ECE228 and SIO209 Machine learning for physical applications, Spring 2019
http://noiselab.ucsd.edu/ECE228/

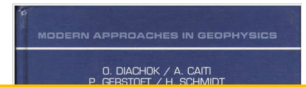
Professor Peter Gerstoft, Spiess Hall 462, Gerstoft@ucsd.edu
TA Siva Prasad Varma Chiluvuri, sivapvarma@gmail.com
TA Harshuk Gupta, h6gupta@eng.ucsd.edu
TA Ruixian Liu, rul188@eng.ucsd.edu

Location: SOLIS 107
Time: Monday and Wednesday 5-6:20pm

We are focused on basic ML methods and their application.

ECE285 Machine Learning for Image Processing is focused on NN

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:

- 170 Papers, H-factor 49 (Scholar).
- 105 Ocean Acoustics, 19 EM, 44 seismics, 45 SP
- Mentoring a diversified (culture, levels, science interest, scien
  person acoustics group.
- Funding ONR, NSF GEO & Polar, DOE, visitors.

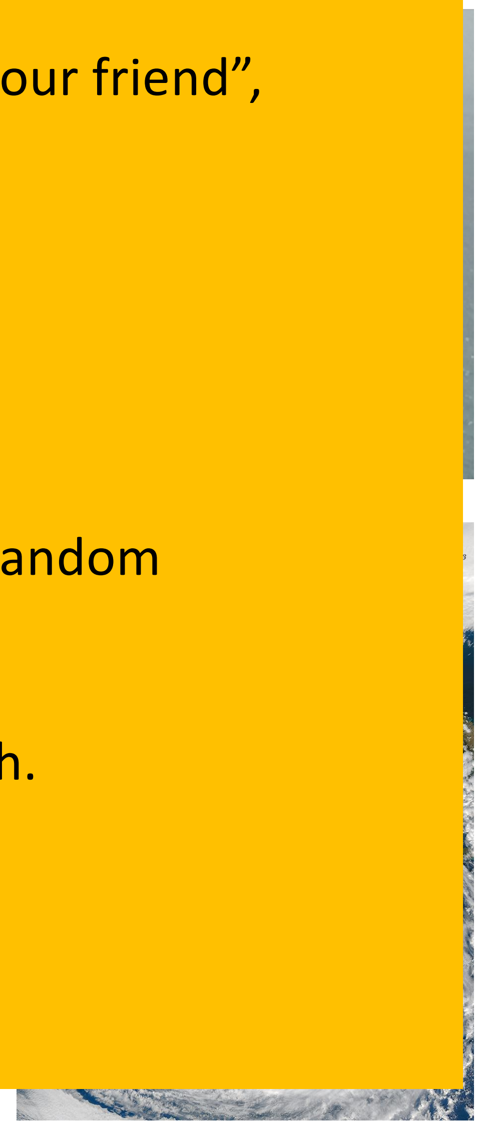
Deterministic,
non-random,
first principles,
stochastic search GA

Random, “Chaos is our friend”,
first principles

first principles
random

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

Always Bayes



2019: 224Students with the following specialization

166 EC, 3 BE, 1 BI, 1 CE, 3 CH, 19 CS, 1 CU, IIR, 9 MC, 1 MA, 1 Na, 2 RS, 5 SE 6 SI 1 PY, 1 UN

2018: 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

Grade 2017: (A+ 19, A 20, A- 13, B+ 7, S 1, W 1)

2018: (A+ 21, A 20, A- 20, B+ 4, B 5)

- 50% Homework, automatic graded
- 50% Project
- 5 class participation

TA (**Siva** Prasad Varma Chiluvuri, **Harshuk** Gupta, **Ruixian** Liu)

- Siva coordinate/lead home work (presentation and Cody)
- Harshuk coordinate/lead Piazza, Jupyter, GPU effort
- Ruixian coordinate projects, present ML to discover PDE
- Office hours on Piazza ECE/SIO, just TA?

Ideal Class 80 min
10 min homework
40 min pre or post homework science.
30 min applications, projects **D2 students please give a presentation instead of projects.**
Light theory initially
Partly reverse class. Stanford
https://www.youtube.com/playlist?list=PL3FW7Lu3i5JvHM8ljYj-zLfQRF3EO8sYv

