ECE228 and SIO209 Machine learning for physical applications, Spring 2018
http://noiselab.ucsd.edu/ECE228/

Professor Peter Gerstoft, Gerstoft@ucsd.edu

TA Mark Wagner, m2wagner@eng.ucsd.edu
TA Paolo Gabriel, pgabriel@eng.ucsd.edu

TA Nima Mirazee, nmirzaee@eng.ucsd.edu

Location: CENTR 109
Time: Monday and Wednesday 5-6:20pm

Research accomplishments. noiselab.ucsd.edu

Genetic Algorithms and Bayesian inversion, sequential fil
- Co-founded geoacoustic inversion (Ross Chapman)
- **Saga**. Combines Bayesian sampling and 7 OA/EM propagation
- Parallel effort in EM atmospheric refractivity Gerstoft (2003).

Ambient noise processing (2004-)=>
- Noise Cross correlation (Sabra, Gerstoft)
- Fathometer (Gerstoft, Siderius)
- Deep impact on seismology

Microseisms (2006-) =>
- Array proc. (Gerstoft 06), body waves (Gerstoft 08),Theory (Tr
- Gerstoft, "Weather bomb" induced seismic signals. **Science** 2
- Antarctic (Bromirski) and Arctic (Worcester) noise

Compressive sensing (2011-)=>
- Yao, Compressive sensing of earthquakes, GRL 2011, **PNAS 20**
- Xenaki, Compressive beamforming, 2014; Yardim (2013), Gers

Machine learning for physical applications

Summary:
- 160 Papers, H-factor 35 (WOS).
- 97 Ocean Acoustics, 19 EM, 44 seismics, 35 SP
- Mentoring a diversified (culture, levels, science interest, scien person acoustics group.
- Funding ONR, NSF GEO & Polar, DOE, visitors.

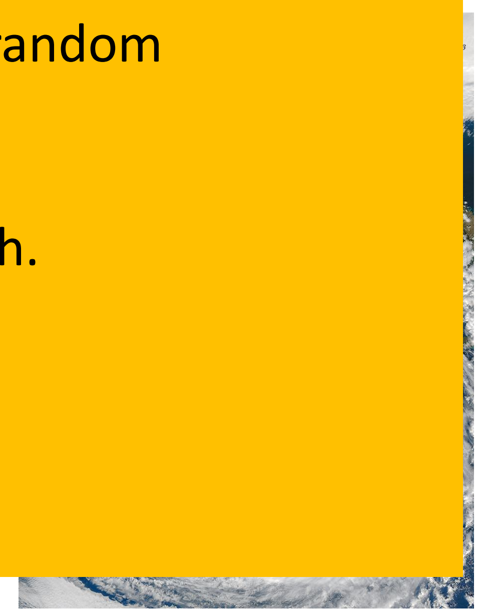
Deterministic,
non-random,
first principles,
stochastic search GA

Random, “Chaos is our friend”,
first principles

first principles
random

Cross-disciplinary, random
Sparse,
random,
deterministic search.

Always Bayes



116 Students with the following specialization 56 EC, 7BE, 1 CE, 4 CS, 6 CU, 1 MA, 15 MC, 5 MC, 1 PY, 3UN
Sit-in students are welcome, but please email me to be signed up for cody

BOOK:
We use **Bishop 2006**, relative to last year Kullback-Leibner, (RNN, LSTM,CNN), RF, sequential estimation.
Murphy 2012 has more detail, but is larger.
Online resources: Sign up for Cosera ML or Stanford Statistical Learning

TA (Mark Wagner, Paolo Gabriel, Nima Mirzaee)

- Mark will coordinate/lead home work (presentation and Cody)
- Paolo coordinate/lead Piazza, Paolo will present his neural brain control, week2
- Nima coordinate projects
- Office hours ECE/SIO, just TA? Wednesday. Peter Maybe on Campus Wednesday

Grade last year (A+ 19, A 20, A- 13, B+ 7, S 1, W 1) [not firm yet]

- 50 Homework
- 50 Project
- 5 class participation

Ideal Class 80 min

- 10 min homework
- 40 min pre or post homework science. With active learning? Bring Clicker
- 30 min applications, projects

Homework

- Automatic graded by Cody in matlab due ABOUT 1 hour before **EVERY** class. First homework April 9
- Please talk about homework, but don’t copy
- Maybe some SciKit Learn on Jupyter Notebook (TA problem)
- Piazza help

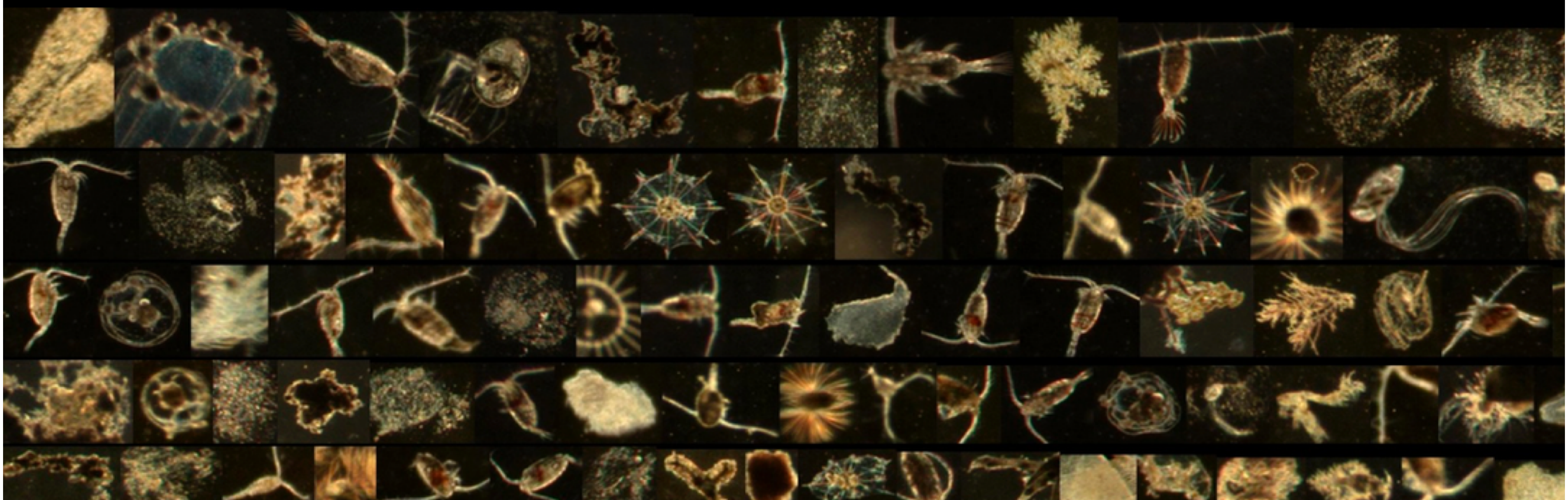
Projects

- **3-4** person groups
- Deliverables: Poster & Report & main code (plus proposal, midterm slide)
- Topics your own or chose form suggested topics
- **Week 3 groups** due to TA Nima (if you don't have a group, ask in week 2 and we can help).
- **Week 5** proposal due. TAs and Peter can approve.
- Proposal: One page: Title, A large paragraph, data, weblinks, references.
- Something physical
- **Week ~7** Midterm slides? Likely presented to a subgroup of class.
- **Week 9/10/11 (likely week 10)** final poster session? Maybe as part of another event.
- Report due Saturday 16 June.

Last years projects:

- Source localization in an ocean waveguide using supervised machine learning, Group3, Group6, Group8, Group10, Group11, Group15 **(from my www)**
- Indoor positioning framework for most Wi-Fi-enabled devices, Group1
- MyShake Seismic Data Classification, Group2 **(from my www)**
- Multi Label Image Classification, Group4. **(Kaggle Use satellite data to track the human footprint in the Amazon rainforest**)
- Face Recognition using Machine Learning, Group7
- Deep Learning for Star-Galaxy Classification, Group9
- Modeling Neural Dynamics using Hidden Markov Models, Group12
- Star Prediction Based on Yelp Business Data And Application in Physics, Group13 **(non physics…)**
- Si K edge X-ray spectrum absorption interpretation using Neural Network, Group14
- Plankton Classification Using VGG16 Network, Group16 **(from my www)**
- A Survey of Convolutional Neural Networks: Motivation, Modern Architectures, and Current Applications in the Earth and Ocean Sciences, Group17 **(NO data, BAD)**
- Use satellite data to track the human footprint in the amazon rainforest, Group18 **(Kaggle Use satellite data to track the human footprint in the Amazon rainforest**)
- Automatic speaker diarization using machine learning techniques, Group19
- Predicting Coral Colony Fate with Random Forest, Group20

Transfer Learning and Deep Feature Extraction for Planktonic Image Data Sets



Eric C. Orenstein[1] and Oscar Beijbom[2]

[1] Scripps Institution of Oceanography – University of California San Diego

[2] Department of Electrical Engineering and Computer Science – University of California Berkeley

IEEE Winter Conference on Applications of Computer Vision 2017 – Paper ID #313

Qingkai Kong is from Berkeley, I have 3GB of data and examples of analysis by students there

