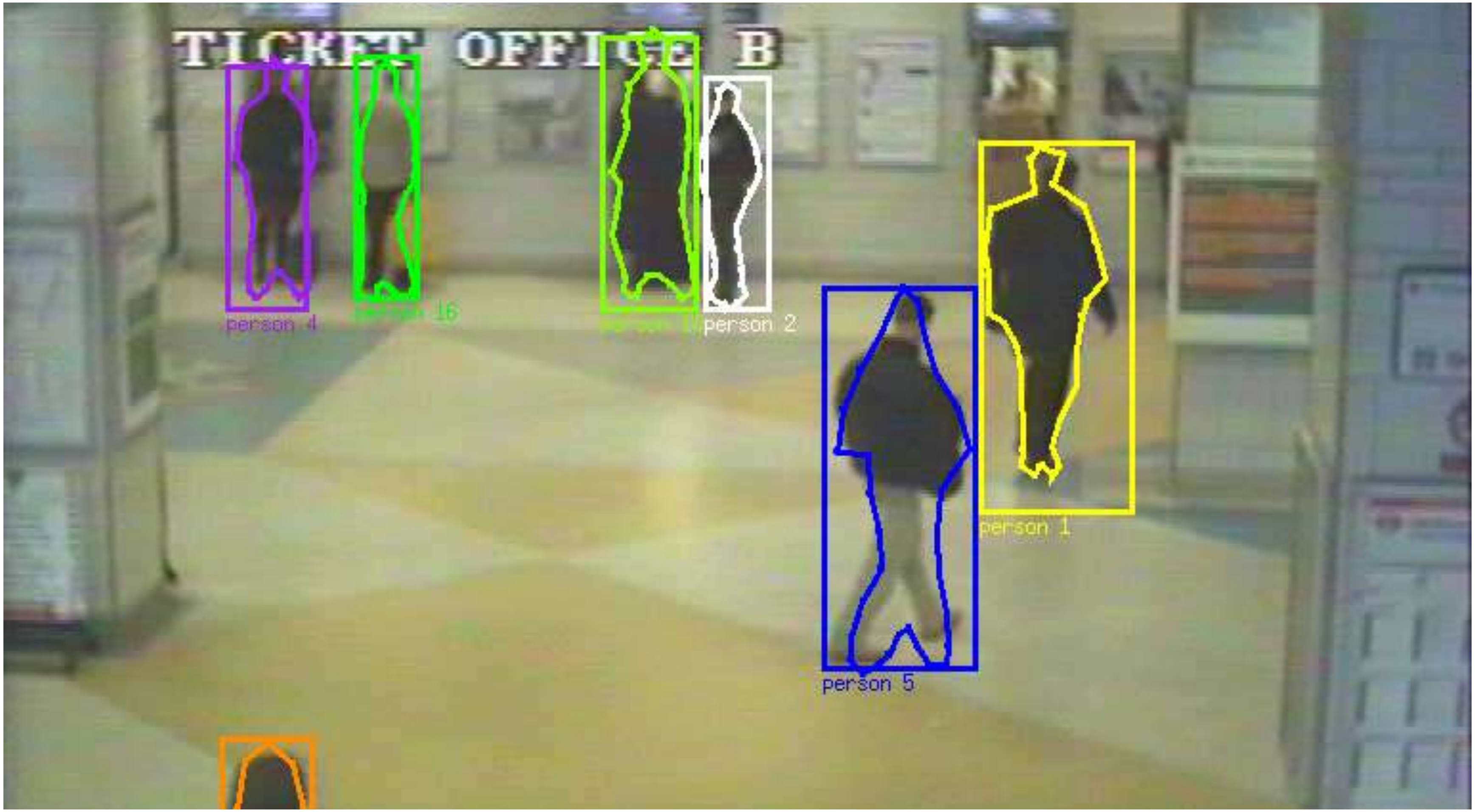
# Stock market

# Surveillance

# Surveillance

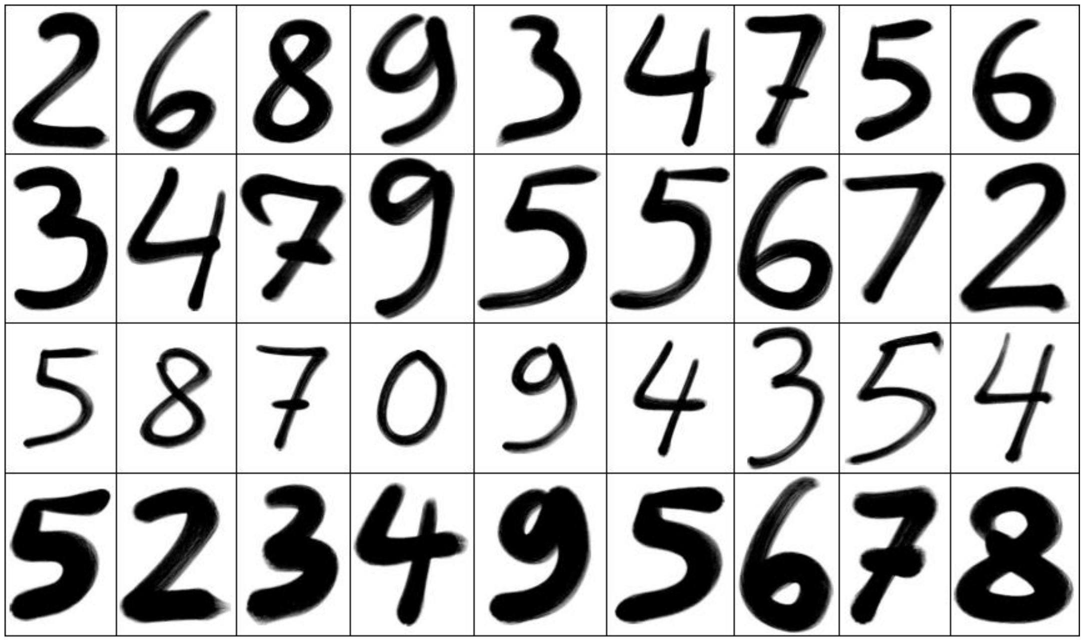


# Surveillance



# 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 |

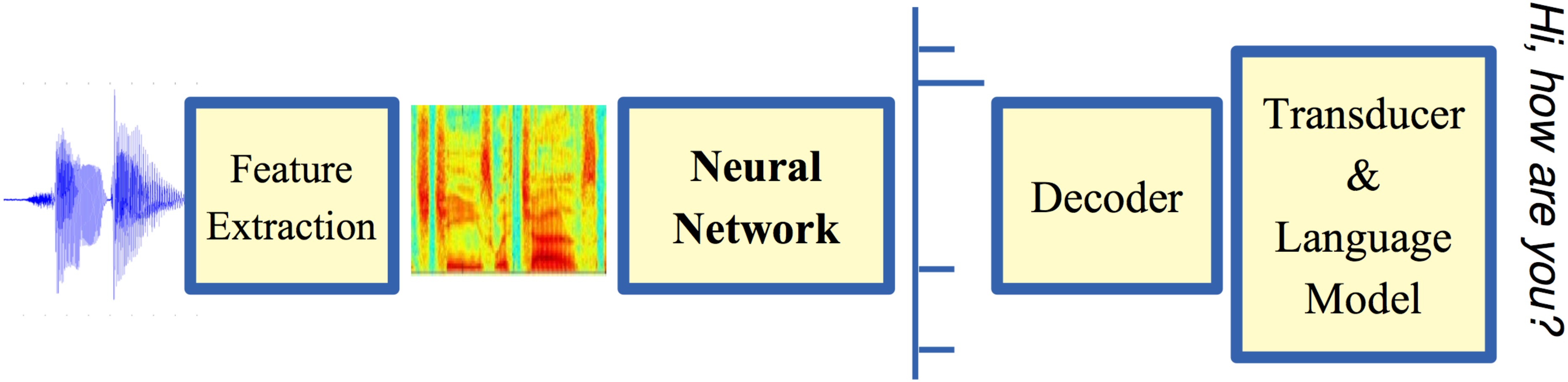
# Speech recognition

# 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



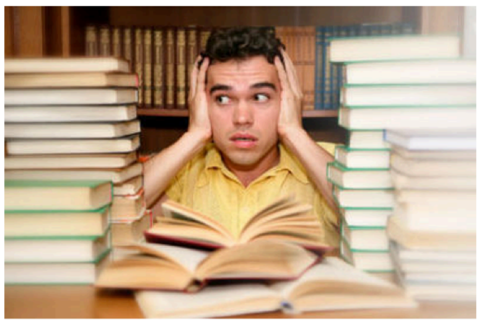

We focus only on the prediction of phone states from short time-windows of spectrogram.