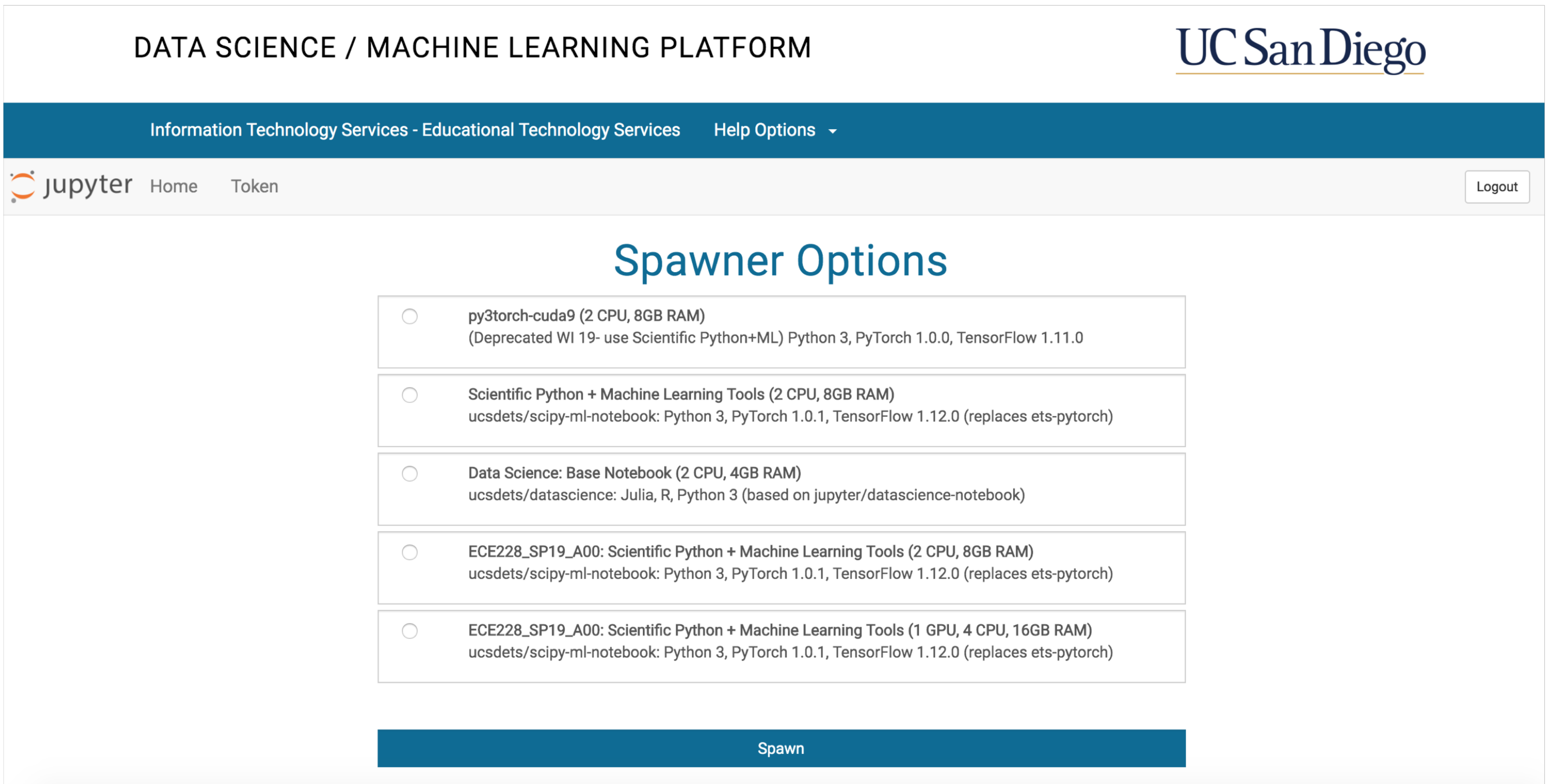
Homework

- 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

GPU datahub.ucsd.edu

https://datahub.ucsd.edu/hub/login
[Documentation]()

TA Harshul
1-2 Homeworks on this
Plus Final project
Tensorflow gave a factor 10 speedup



Projects

- **3-4** person groups
- Deliverables: Poster & Report & main code (plus proposal, midterm slide)
- **Topics** your own or chose form suggested topics
- **Week 4 groups** due to TA Ruixian (if you don’t have a group, ask in week 3 and we can help).
- **May 5** proposal due. TAs and Peter can approve.
- Proposal: One page: Title, A large paragraph, data, weblinks, references.
- Something physical
- **May 20** Midterm slide presentation. Presented to a subgroup of class.
- **June 5** final poster. Uploaded June 3
- Report and code due **Saturday 15 June.**

Final Projects 2018

Group	Topic	Authors	Poster	Report
1	**Reimplementation of source localization in an ocean waveguide using supervised learning**	Jinzhao Feng, Zhuoxi Zeng, Yu Zhang	Poster	Paper
2	**Machine learning methods for ship detection in satelite images**	Yifan Li, Huadong Zhang, Xiaoshi Li, Quianfeng Guo	Poster	Paper
3	**Transparent Conductor Prediction**	Yan Sun, Yiyuan Xing, Xufan Xiong, Tianduo Hao	Poster	Paper
4	**Ship identification in sateklite Images**	Weilun Zhang, Zhaoliang Zheng, Mingchen Mao,	Poster	Paper
5	**Fruit Recognition**	Eskil Jarslkog, Richard Wang, Joel Andersson	Poster	Paper
6	**RSNA Bone Age Prediction**	Juan Camilo Castillo, Yitian Tong, Jiyang Zhao, Fengcan Zhu	Poster	Paper
7	**Facial Expression Classification into Emotions**	David Orozco, Christopher Lee, Yevgeniy Arabadzhi, Deval Gupta	Poster	Paper
8	**Urban Scene Segmentation for Autonomous Vehicles**	Hsiao-Chen Huang, Eddie Tseng, Ping-Chun Chiang, Chih-Yen Lin	Poster	Paper
9	**Face Detection Using Deep Learning**	Yu Shen, Kuan-Wei Chen, Yizhou Hao, Min Hsuan Wu	Poster	Paper
10	**Understanding the Amazon Rainforest using Neural Networks**	Naveen Dharshana Ketagoda, Christian Jonathan Koguchi, Niral Lalit Pathak, Samuel Sunarjo	Poster	Paper
11	**Mercedes-Benz Bench Test Time Estimation**	Lanjihong Ma, Kexiong Wu, Bo Xiao, Zihang Yu	Poster	Paper
12	**Vegetation Classification in Hyperspectral Image**	Osman Cihan Kilinc, Kazim Ergun, Yuming Qiao, Fengjunyan Li	Poster	Paper
13	**Threat Detection Using AlexNet on TSA scans**	Amartya Bhattacharyya, Christine H Lind, Rahul Shirpurkar	Poster	Paper
14	**Flagellates Classification via Transfer Learning**	Eric Ho, Brian Henriquez, Jeffrey Yeung	Poster	Paper
15	**Biomedical Image Segmentation**	Lucas Tindall, Amir Persekian, Max Jiao	Poster	Paper
16	**"Deep Fakes" using Generative Adversarial Networks (GAN)**	Tianxiang Shen, Ruixian Liu, Ju Bai, Zheng Li	Poster	Paper
17	**Dog Breed Classification via Convolutional Neural Network**	Yizhou Chen; Xiaotong Chen; Xuanzhen Xu	Poster	Paper
18	**Dog Breed Identification**	Wenting Shi, Jiaquan Chen, Fangyu Liu, Muyun Liu	Poster	Paper
19	**Impact of Skewed Distributions on an Automated Plankton Classifier**	Will Chapman, Emal Fatima, William Jenkins, Steven Tien, Shawheen Tosifian	Poster	Paper
20	**Blood Cell Detection using Single shot MultiBox Detector**	Inyoung Huh	Poster	Paper