RESEARCH ARTICLE

EARTH SCIENCES

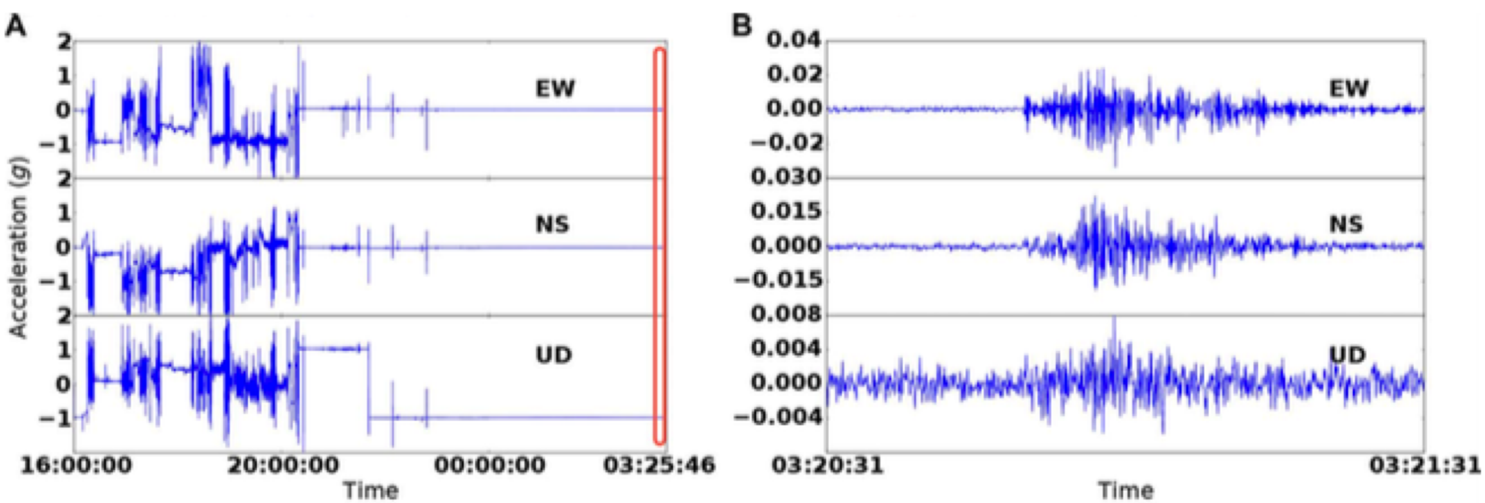
MyShake: A smartphone seismic network for earthquake early warning and beyond

10.1126/sciadv.1501055

Qingkai Kong,[1]* Richard M. Allen,[1] Louis Schreier,[2] Young-Woo Kwon[3]

Large magnitude earthquakes in urban environments continue to kill and injure tens to hundreds of thousands of people, inflicting lasting societal and economic disasters. Earthquake early warning (EEW) provides seconds to minutes of warning, allowing people to move to safe zones and automated slowdown and shutdown of transit and other machinery. The handful of EEW systems operating around the world use traditional seismic and geodetic networks

RESEARCH ARTICLE

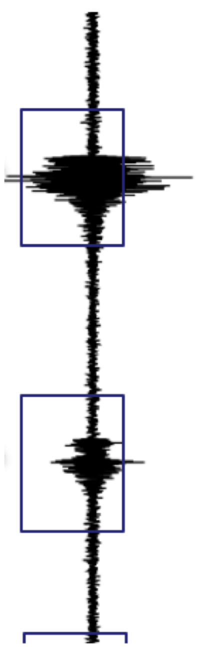


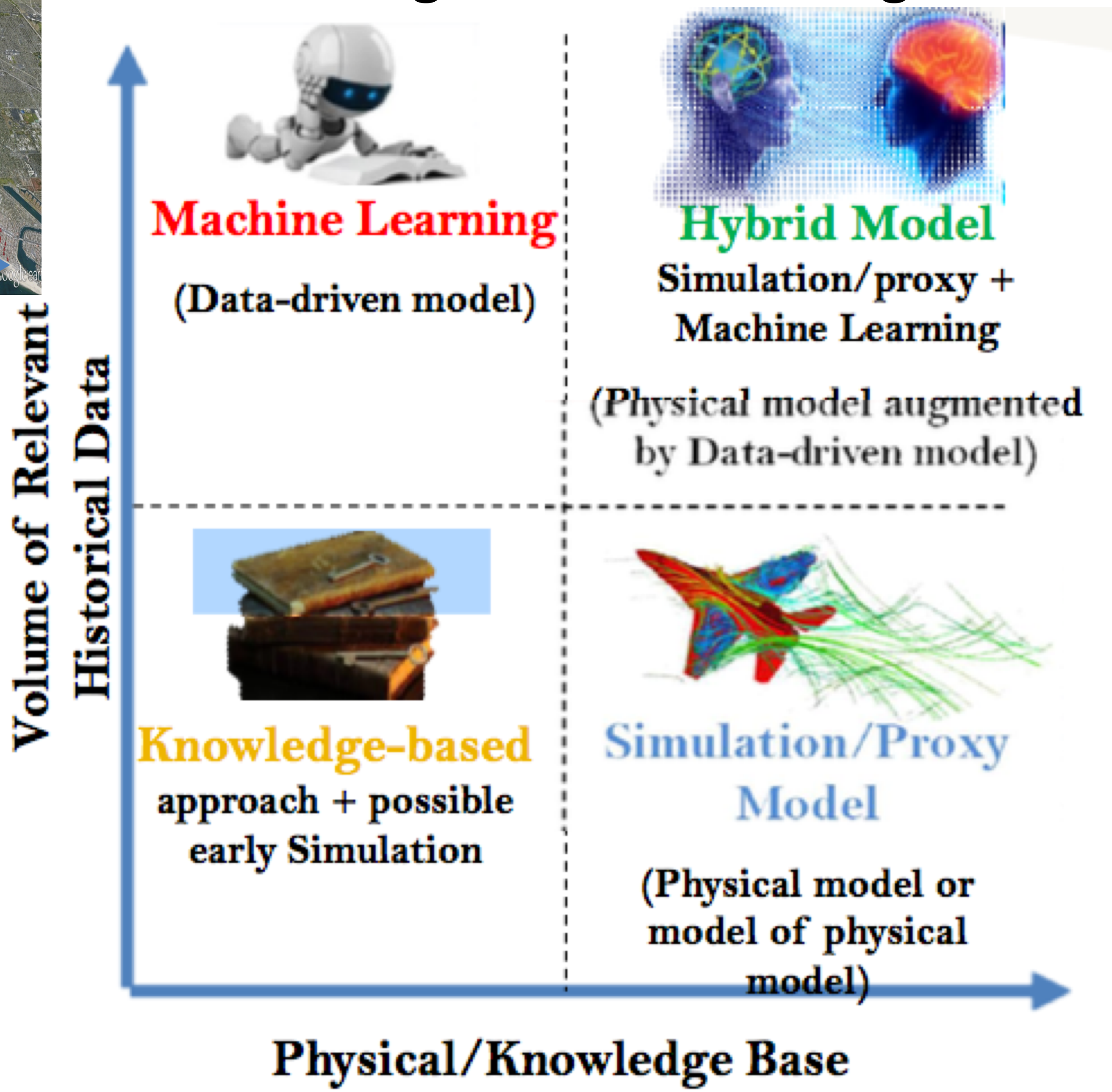
First principles vs Data driven

	First principles	Data driven
Data	Small data	Big data to train
Domain expertise	High reliance on domain expertise	Results with little domain knowledge
Fidelity/ Robustness	Universal link can handle non-linear complex relations	Limited by the range of values spanned by training data
Adaptability	Complex and time consuming derivation to use new relations	Rapidly adapt to new problems
Interpretability	Parameters are physical!	Physically agnostic, limited by the rigidity of the functional form