(LeCunn & Ranzato , ICML 2013)

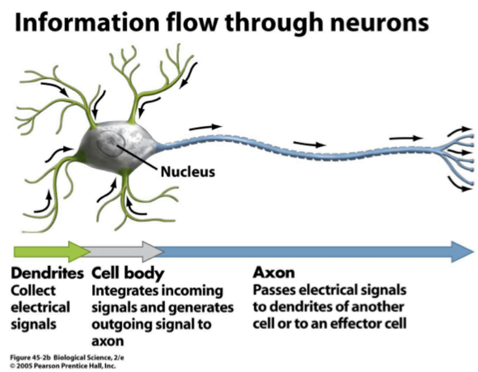
Note: The brain itself is a remarkable pattern recognition system -
and it has provided significant inspiration on artificial recognition systems

# Krizhevsky et al., NIPS 2012

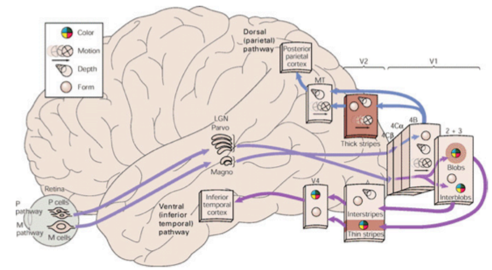
## Deep Learning: Motivated from Human Learning



*Learn* massive data

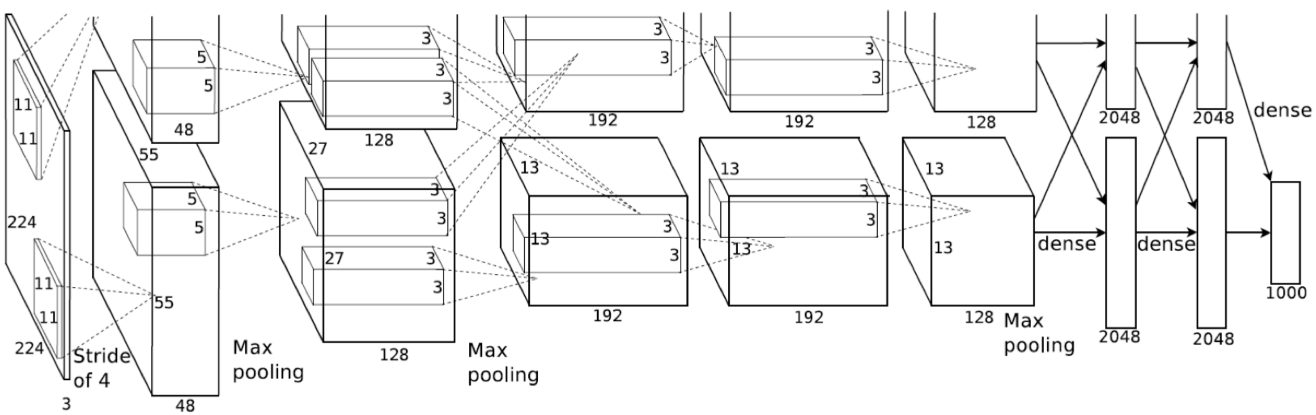



*simple functions*




(Van Essen&Gallant, 1994)

*Multi-layered*




(Krizhevsky et al., 2012)

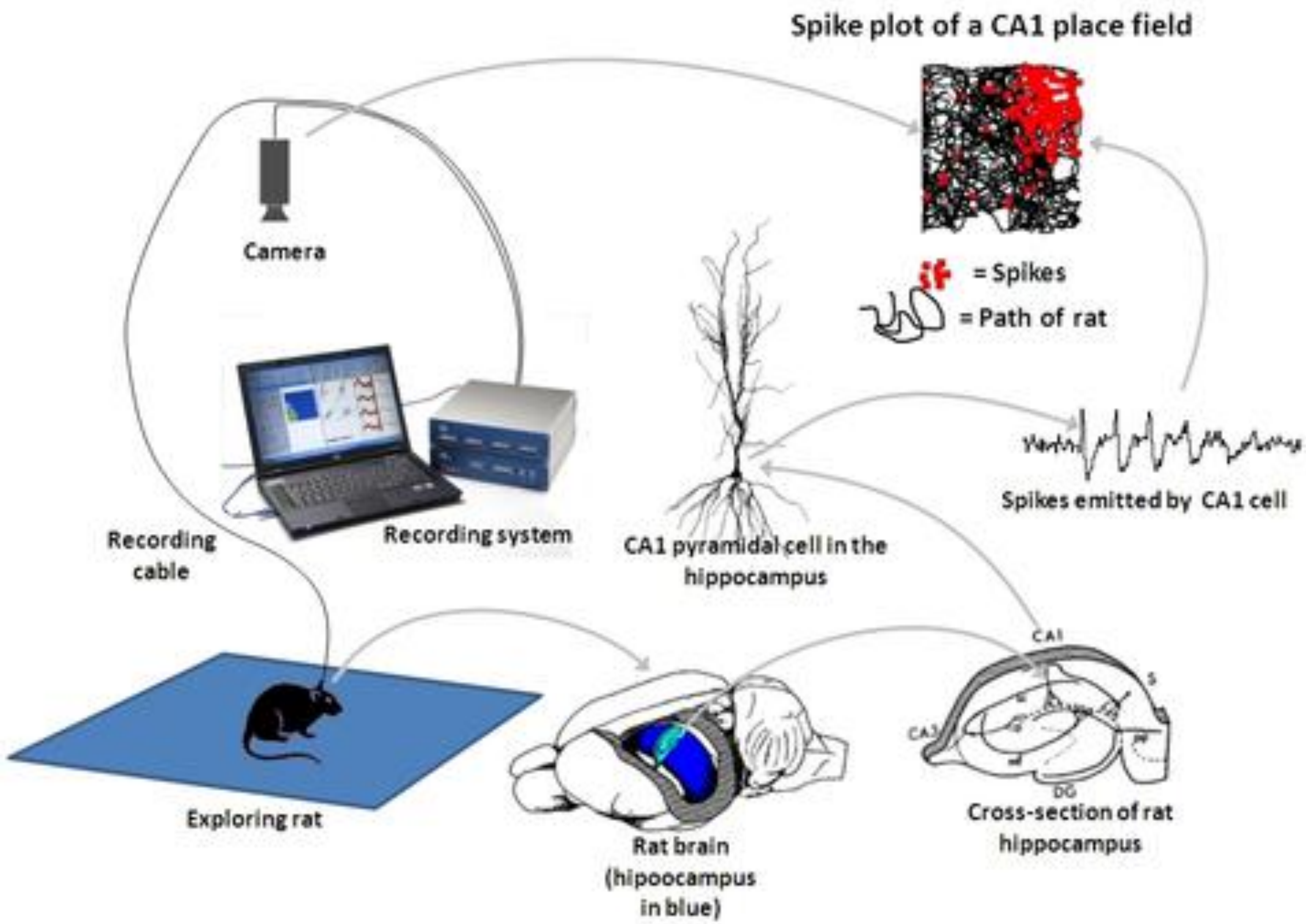
# Try it out for yourself!

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

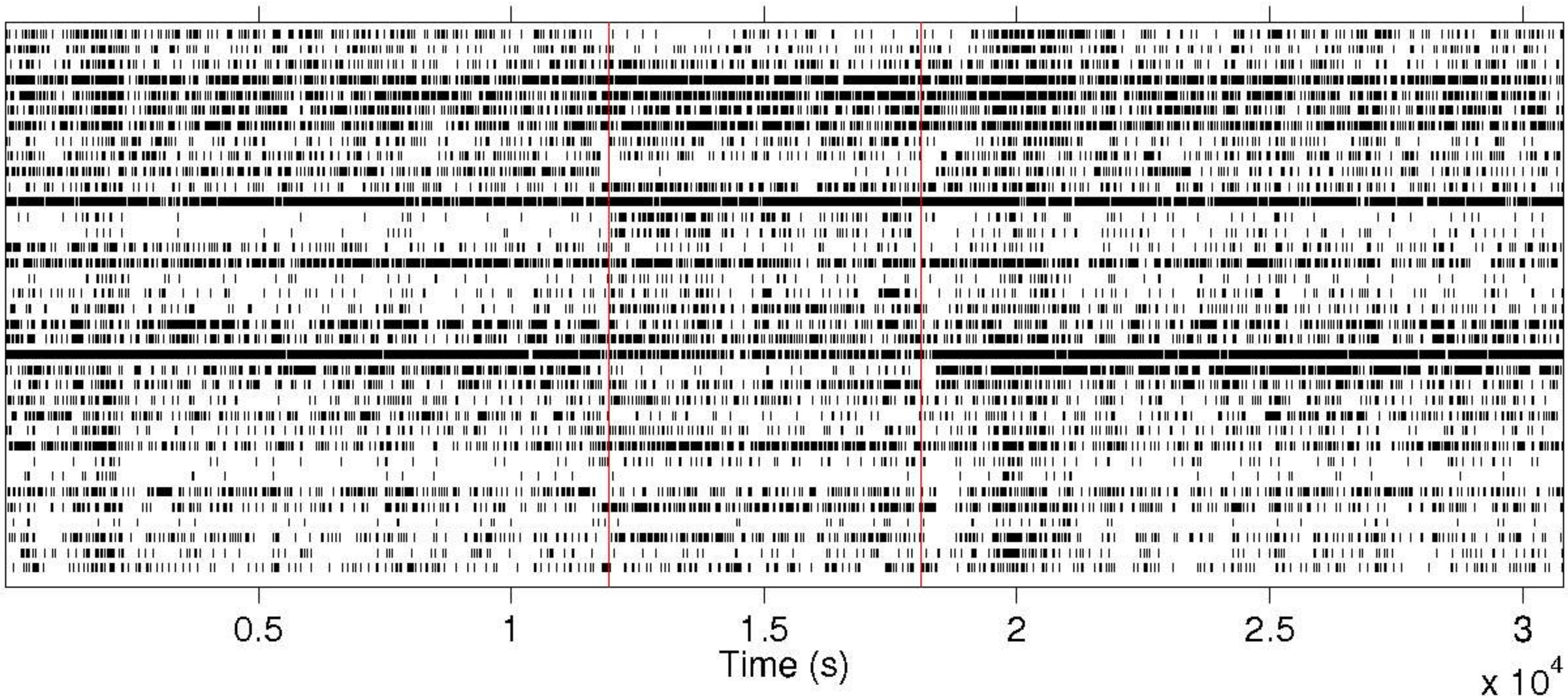
Upload your favorite image and let it recognize its content and/or find similar images!