2017 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

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

RESEARCH ARTICLE

EARTH SCIENCES

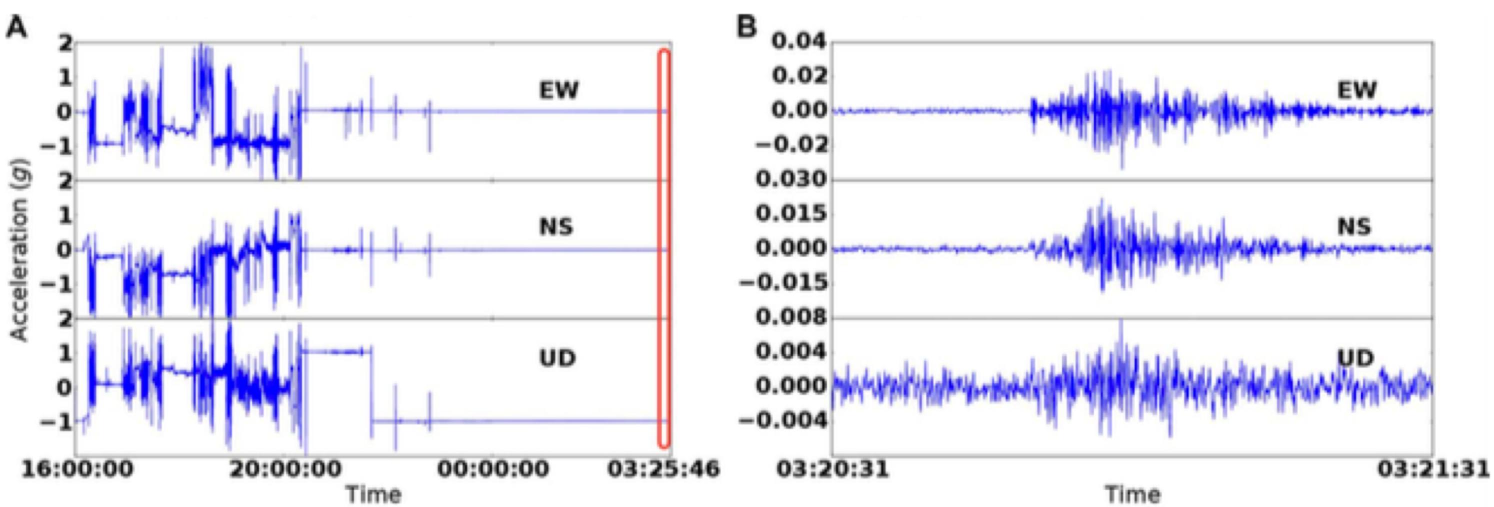
MyShake: A smartphone seismic network for earthquake early warning and beyond

10.1126/sciadv.1501055

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

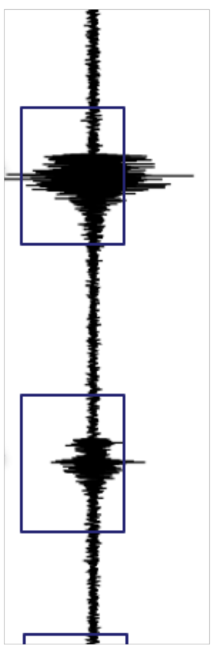
Large magnitude earthquakes in urban environments continue to kill and injure tens to hundreds of thousands of people, inflicting lasting societal and economic disasters. Earthquake early warning (EEW) provides seconds to minutes of warning, allowing people to move to safe zones and automated slowdown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

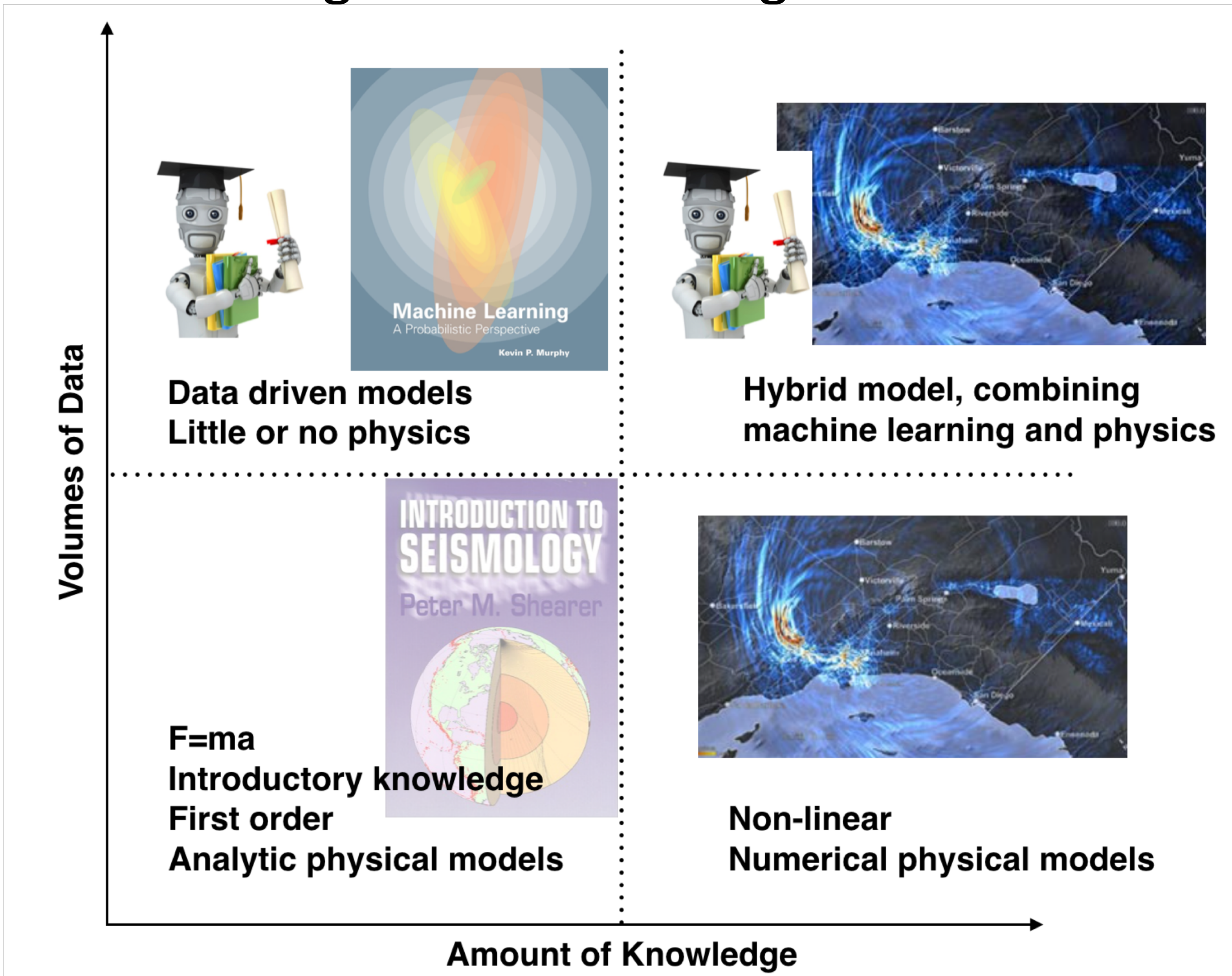
RESEARCH ARTICLE



	First principles vs	**Data driven**
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** **Signal-Proc**	**Peter** **Google**