Perceived Importance. **SIO** **SignalProc** **Peter** **Google**

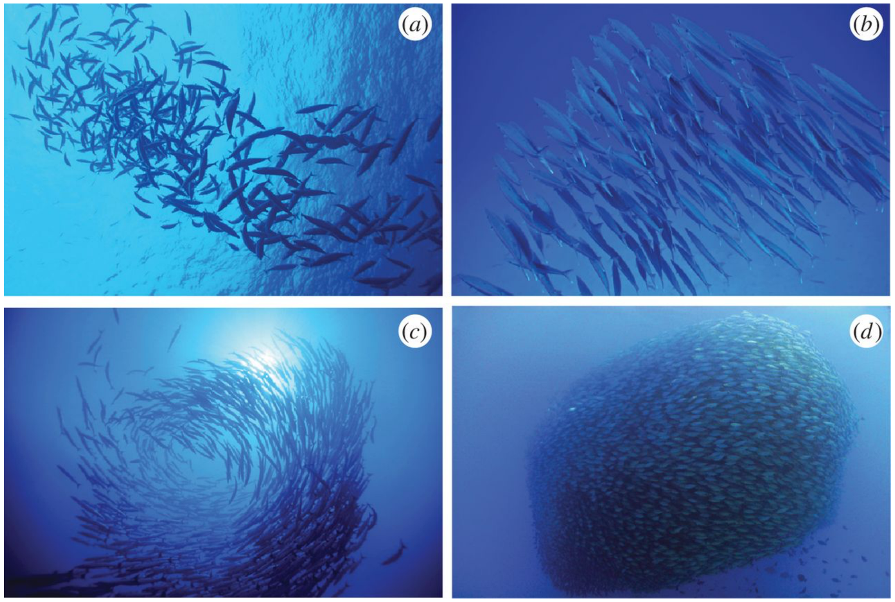
Machine learning versus knowledge based

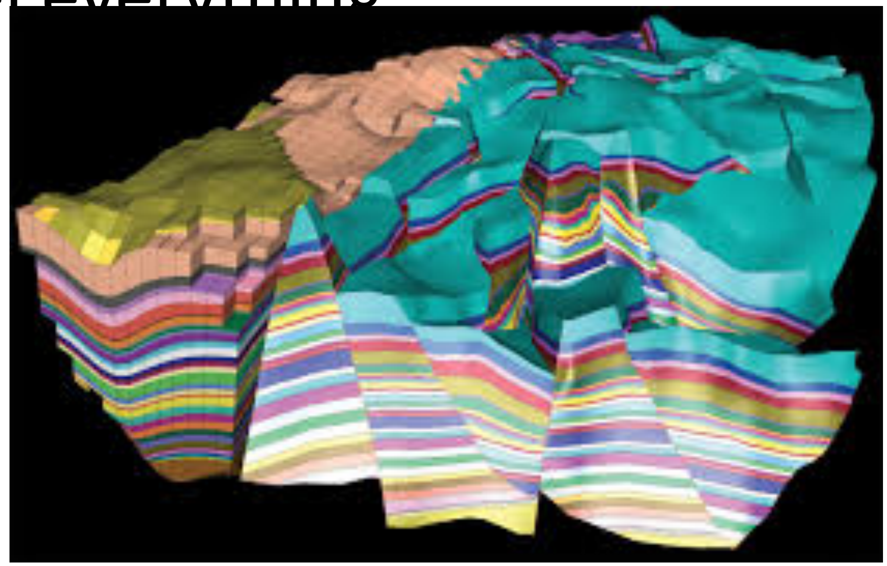




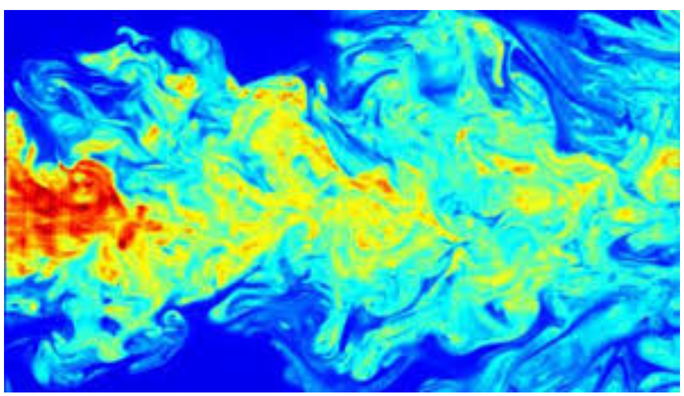
3D spectral elements

We can’t model everything



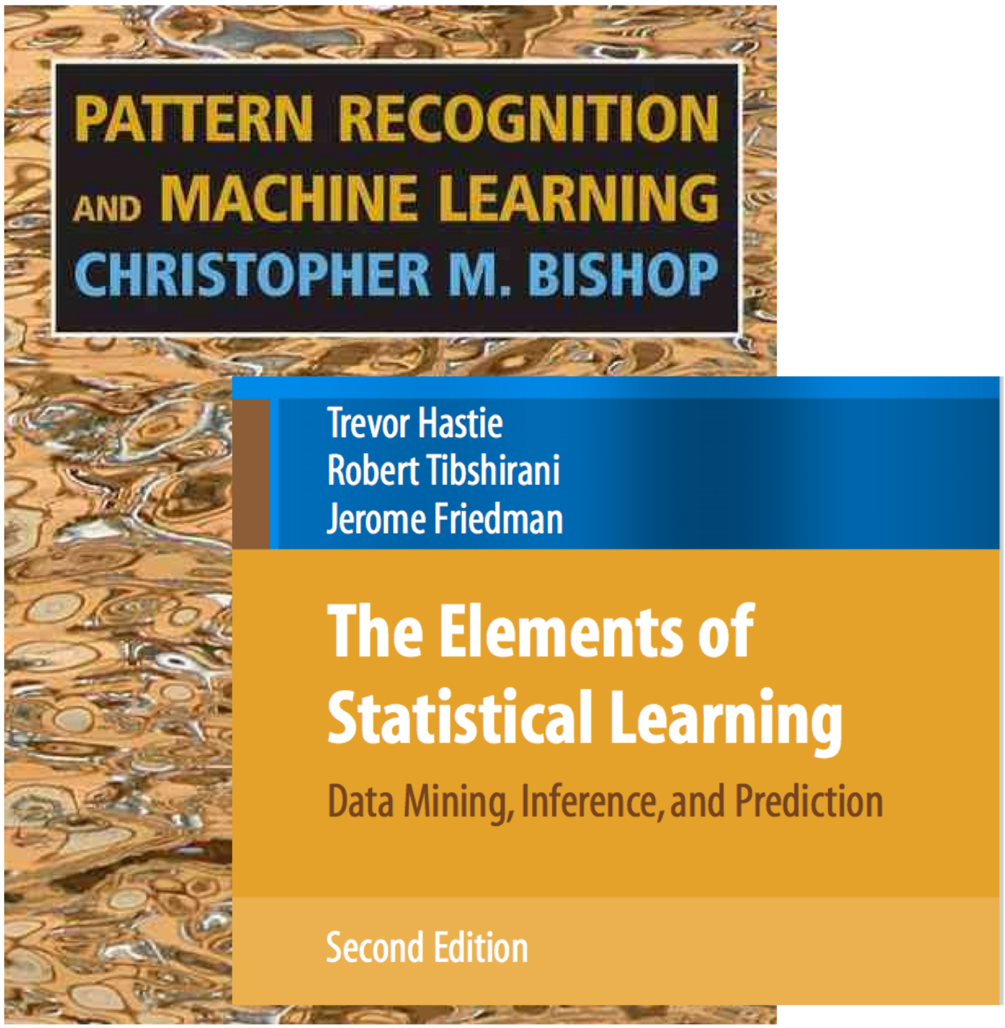


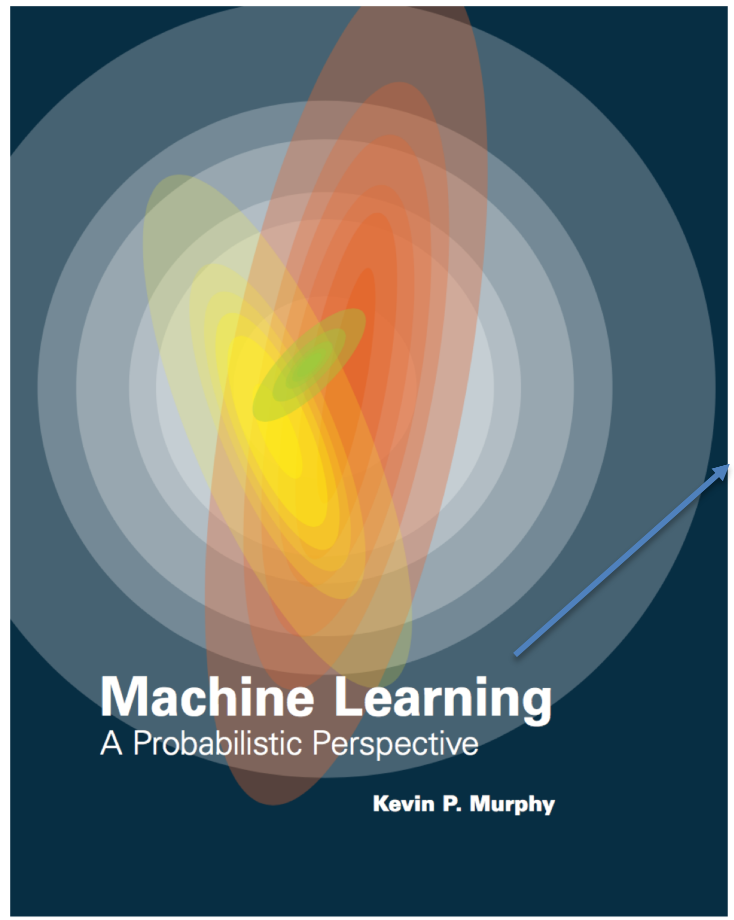
Predict received acoustic field in **turbulent** media



Machine Learning for physical Applications
noiselab.ucsd.edu

Murphy: "…**the best way to make machines that can learn from data is to use the *tools of probability theory*, which has been the mainstay of statistics and engineering for centuries.**"





Learning:
The view from different fields

- Engineering: signal processing, system identification, adaptive and optimal control, information theory, robotics, ...
- Computer Science: Artificial Intelligence, computer vision, information retrieval, ...
- Statistics: learning theory, data mining, learning and inference from data, ...
- Cognitive Science and Psychology: perception, movement control, reinforcement learning, mathematical psychology, computational linguistics, ...
- Computational Neuroscience: neuronal networks, neural information processing, ...
- Economics: decision theory, game theory, operational research, ...

Physical science is missing!
ML cannot replace physical understanding.
It might improve or find additional trends

Machine learning is interdisciplinary focusing on both mathematical foundations and practical applications of systems that learn, reason and act.

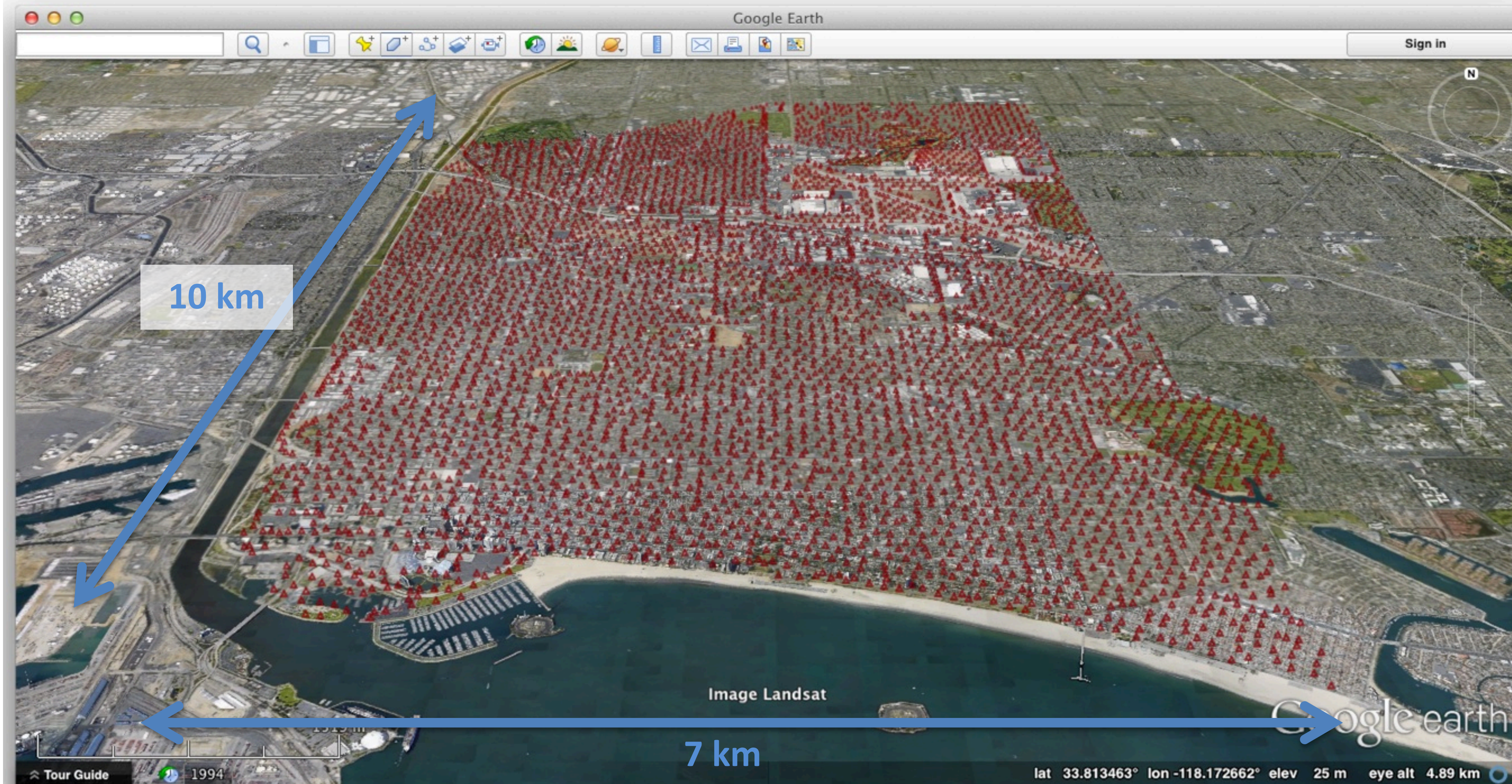
What is Machine Learning?