# Measuring brain activity

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



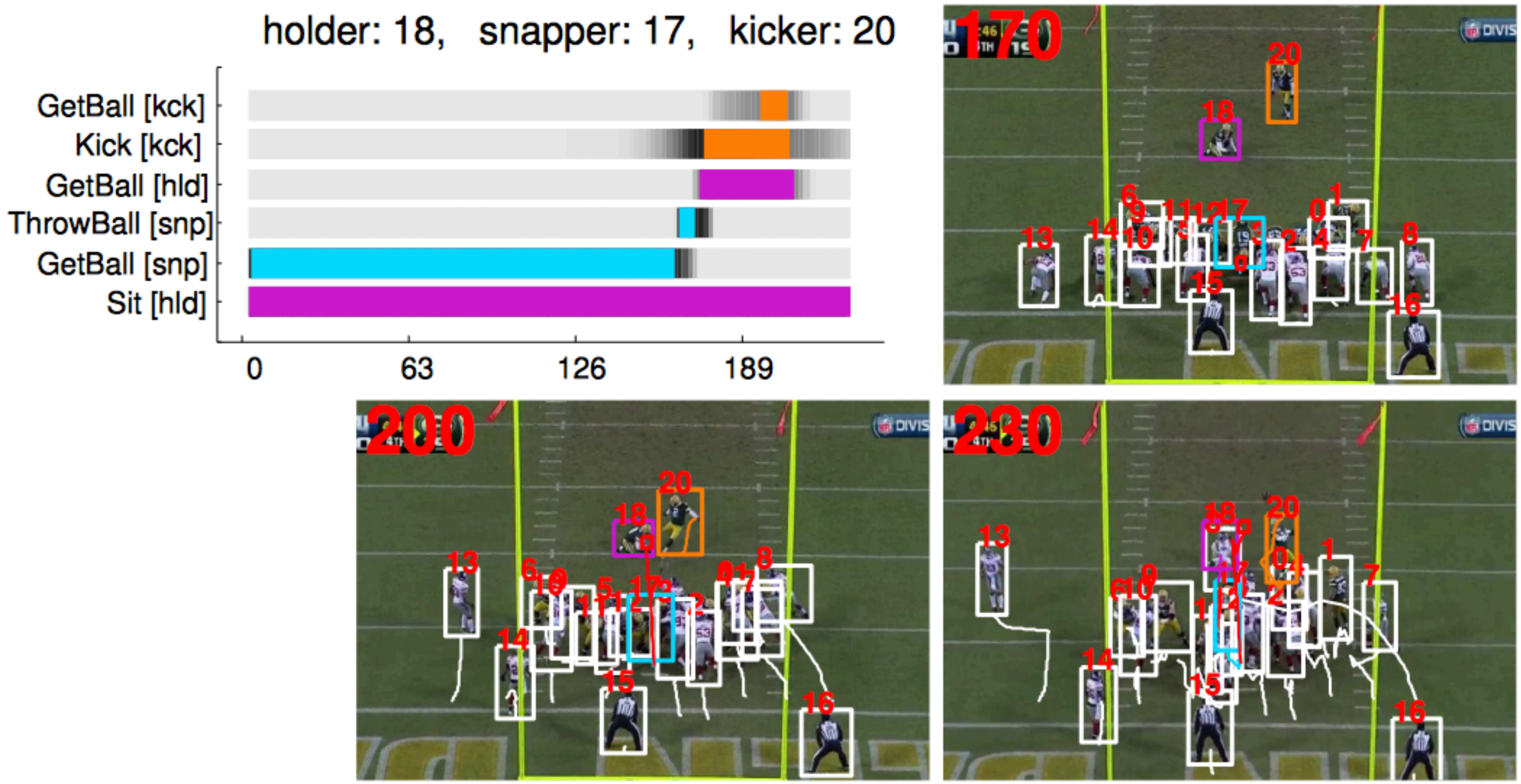
# Brain activity




- 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.