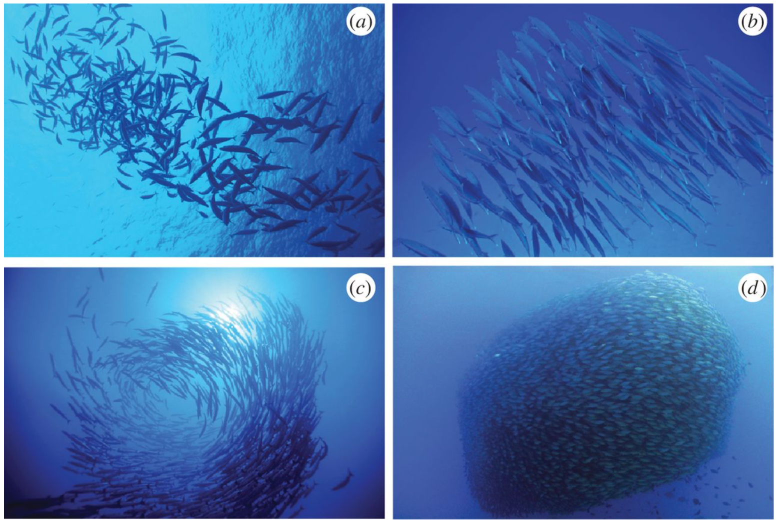
Machine learning versus knowledge based



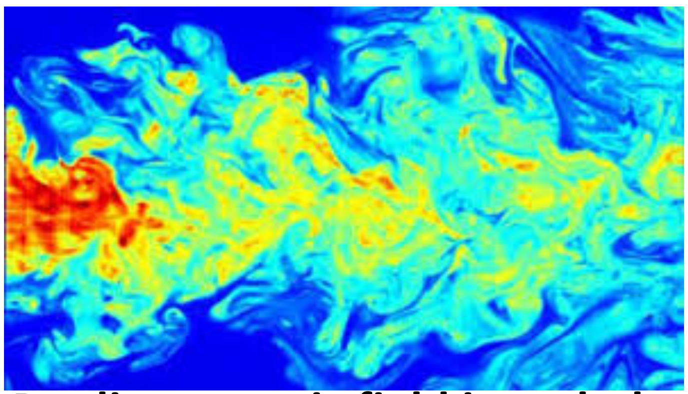


3D spectral elements

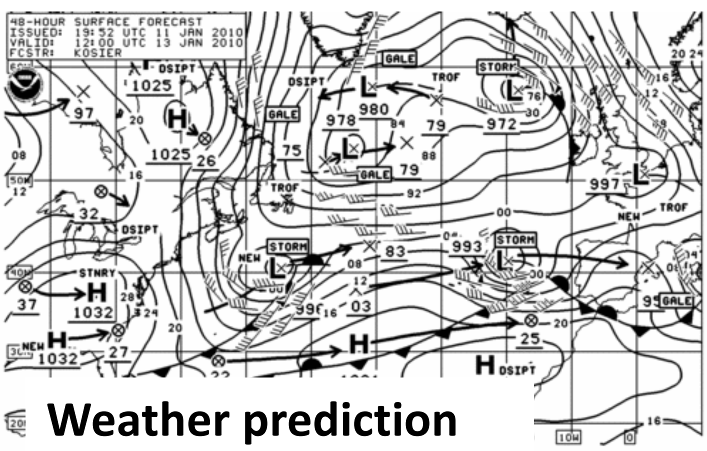
We can’t model everything…



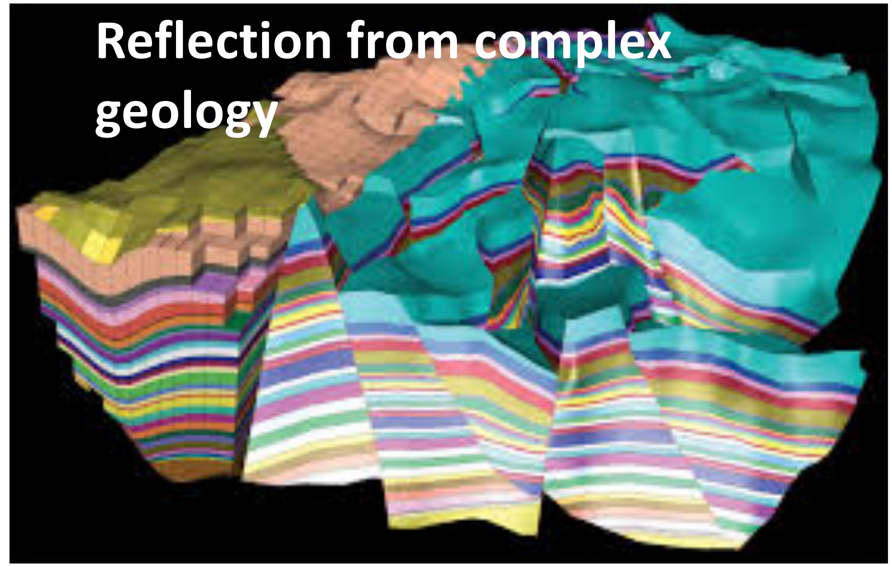
Back scattering from fish school



Predict acoustic field in turbulence



Weather prediction



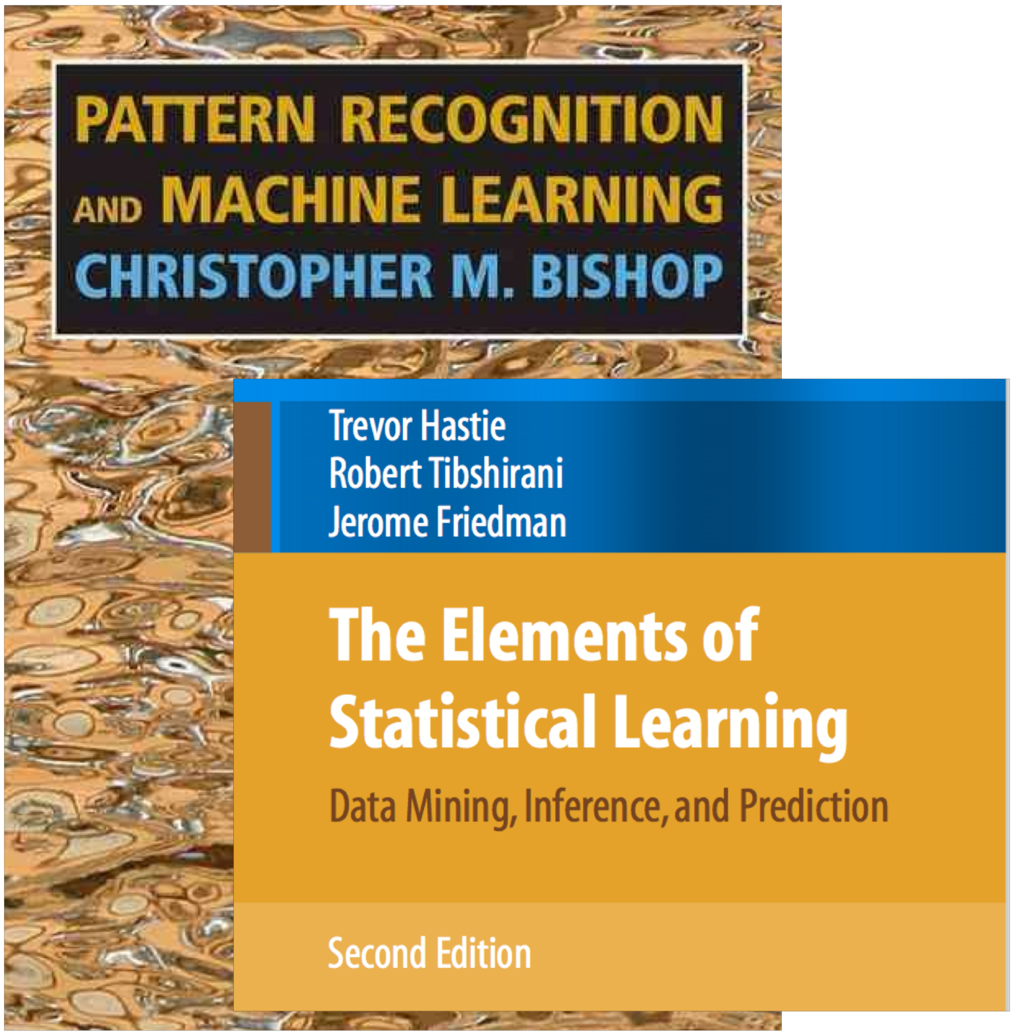
Reflection from complex geology

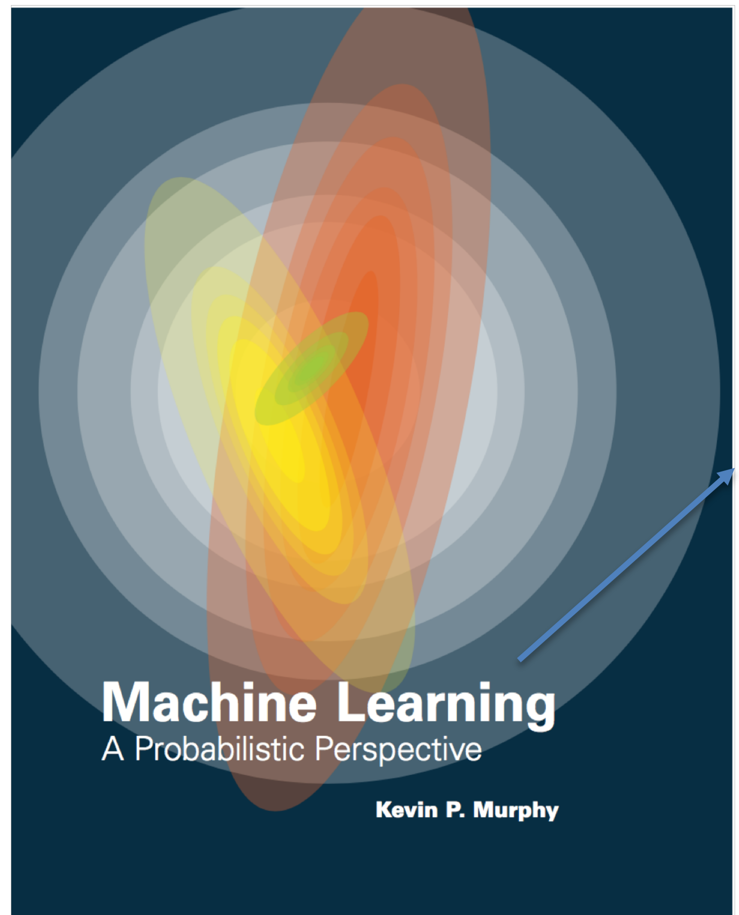
Detection of mines. Navy uses dolphins to assist in this.

Dolphins = real ML!

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

Machine learning in Physical Sciences

Peter Gerstoft, Mike Bianco, Emma Ozanich, Haiqiang Niu

http://noiselab.ucsd.edu/. SIO, UCSD

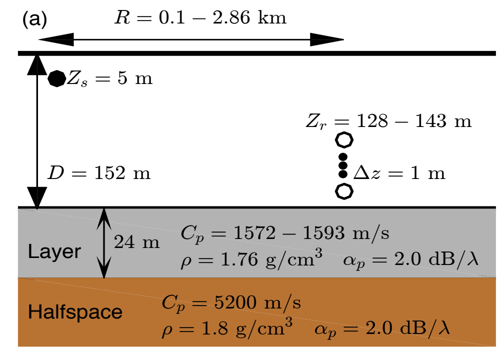
Summary

- Machine learning, big data, data science, artificial intelligence are about the same.
- Data science has lots of **opportunities** in physics.
- **Neural networks** is one method. Similar are methods are **Support Vector Machines** (SVM) and **Random Forest** (RF). Use the latter for a first implementation.
- **Unsupervised learning** is more challenging than **supervised learning**
- Coding: Matlab OK, Jupyter notebook is nice**.**
- I like **graph signal processing** methods, **dictionary learning, sequential estimation**
- **Following the trend, here we use RF, SVM, FNN, CNN, LSTM, ResNet**

Relevant papers ML in ocean acoustics: (FNN)
Niu, Reeves, Gerstoft (2017) JASA **142**. (Noise09)
Niu, Ozanich, Gerstoft (2017) JASA-EL **142**. (SBC)
Ozanich, Niu Gerstoft (2019?) JASA
Niu, Ozanich, Gerstoft (2019?) JASA.
Michalopoulou, Gerstoft (2019), JOE in press.
Bianco 2019? **Review paper**