Many related terms:

- Pattern Recognition
- Neural Networks
- Data Mining
- Adaptive Control
- Statistical Modelling
- Data analytics / data science
- Artificial Intelligence
- Machine Learning

Big data

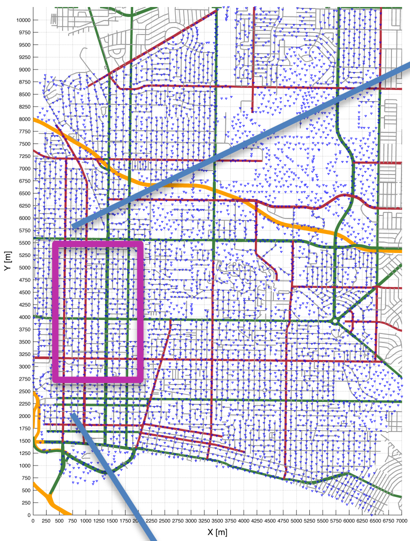
Why we got interested in traffic

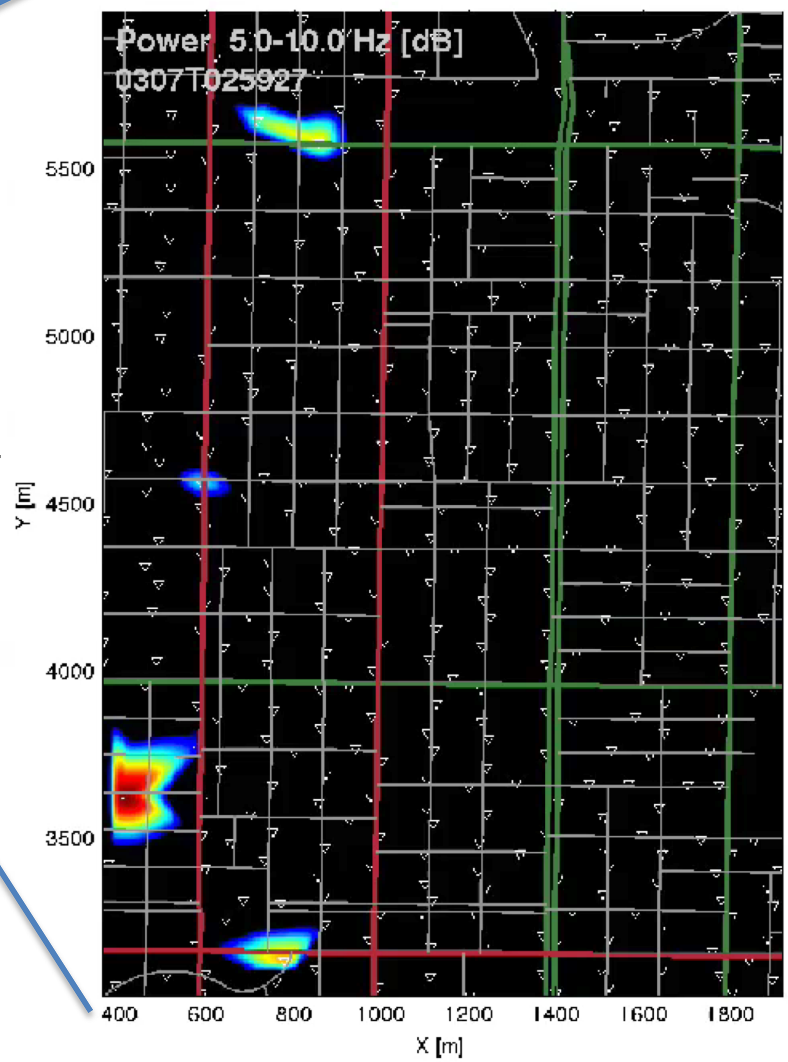


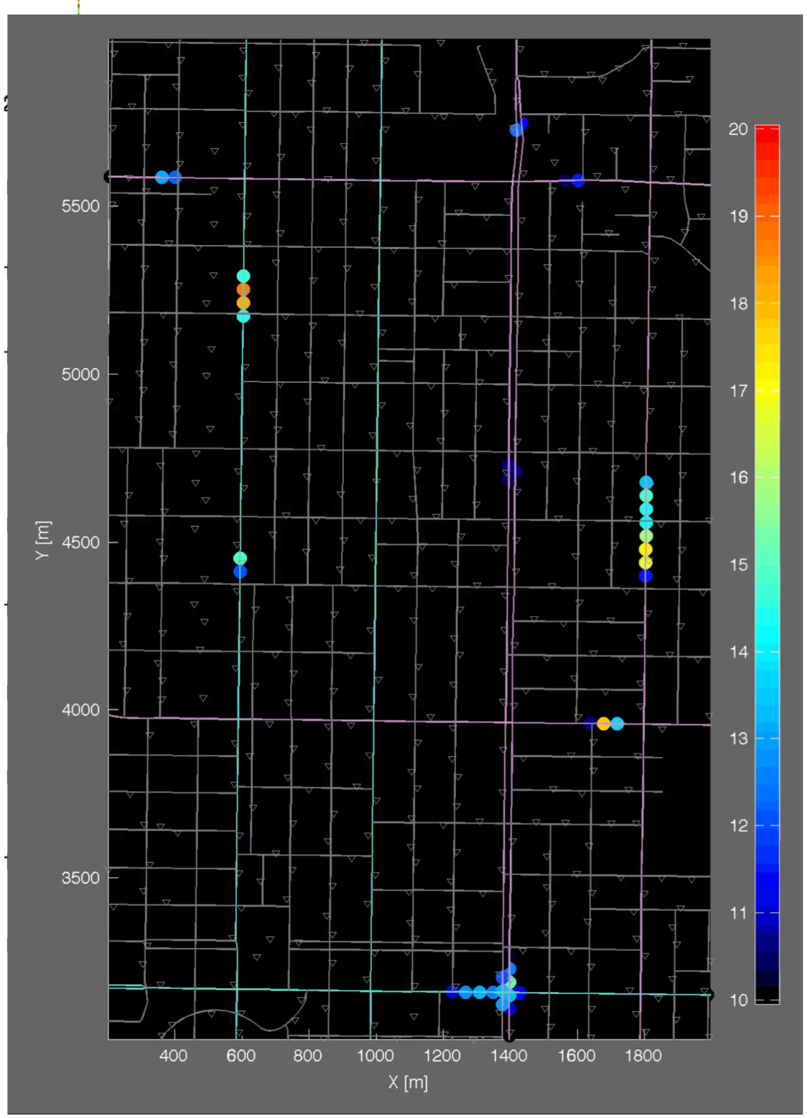
March 5—12, 2011

Noise Tracking of Cars/Trains/Airplanes

5200 element Long Beach array (Dan Hollis)