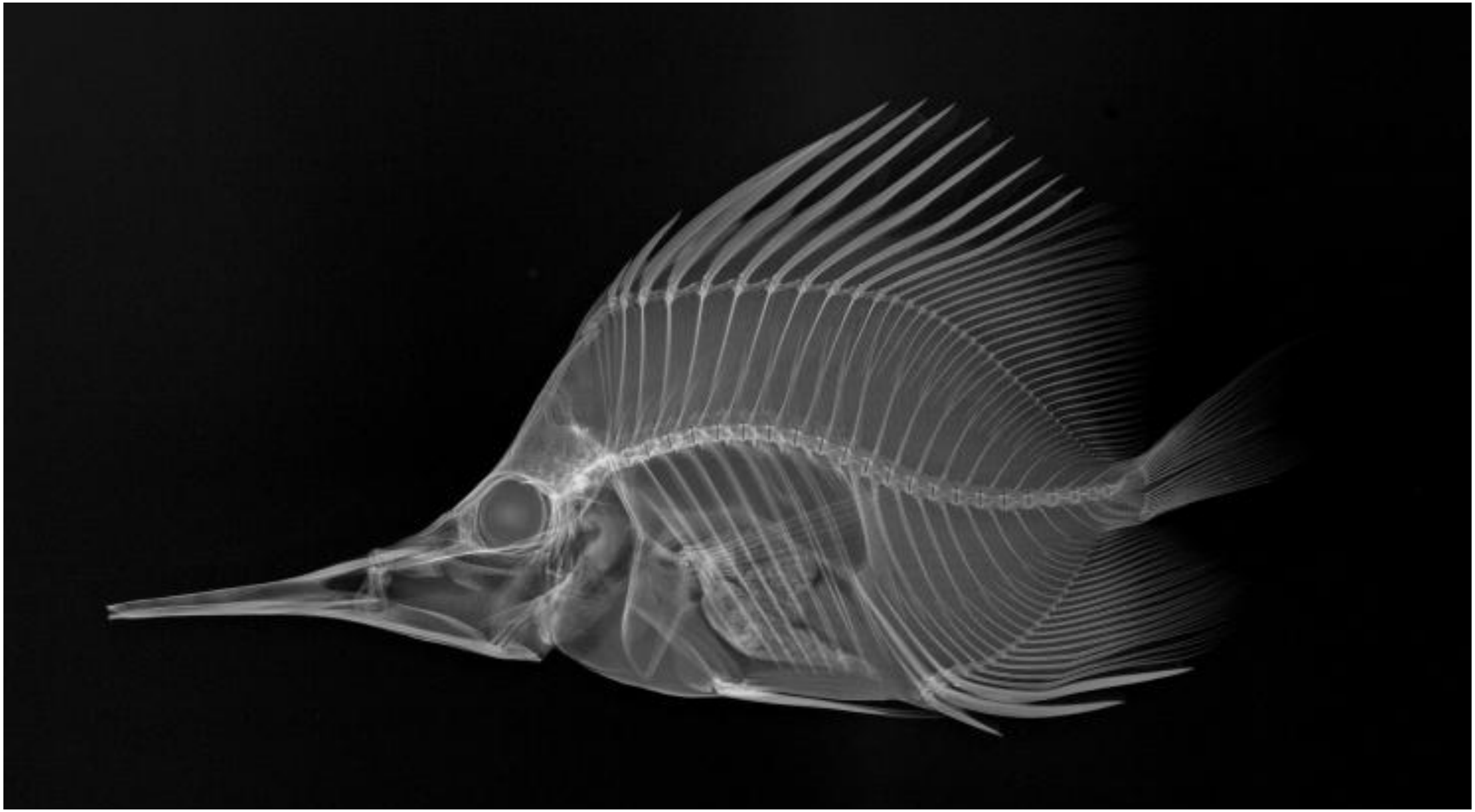
# Object Recognition in Sports

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

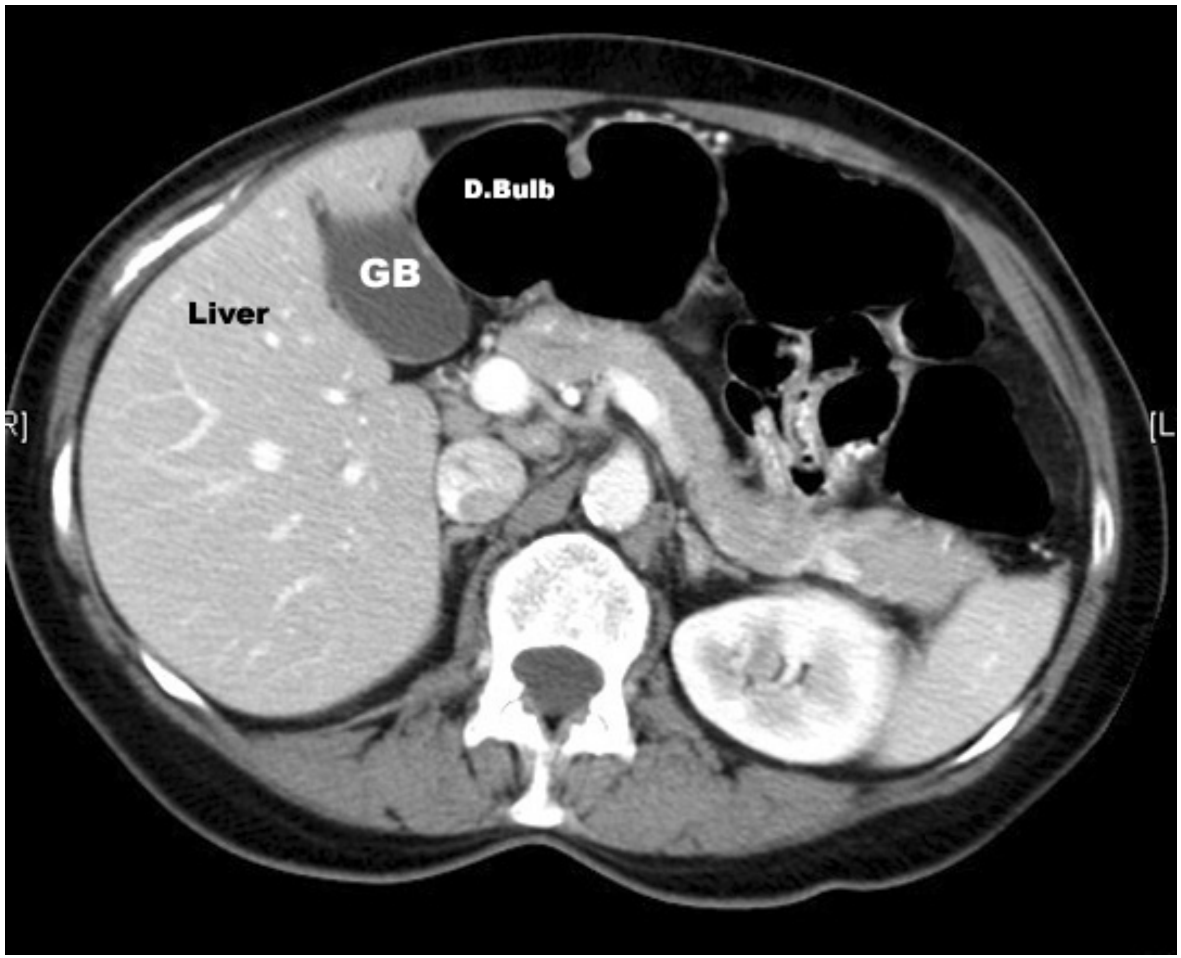



(Kwak et al., CVPR 2013)

# Medical applications



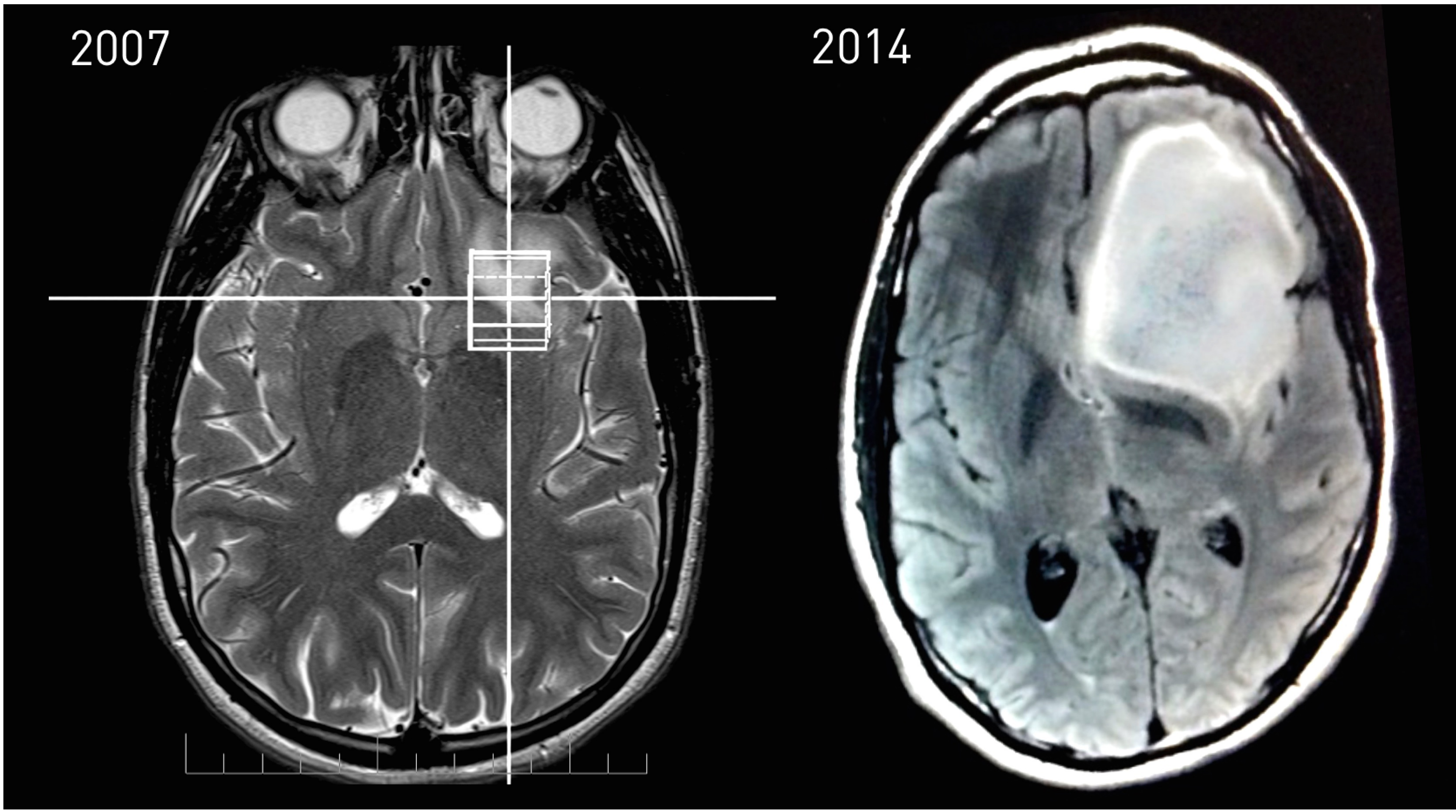
# Computed Tomography (CT)