ML in seismics
Riahi 2017 (**Graph processing**)
Bianco 2017, 2018,2019? (**Tomography/ Dictionary Learning**)
Kong 2019 **Review paper**

Matched-Field Processing on test data 1

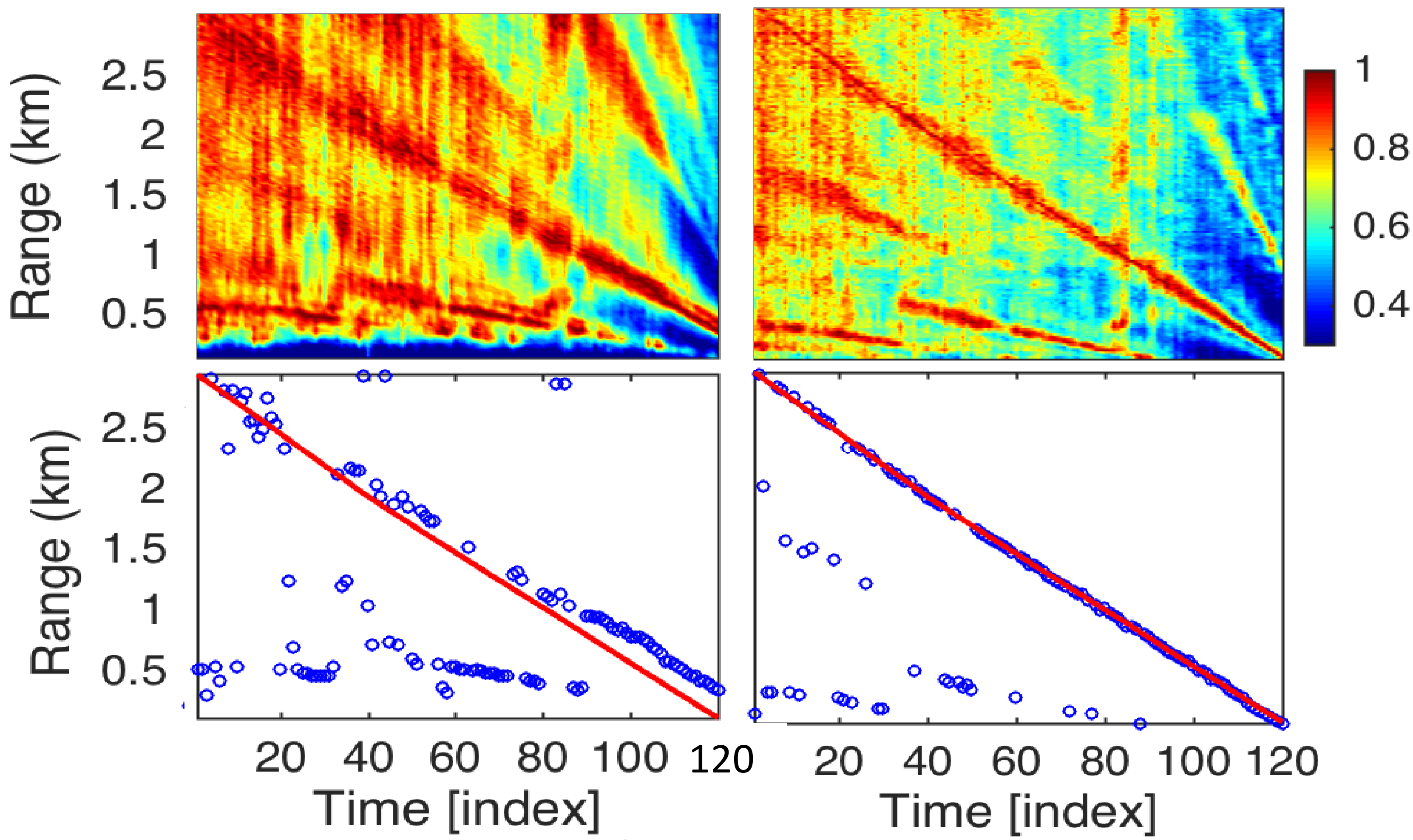


Frequencies [300:10:950]Hz

$$E_{\mathrm{MAPE}} = \frac{100}{N} \sum_{i=1}^{N} \left| \frac{Rp_i - Rg_i}{Rg_i} \right|$$

$$B = \mathbf{p}^H \mathbf{C} \mathbf{p}$$

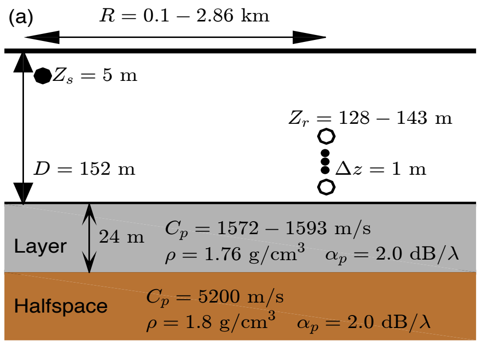
synthetic replicas. measured replicas

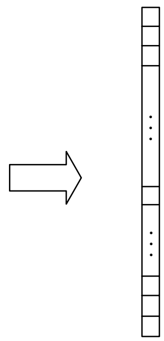


Mean Absolute Percentage Error error of MFPs: ***55%*** and ***19%***

Classification versus regression

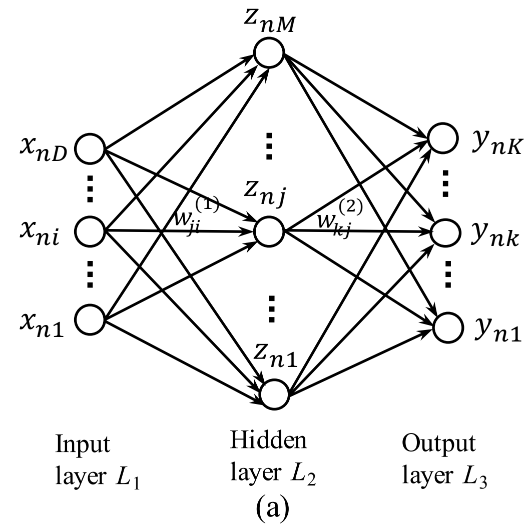
Classification:



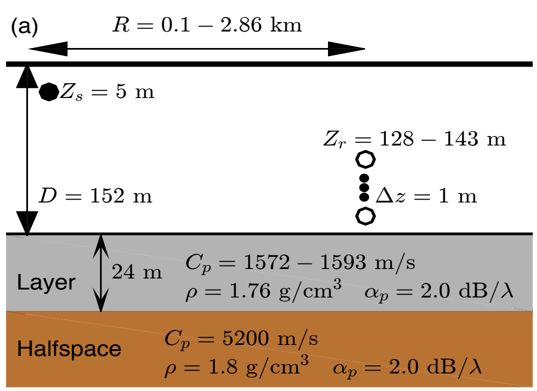


N potential source ranges

$$R = \{r_1, \dots, r_N\}$$

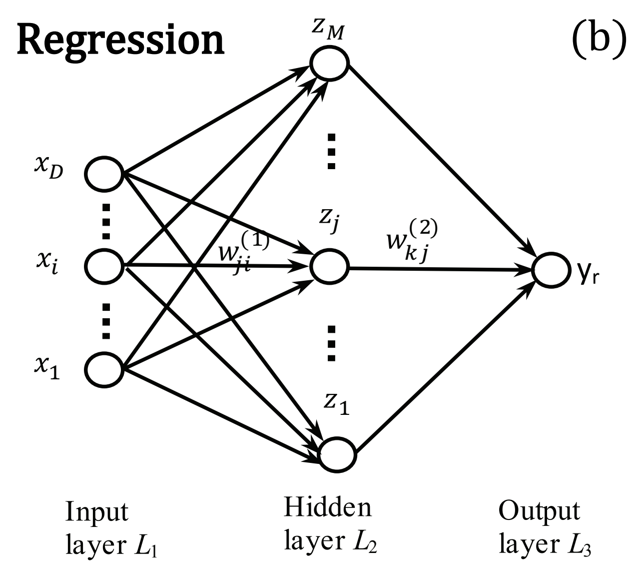


Regression:



one source continuous range

Regression is harder



Number of parameters

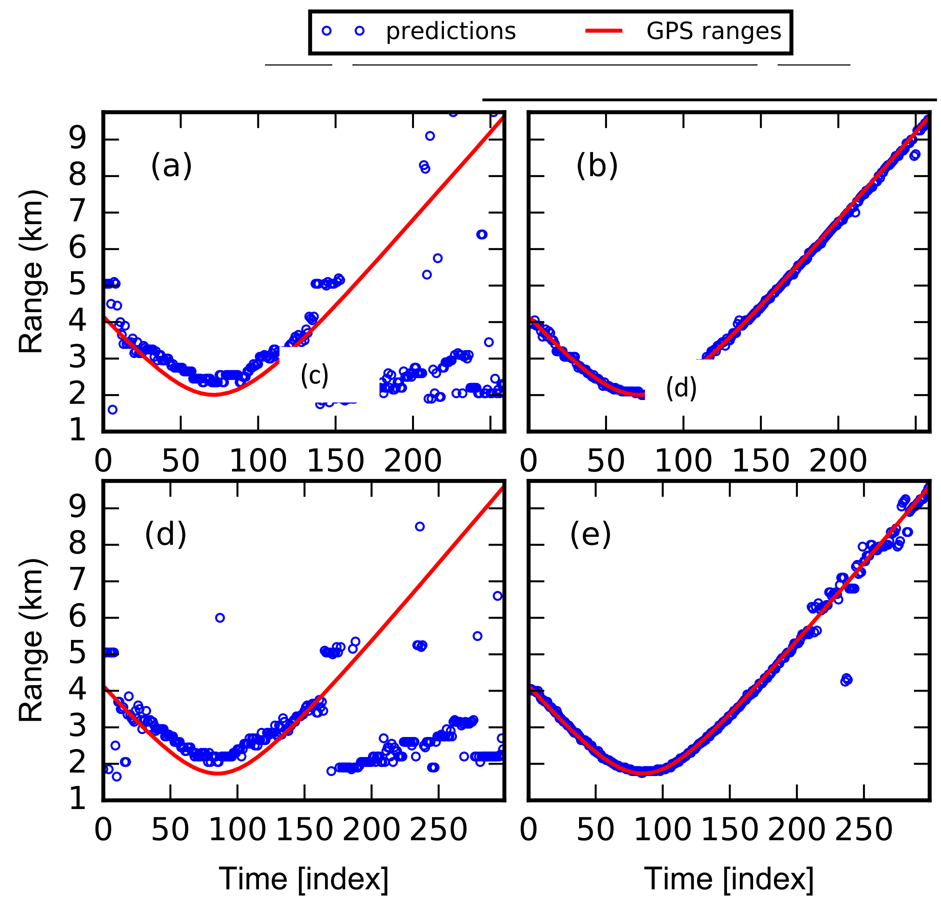
MFP: O(10)

ML: 400*1000+ 1000*1000+1000*100

= O(1000000)

So far…

- Can machine learning learn a nonlinear noise-range relationship?
  - **Yes:** *Niu et al. 2017, "Source localization in an ocean waveguide using machine learning."*
- We can use different ships for training and testing ?
  - **Yes**: *Niu et a. 2017, "Ship localization in Santa Barbara Channel using machine learning classifiers."* (see figure)



Ship range localization using (a,c) MFP and (b,d) Support Vector Machine (rbf kernel).