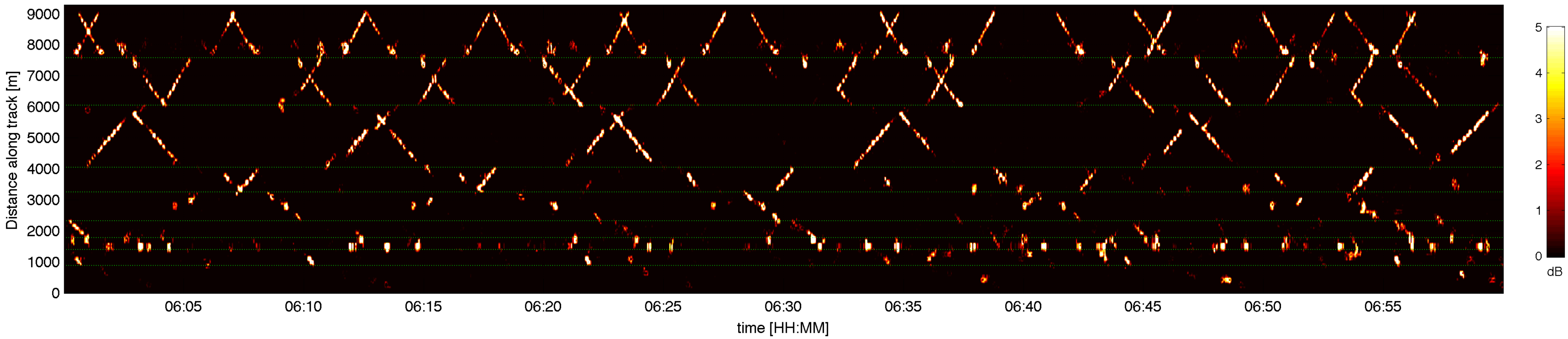


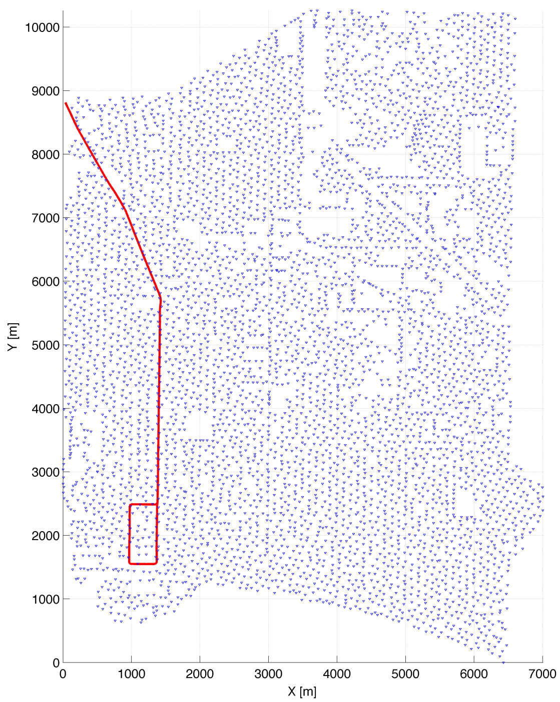




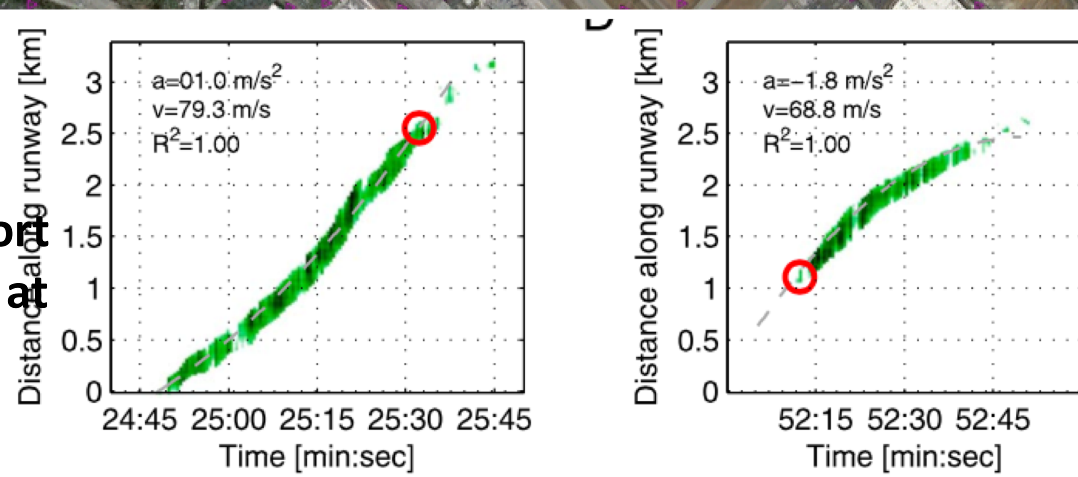
Nima Riahi 2014

Noise Tracking of Cars/Trains/Airplanes





March 7th, 6-7am, rush hour, Blue Line



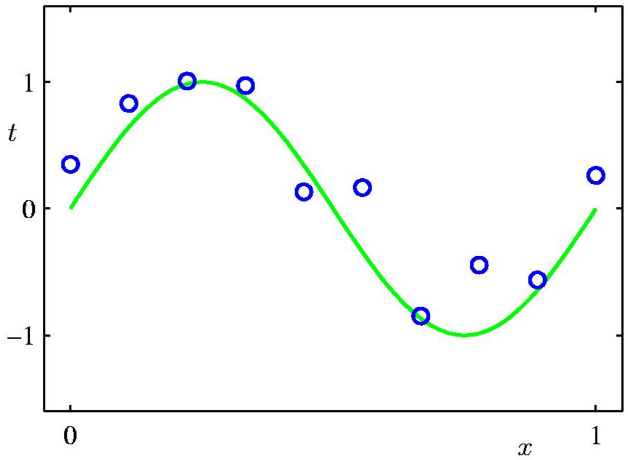
Accelerating airplane on Long Beach Airport runway, moving northwest and taking off at about 120 mi/h.

Riahi, Gerstoft, GRL 2015

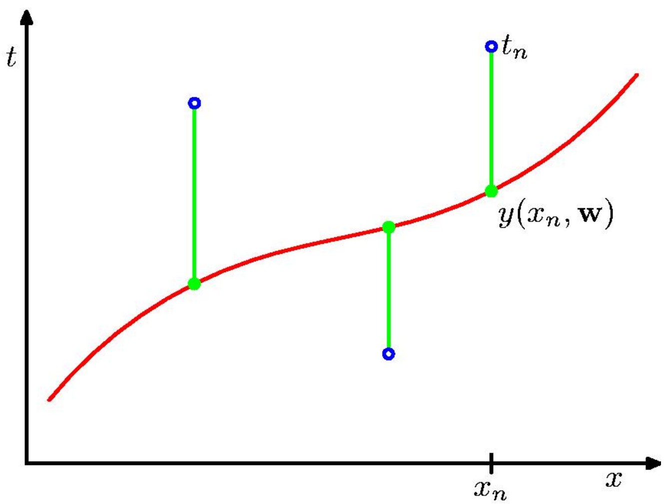
- BISHOP 1.2

Polynomial Curve Fitting

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

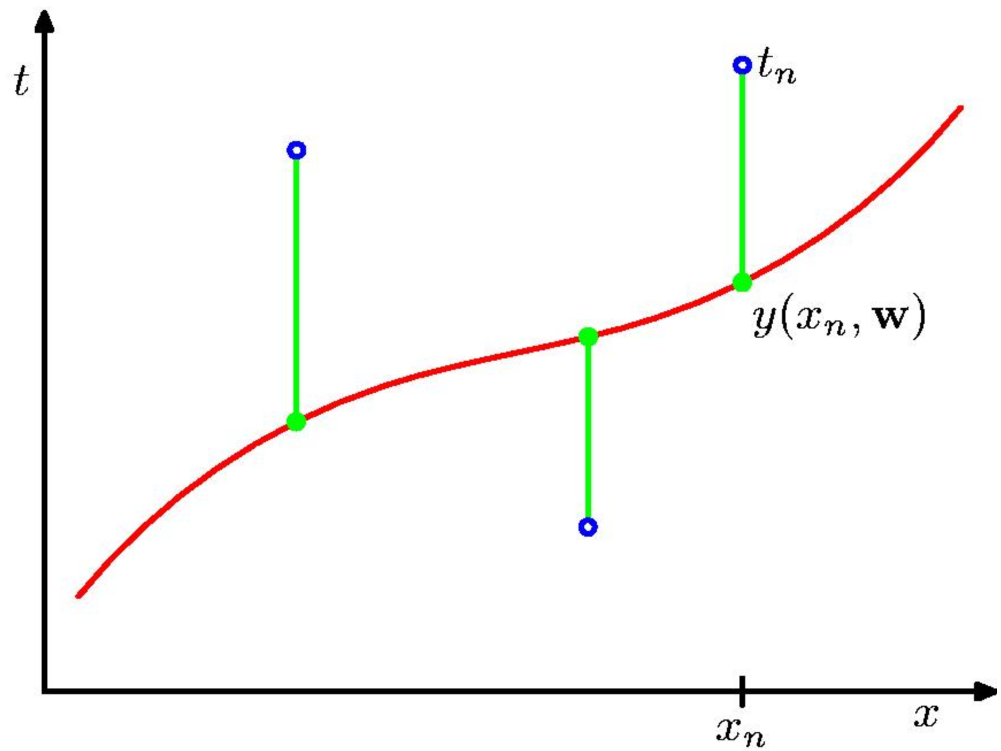


Sum-of-Squares Error Function



$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

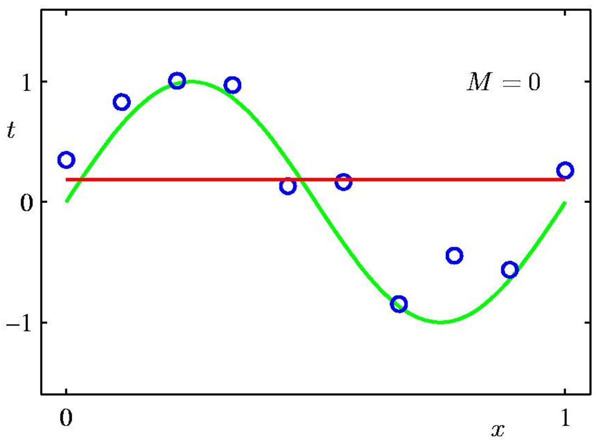
Sum-of-Squares Error Function



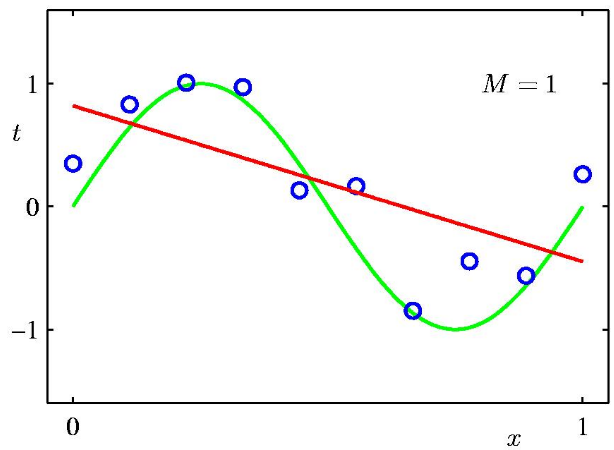
$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