# 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…

# 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.

# Regression and classification

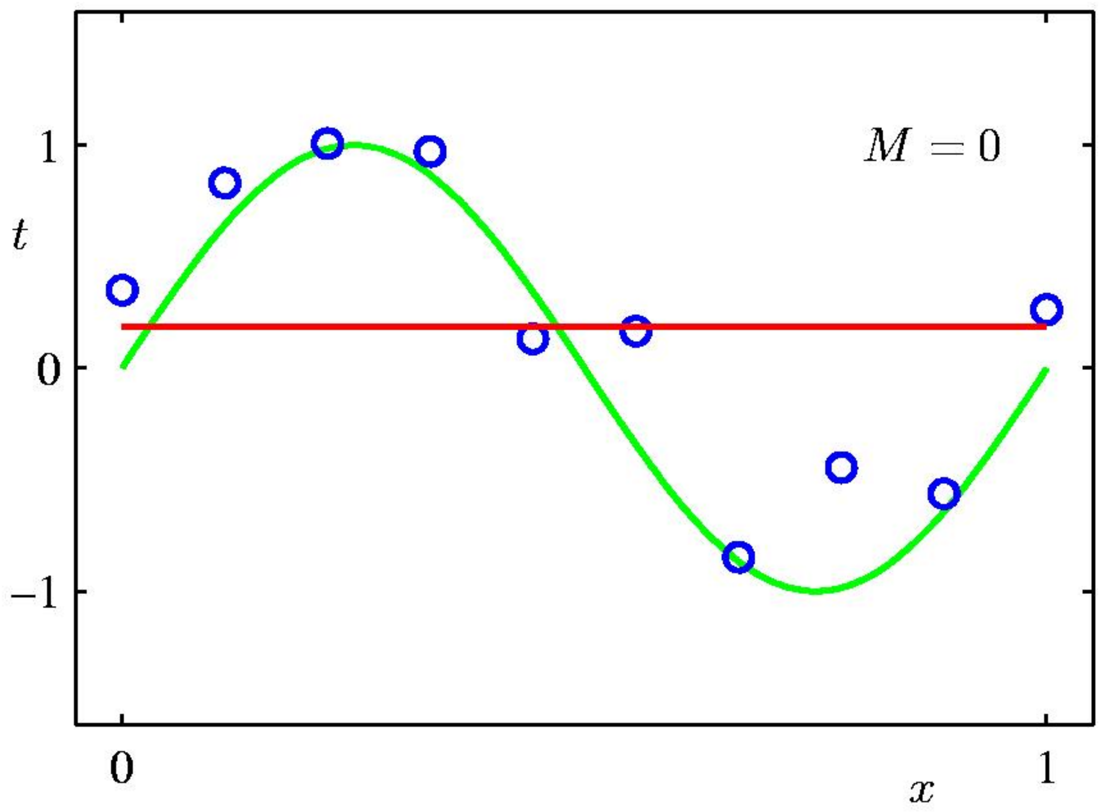
- Regression:

$$x \in [-\infty, \infty], t \in [-\infty, \infty]$$

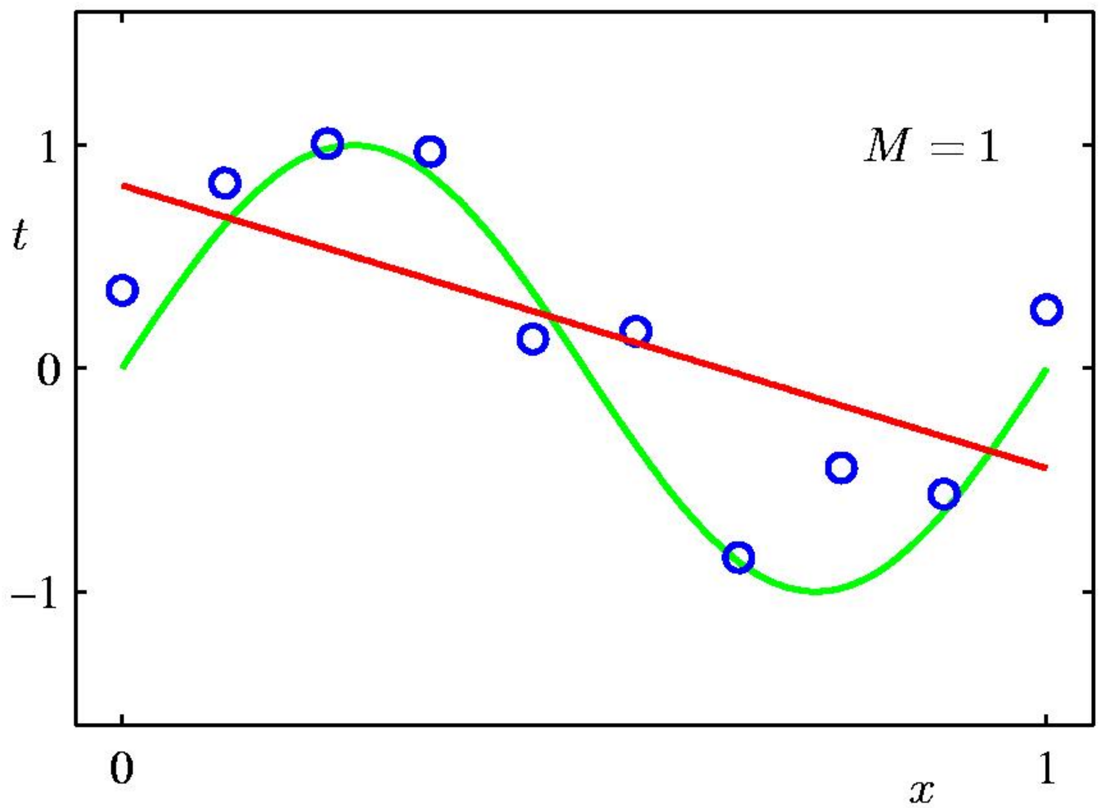
- Classification:

$$x \in [-\infty, \infty], t \in \{0, 1\}$$

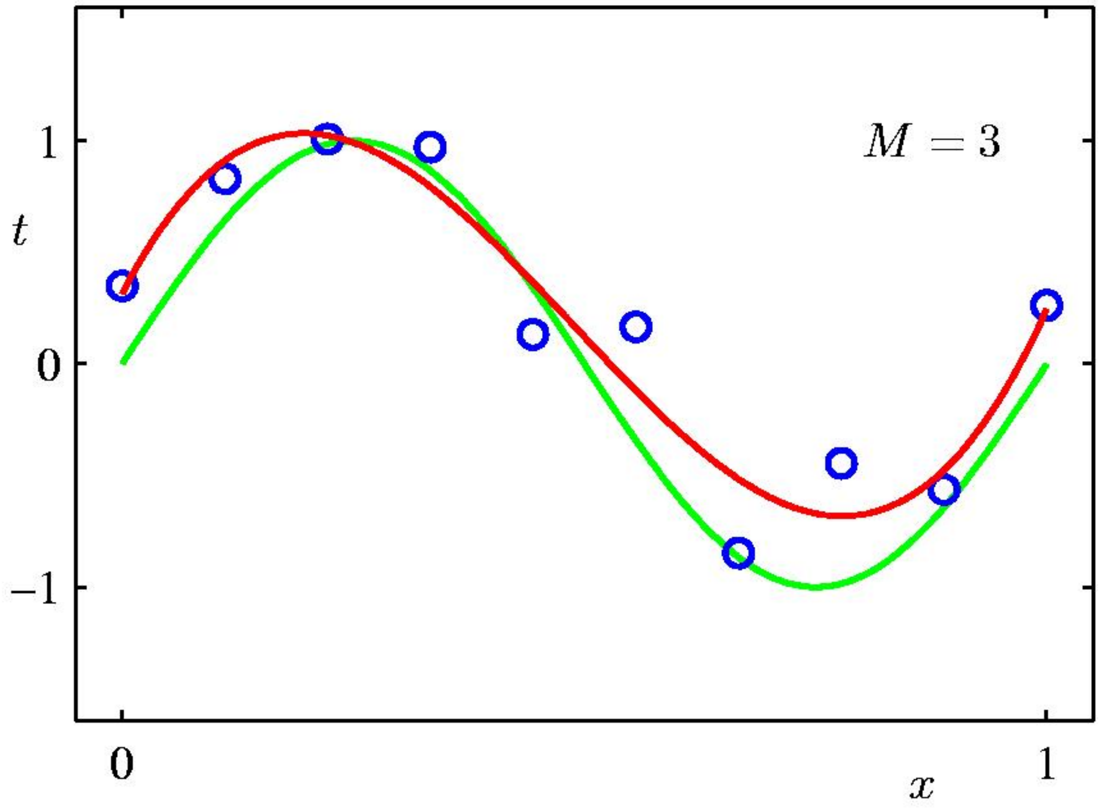
# $0^{th}$ Order Polynomial



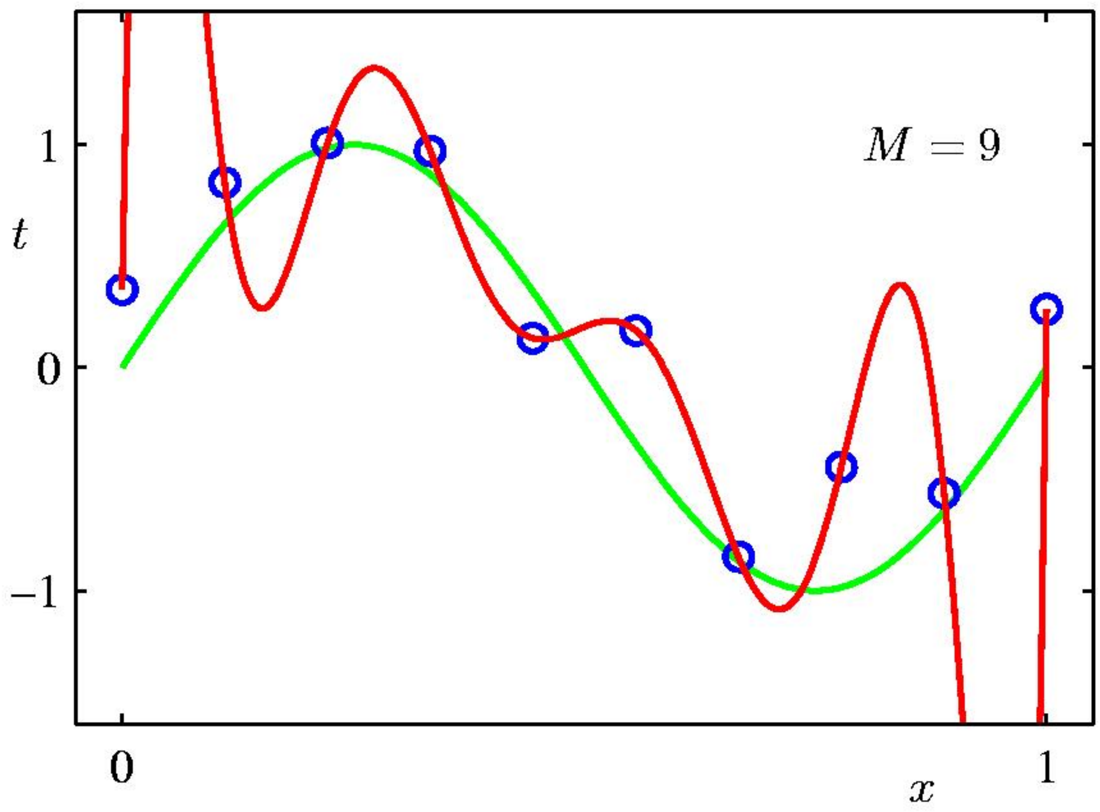
# $1^{st}$ Order Polynomial



# 3rd Order Polynomial



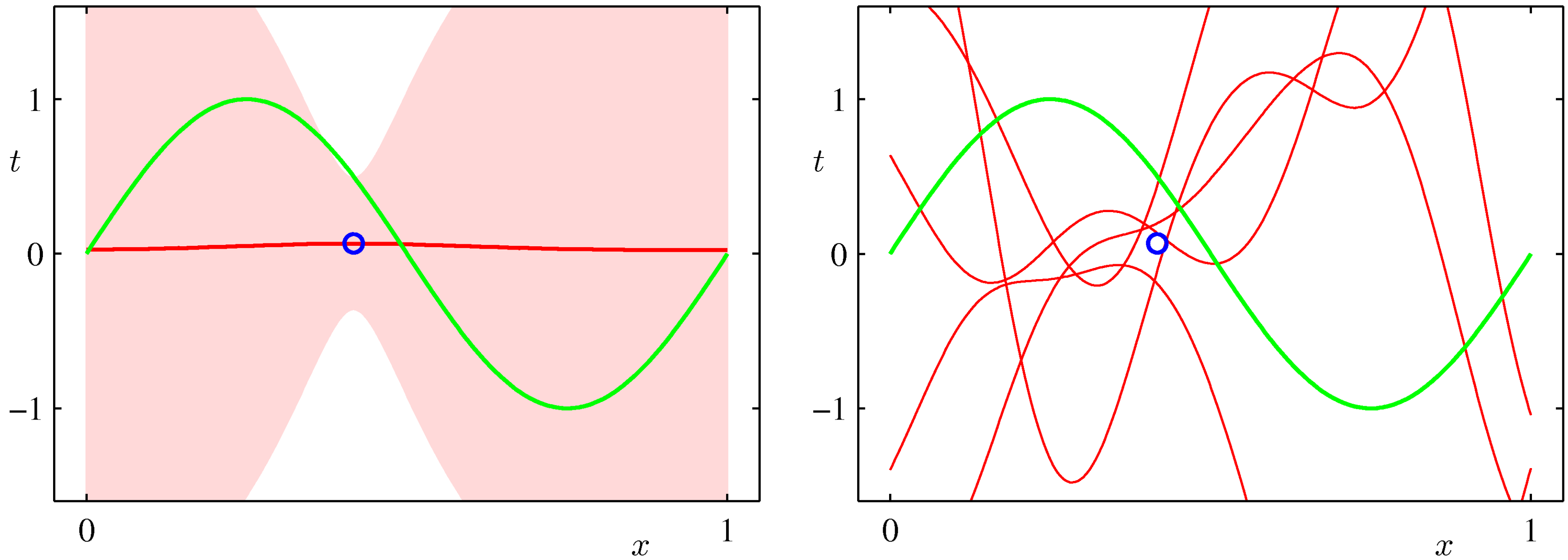
# $9^{th}$ Order Polynomial




Overfitting, a major issue!