M Order Polynomial Fit

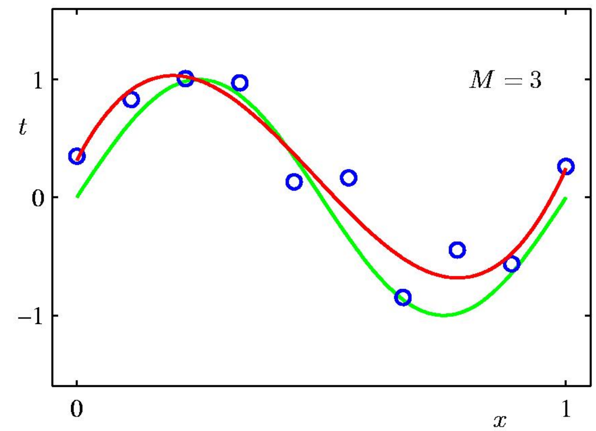
0 Order Polynomial



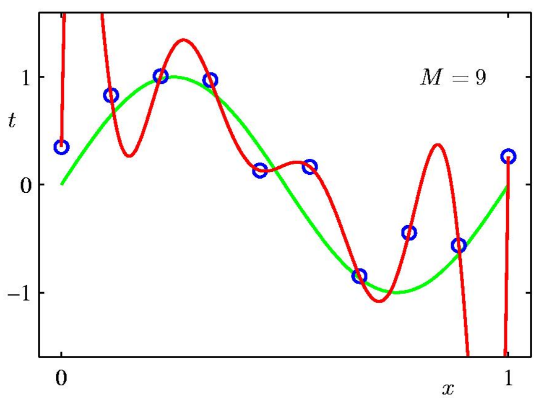
1st Order Polynomial

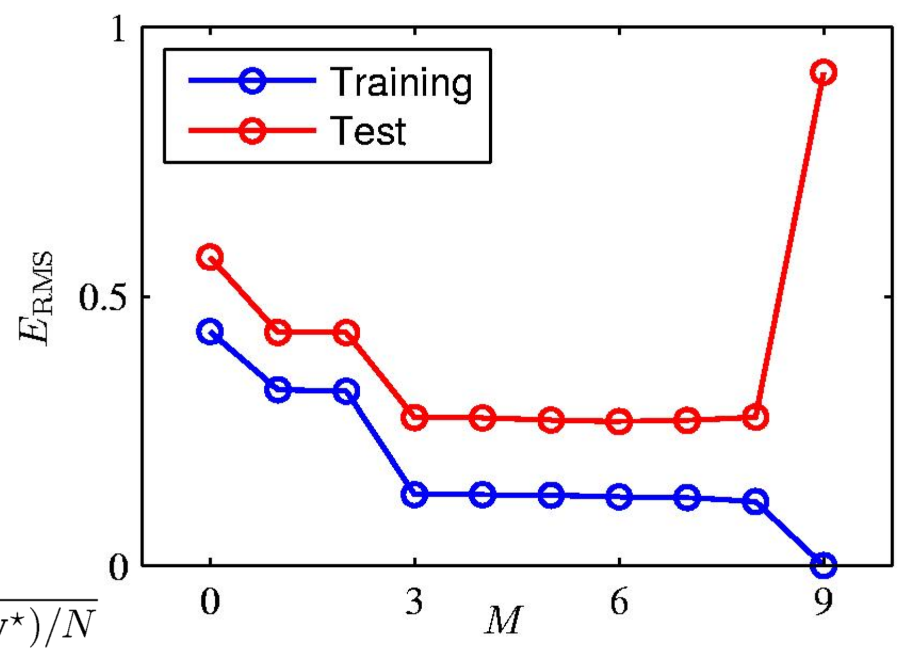


3 Order Polynomial



9 Order Polynomial



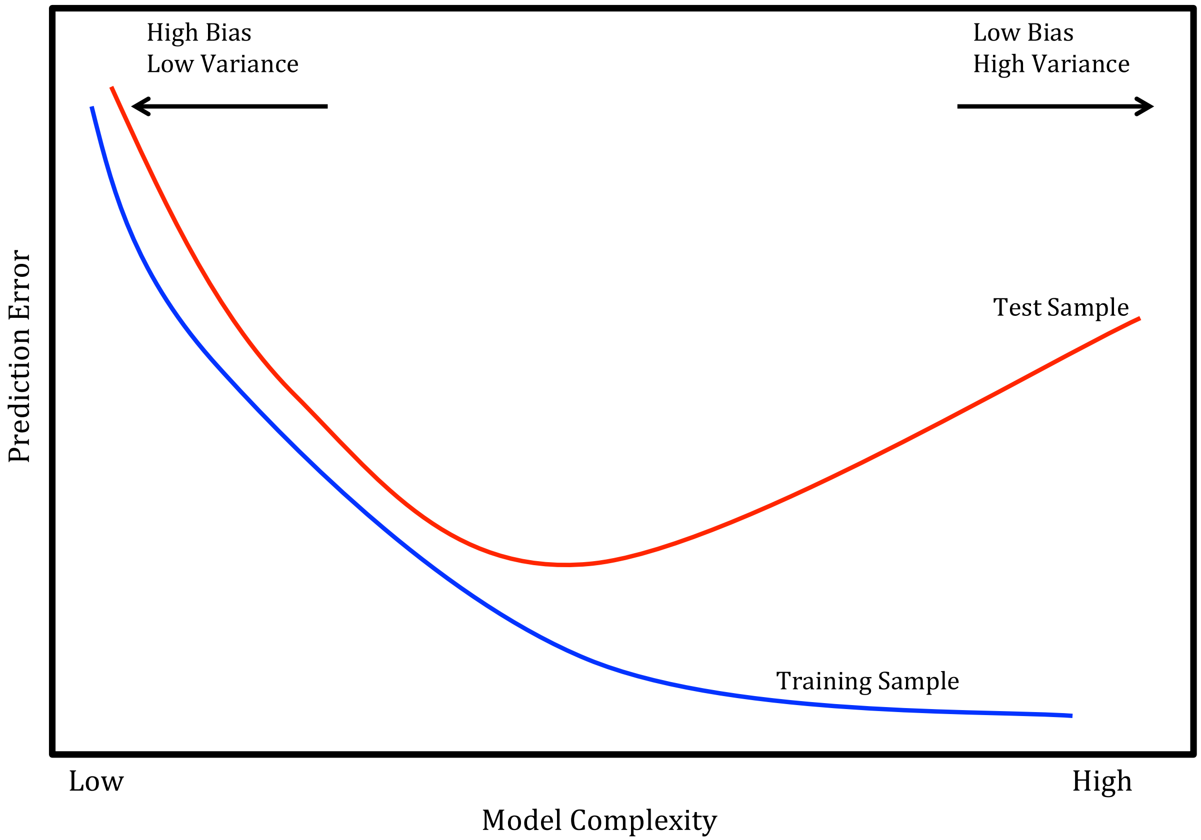


Root-Mean-Square (RMS) Error:

$$E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^{\star})/N}$$

Bias-variance tradeoff

Concept: Complex models can learn data-label relationships well, but may not extrapolate to new cases.