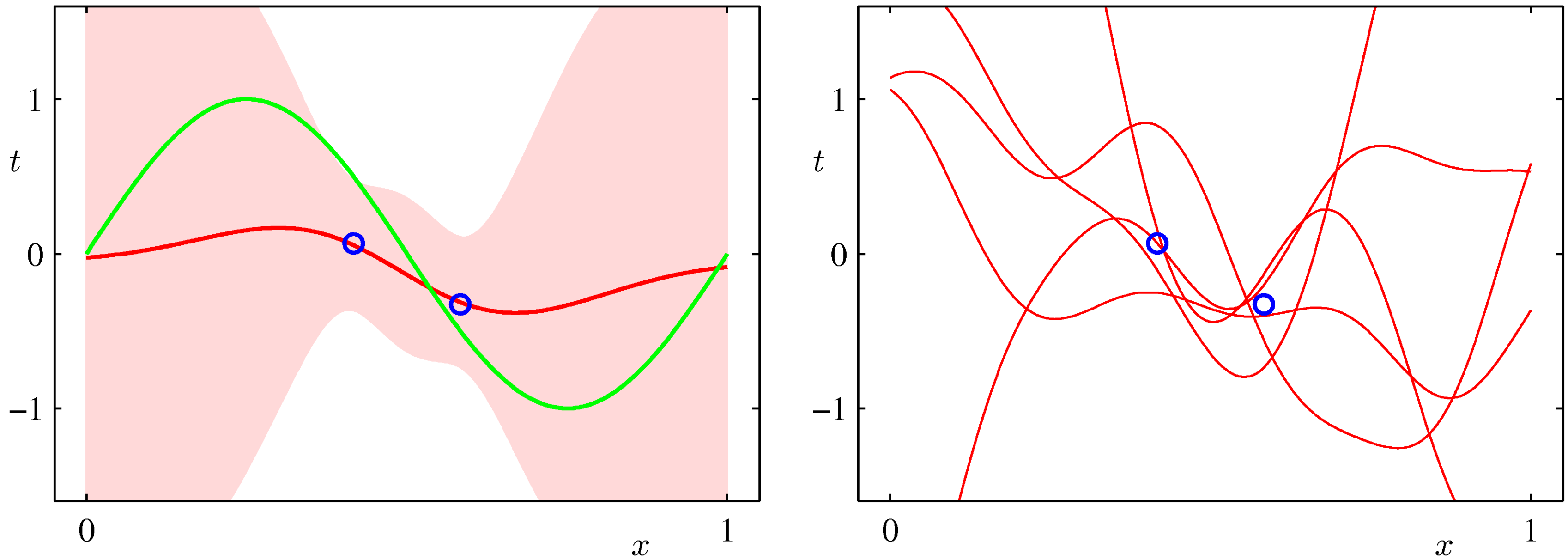
# Bayesian Curve Fitting

- Example: Sinusoidal data, 9 Gaussian basis functions, 1 data point



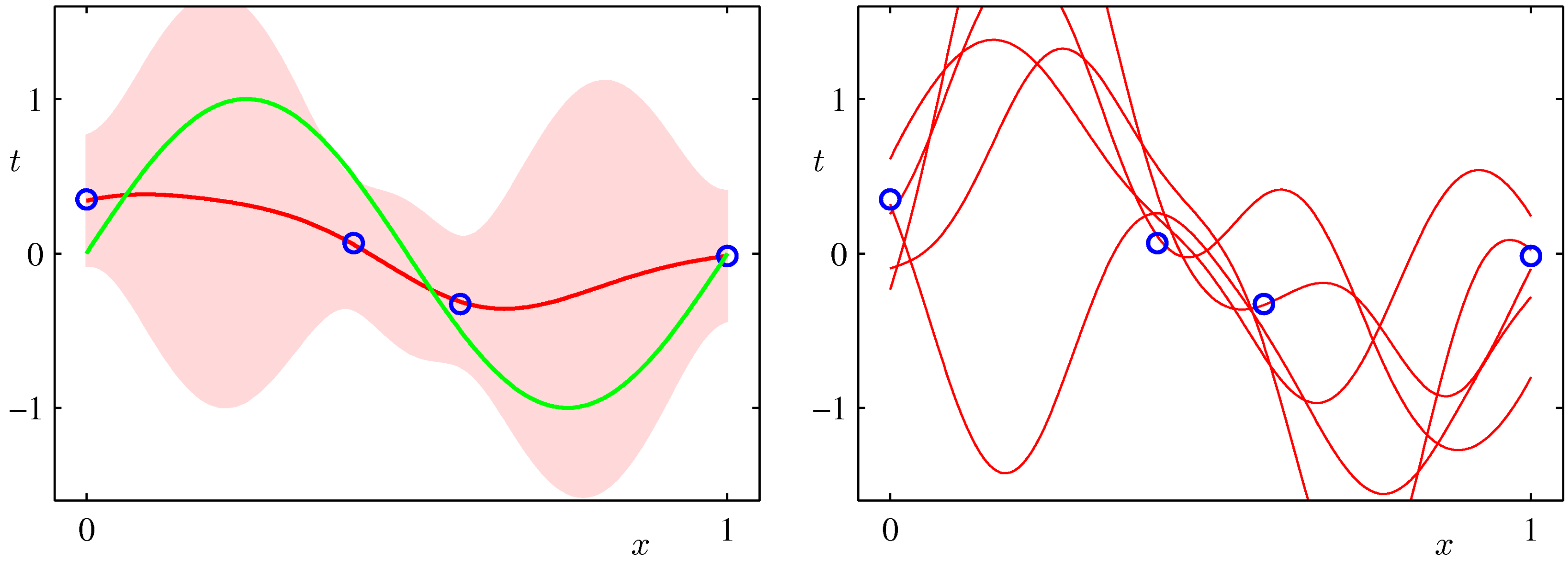
# Bayesian Curve Fitting

- Example: Sinusoidal data, 9 Gaussian basis functions, 2 data points



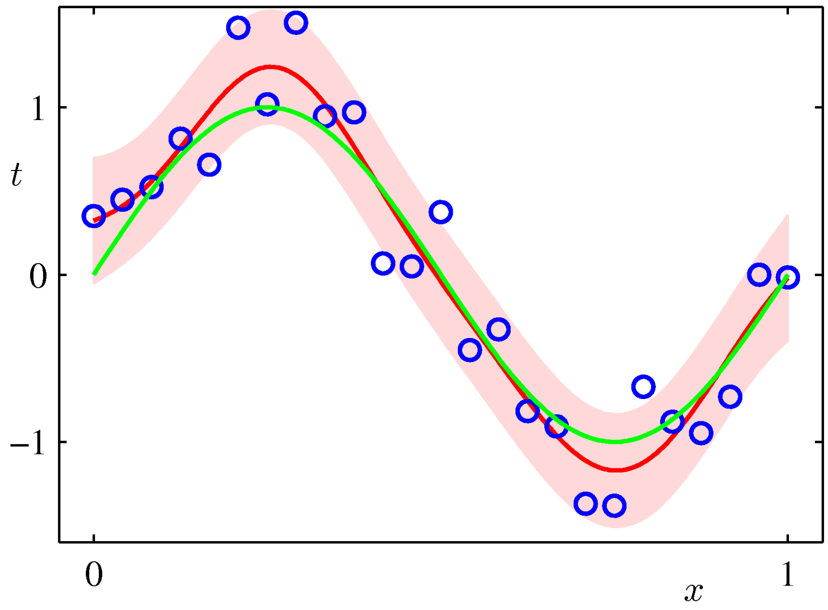
# Bayesian Curve Fitting

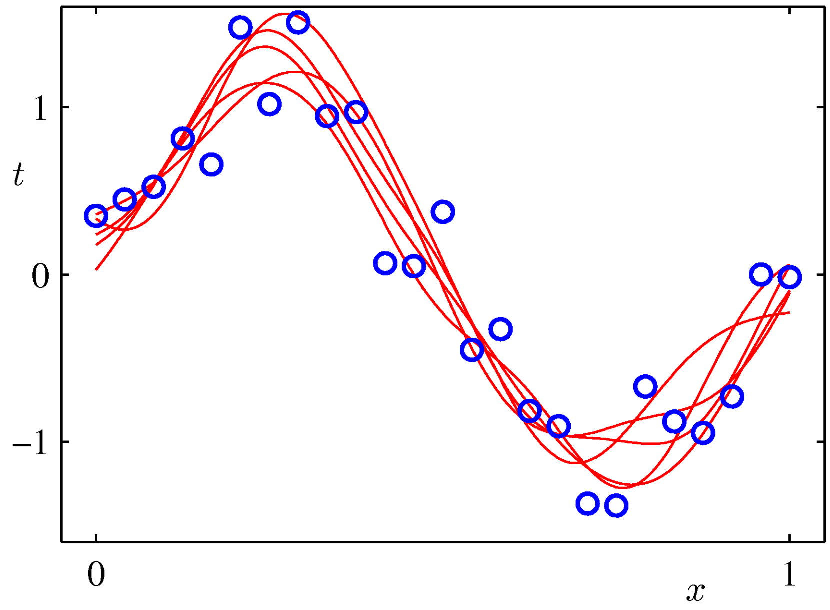
- Example: Sinusoidal data, 9 Gaussian basis functions, 4 data points



# Bayesian Curve Fitting

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



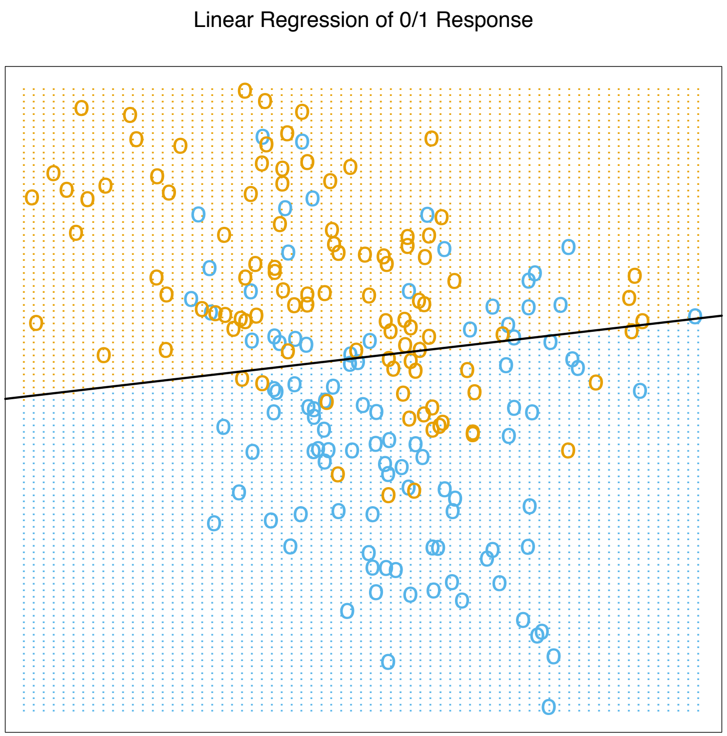



**FIGURE 2.1.** *A classification example in two dimensions. The classes are coded as a binary variable* (BLUE $= 0$, ORANGE $= 1$)*, and then fit by linear regression. The line is the decision boundary defined by* $x^T\hat{\beta} = 0.5$*. The orange shaded region denotes that part of input space classified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

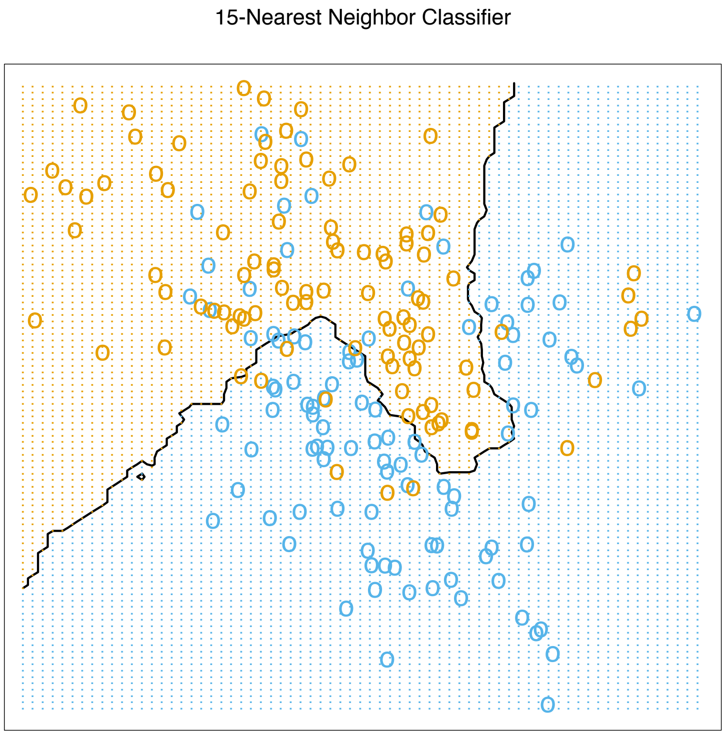



**FIGURE 2.2.** *The same classification example in two dimensions as in Figure 2.1. The classes are coded as a binary variable* (BLUE $= 0$, ORANGE $= 1$) *and then fit by* 15*-nearest-neighbor averaging as in (2.8). The predicted class is hence chosen by majority vote amongst the* 15*-nearest neighbors.*