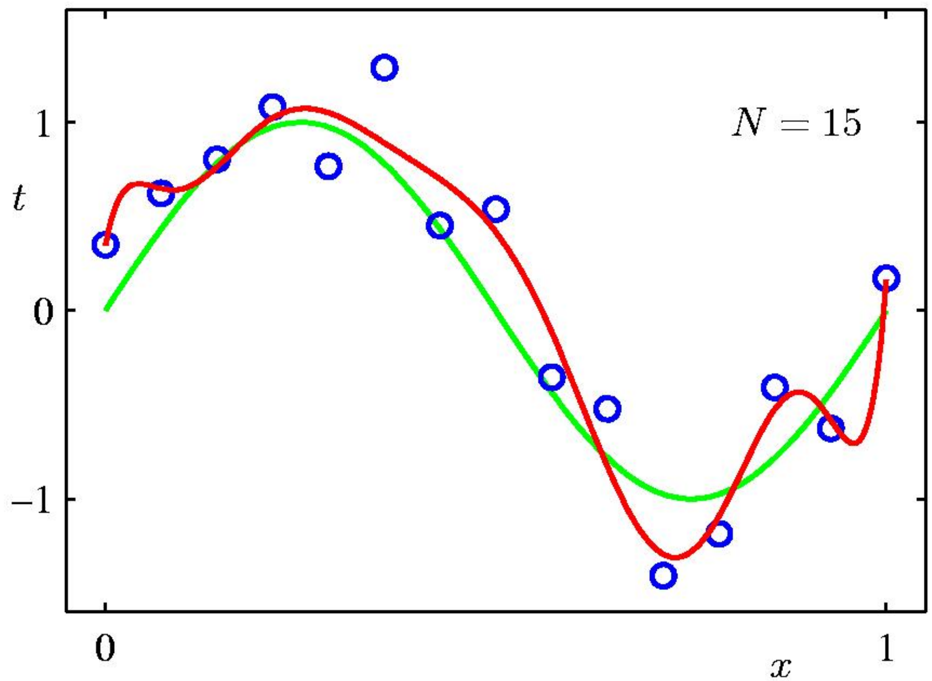
Polynomial Coefficients

	$M=0$	$M=1$	$M=3$	$M=9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

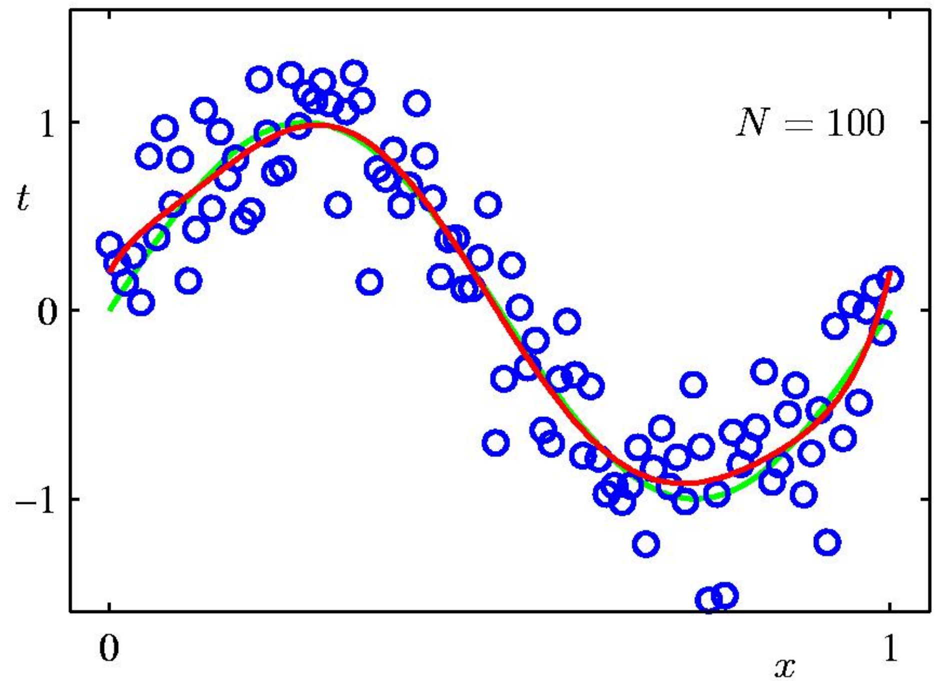
Data Set Size:

9th Order Polynomial

$N = 15$



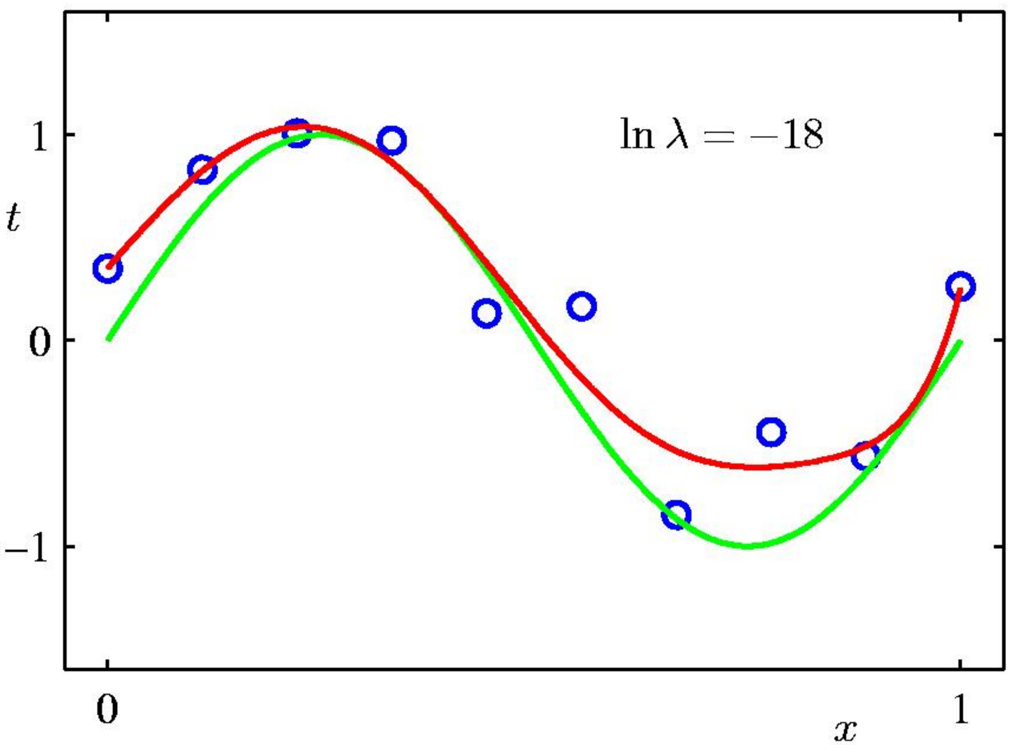
$N = 100$

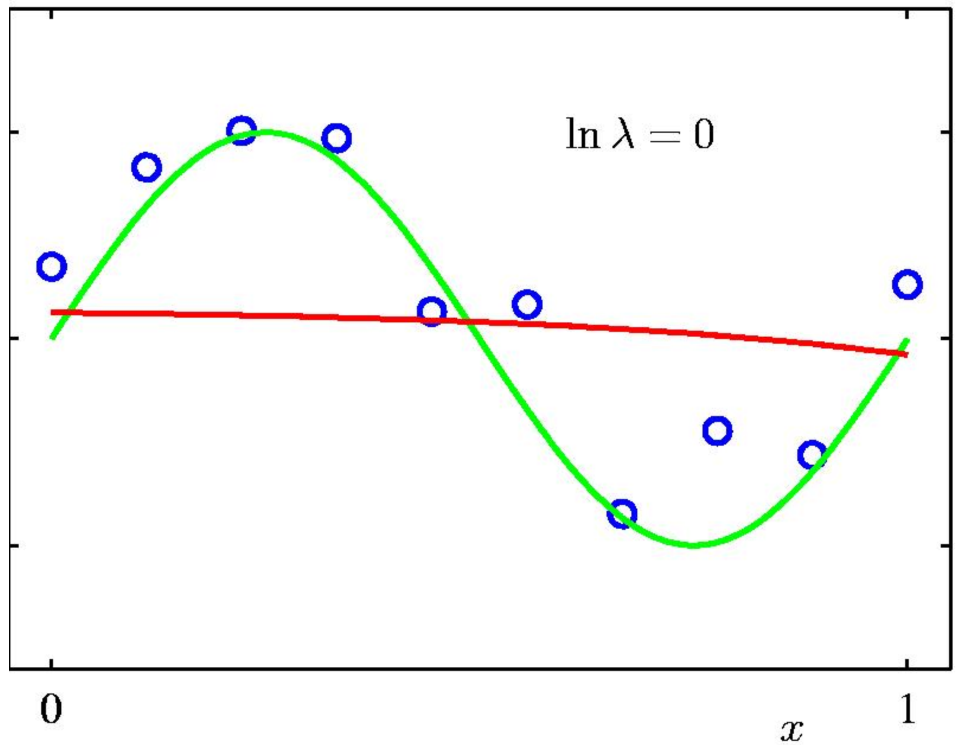


Regularization

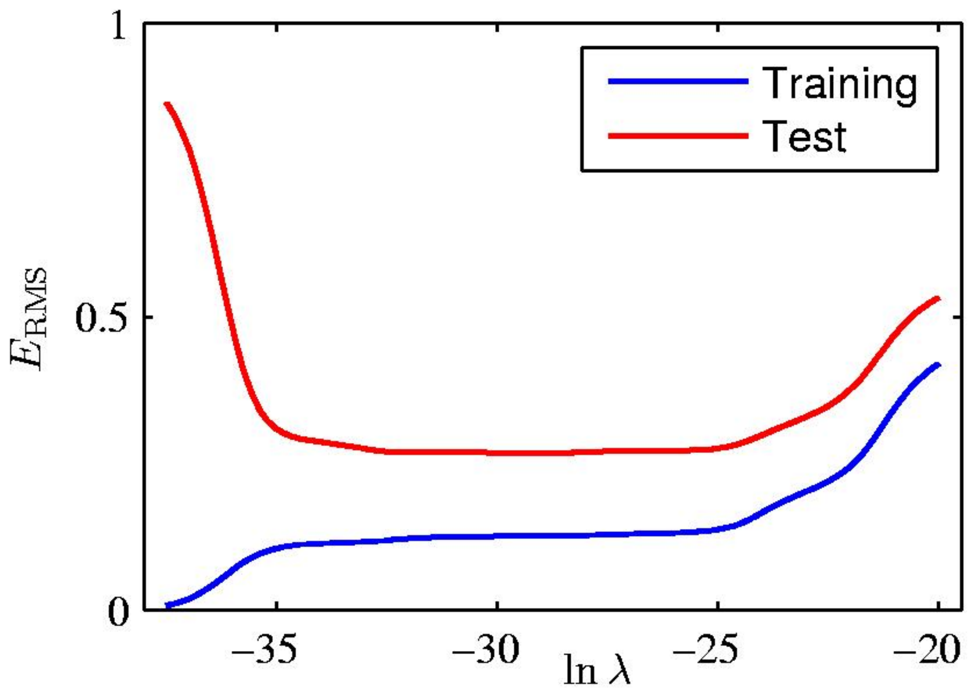
- Penalize large coefficient values

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$





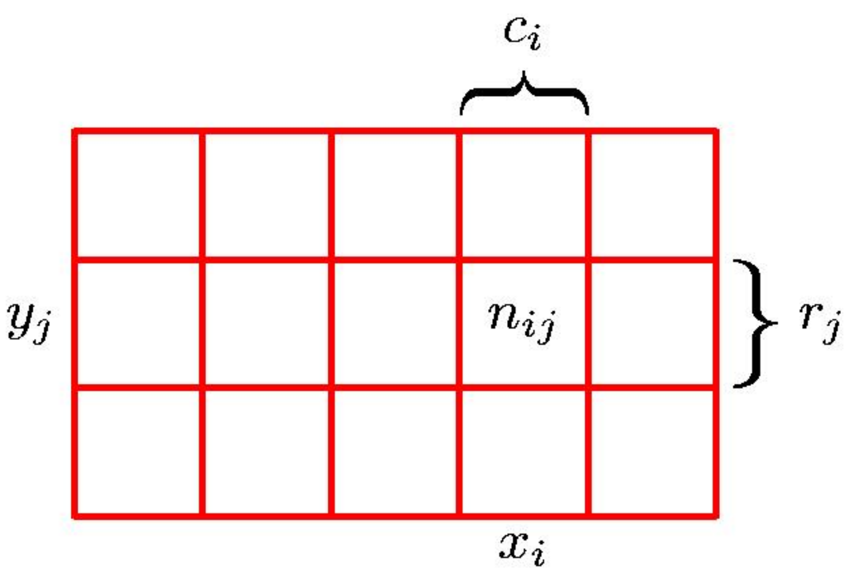
Regularization: E_{RMS} vs. $\ln \lambda$