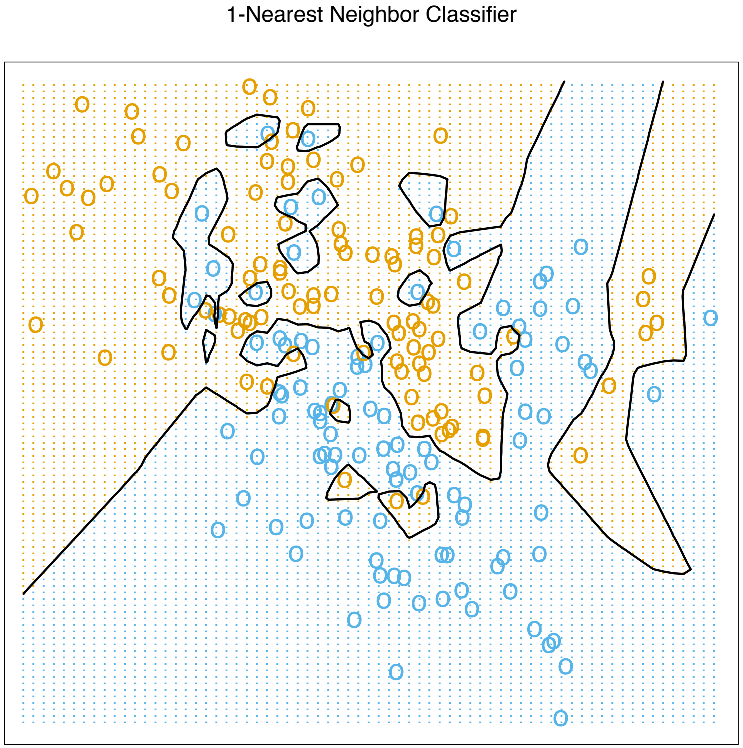



**FIGURE 2.3.** *The same classification example in two dimensions as in Figure 2.1. The classes are coded as a binary variable* ($\texttt{BLUE} = 0, \texttt{ORANGE} = 1$)*, and then predicted by* 1*-nearest-neighbor classification.*

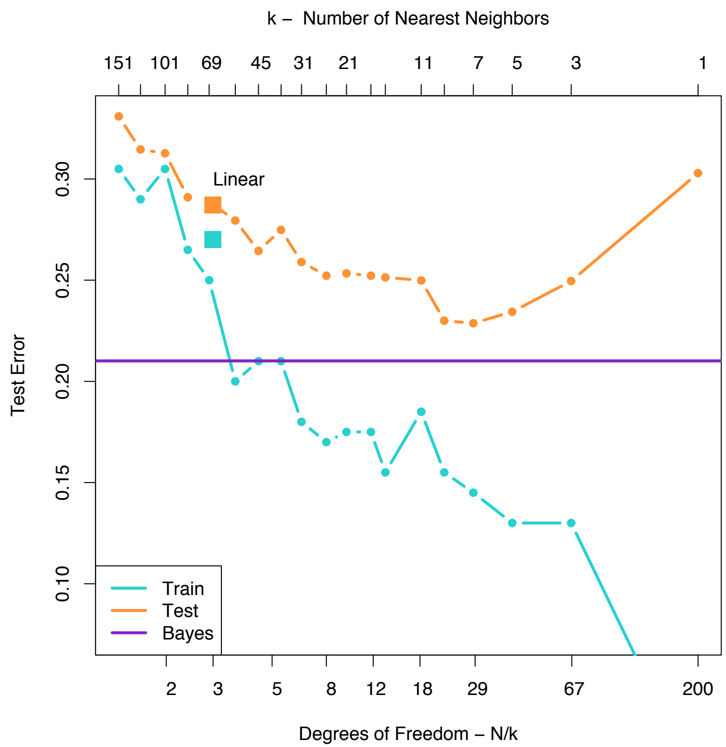



**FIGURE 2.4.** *Misclassification curves for the simulation example used in Figures 2.1, 2.2 and 2.3. A single training sample of size* 200 *was used, and a test sample of size* 10,000*. The orange curves are test and the blue are training error for k-nearest-neighbor classification. The results for linear regression are the bigger orange and blue squares at three degrees of freedom. The purple line is the optimal Bayes error rate.*

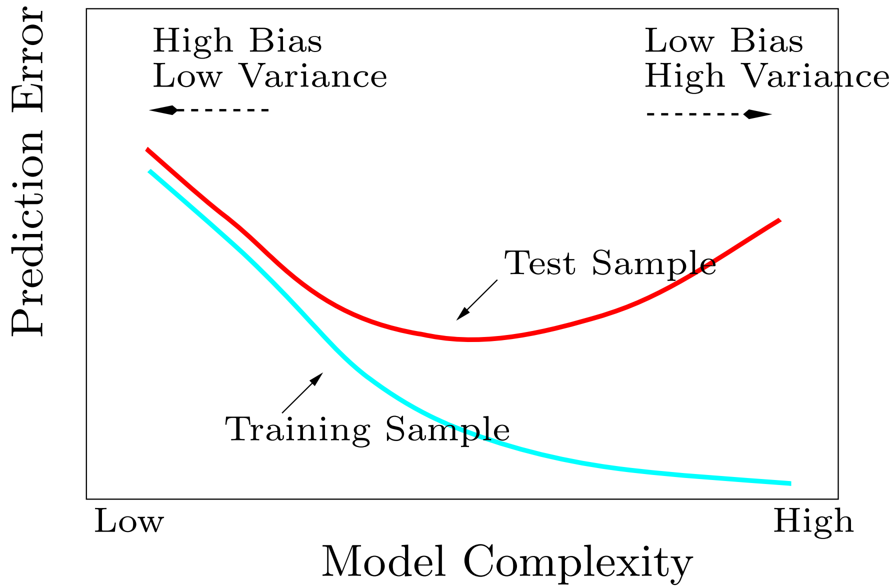



**FIGURE 2.11.** *Test and training error as a function of model complexity.*

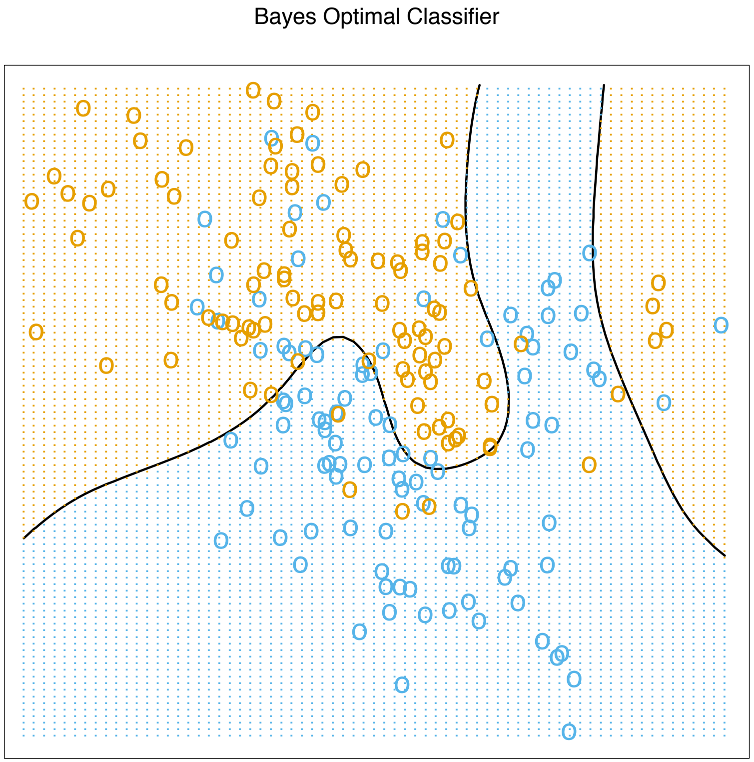



**FIGURE 2.5.** *The optimal Bayes decision boundary for the simulation example of Figures 2.1, 2.2 and 2.3. Since the generating density is known for each class, this boundary can be calculated exactly (Exercise 2.2).*

# Content of course (tentative)

- Intro, recap Machine Learning I

April 18 – May 9: (Prof. Dr. M. Kaschube) (dimensionality reduction, neural networks)

- Continue recap
- Clustering, continuous latent variables
- Neural Networks

May 16 through June 13: (Prof. Dr. V. Ramesh) (Deep Learning and Bayesian Machine Learning, including applications)

- Deep Learning
- Graphical Models and Machine Learning (MRF, CRF, HMMs, Nonparametric Bayesian Models)

June 20 through July 11: (Prof. Dr. N. Bertschinger) (Sampling methods)

- Intro: From rejection to importance Sampling
- Markov-Chain Monte Carlo
- Example: Gibbs sampling for Gaussian mixtures
- Hamiltonian Monte Carlo with short demo of Stan (linear regression or neural net)