Polynomial Coefficients

	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01

Probability Theory



Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

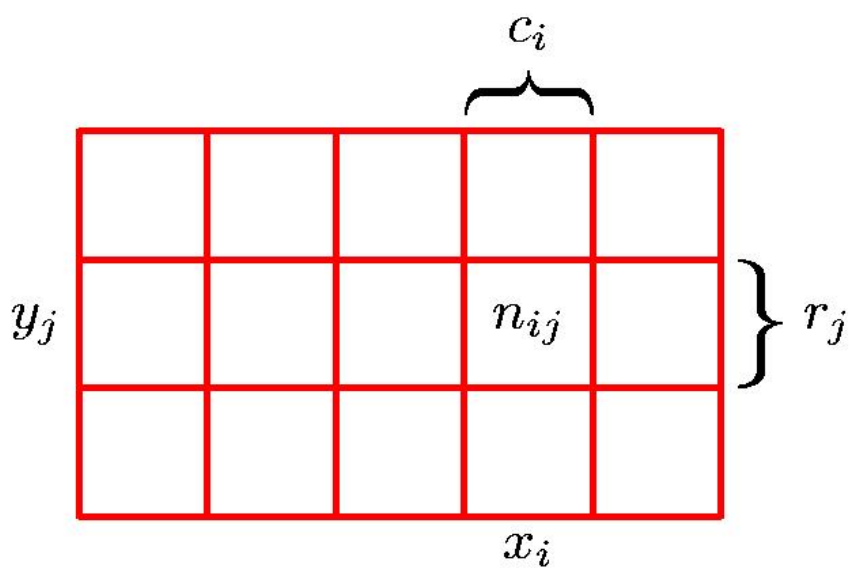
Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory



- Sum Rule

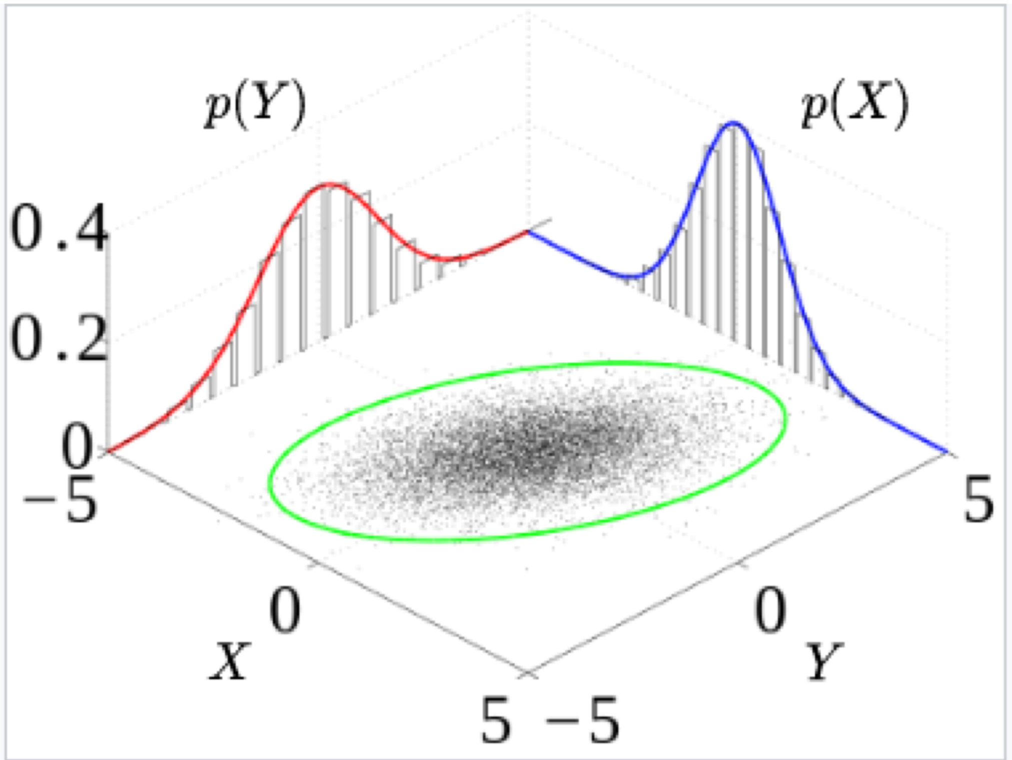
$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

Probability Theory



Joint Probability

Marginal Probability

Conditional Probability

The Rules of Probability

- Sum Rule

$$p(X) = \sum_{Y} p(X, Y)$$

- Product Rule

$$p(X, Y) = p(Y|X)p(X)$$

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior $\propto$ likelihood × prior

Bayes Rule

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{data}|\text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

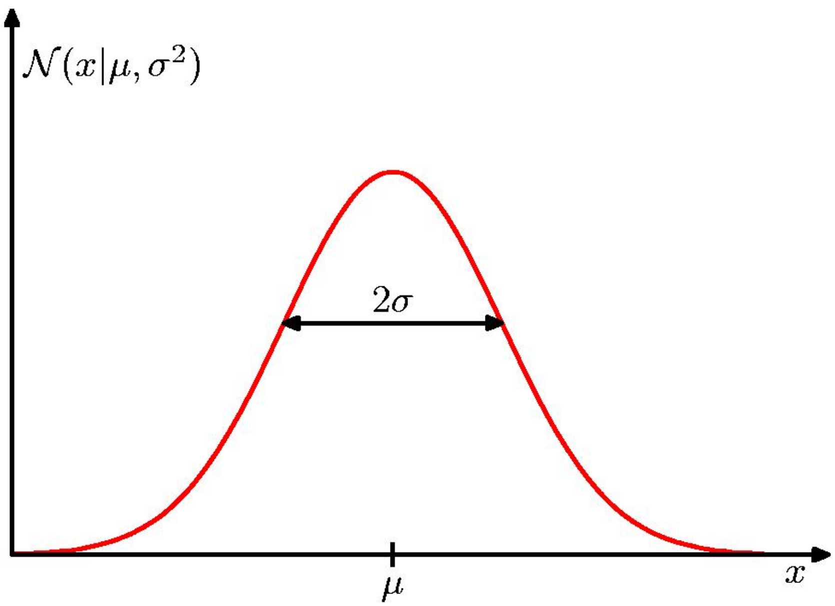
Rev'd Thomas Bayes (1702–1761)

- Bayes rule tells us how to do inference about hypotheses from data.

- Learning and prediction can be seen as forms of inference.

The Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) \,\mathrm{d}x = 1$$

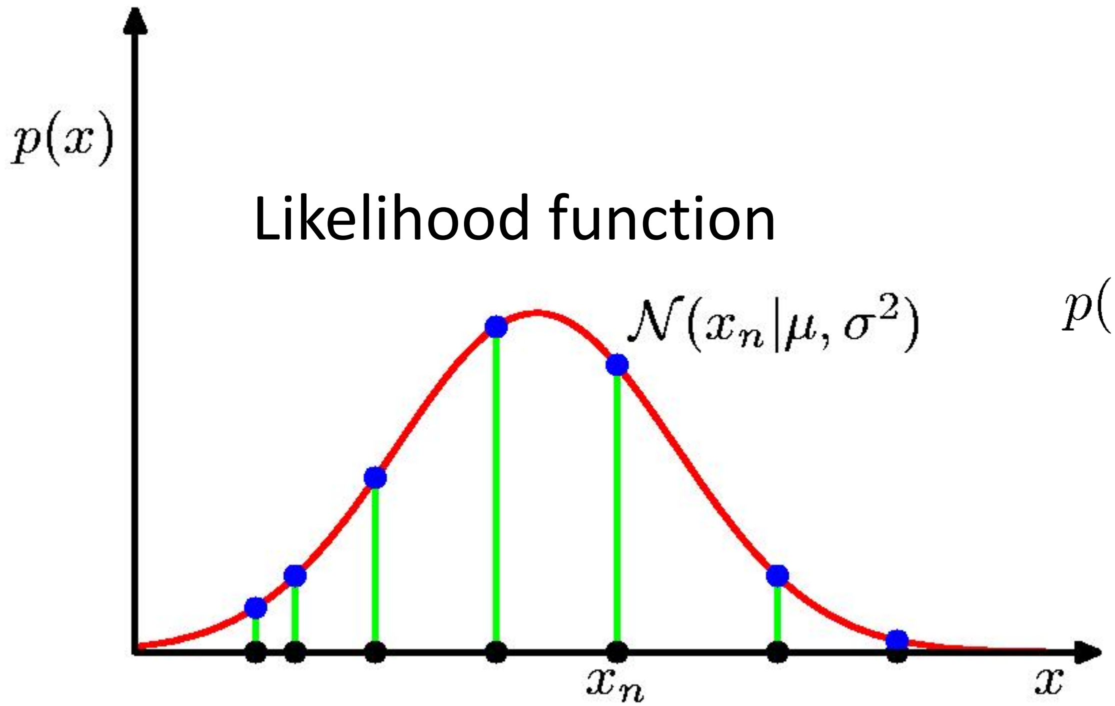
Gaussian Mean and Variance

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \,\mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \,\mathrm{d}x = \mu^2 + \sigma^2$$

$$\operatorname{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

Gaussian Parameter Estimation



$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu,\sigma^2\right)$$

Maximum (Log) Likelihood

$$\ln p\left(\mathbf{x}|\mu,\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad \sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2$$

ML std is biased.

Curve Fitting Re-visited