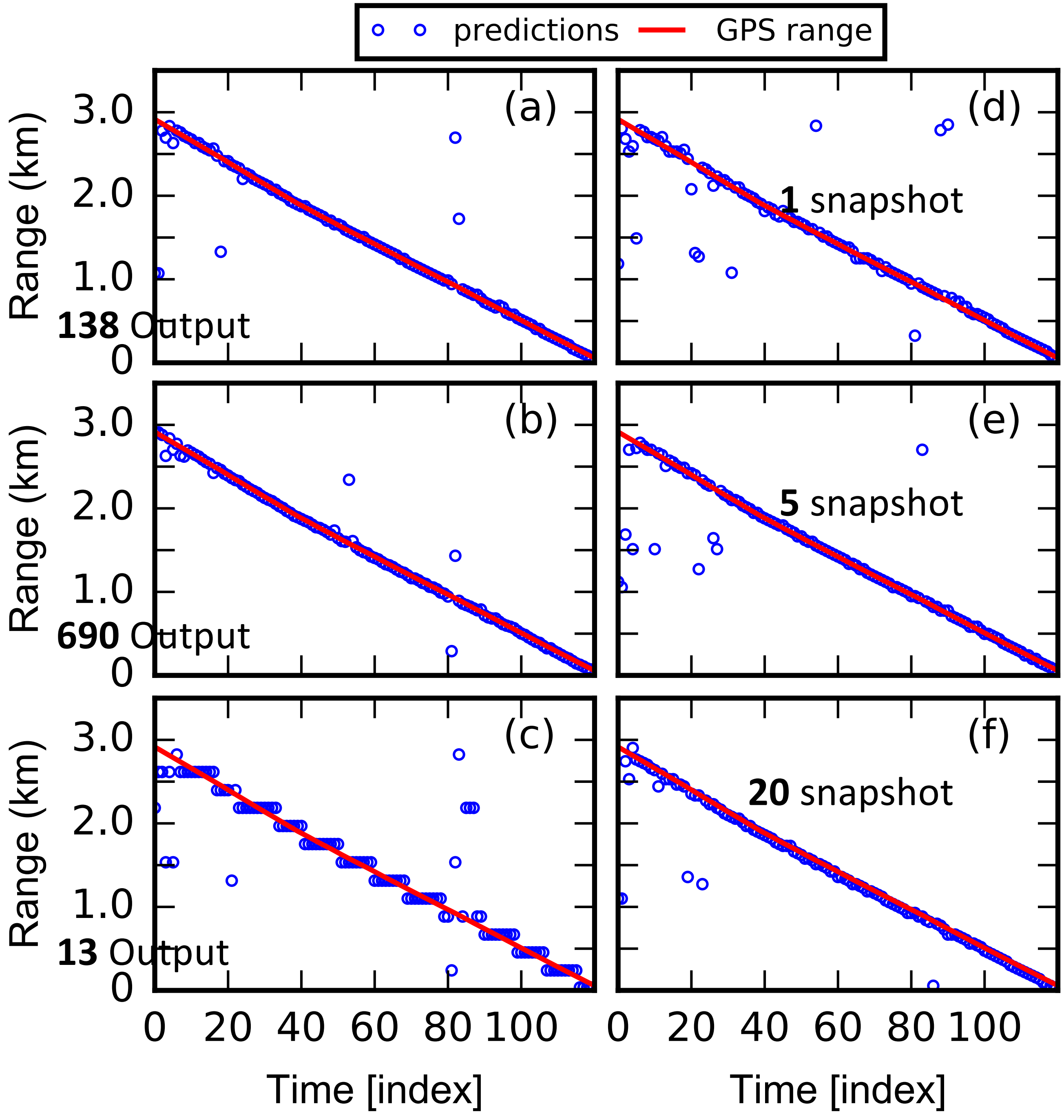
NN, SVM, and random forest
Perform about similar

60s Science
Scientic Am

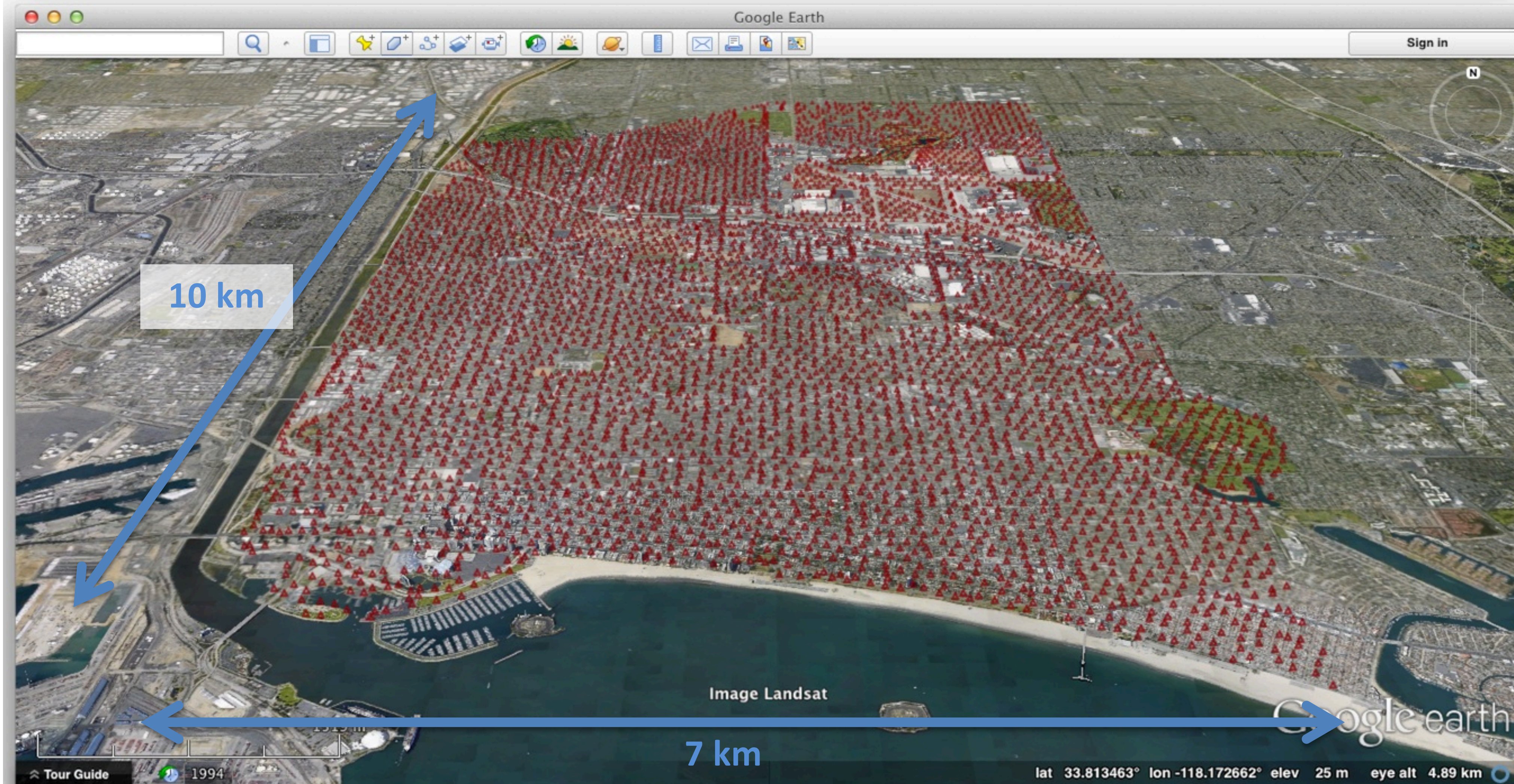
Other parameters: FNN

Conclusion

- Works better than MFP
- Classification better than regression
- FNN, SVM, RF works.
- Works for:
    - multiple ships,
  - Deep/shallow water
  - Azimuth from VLA



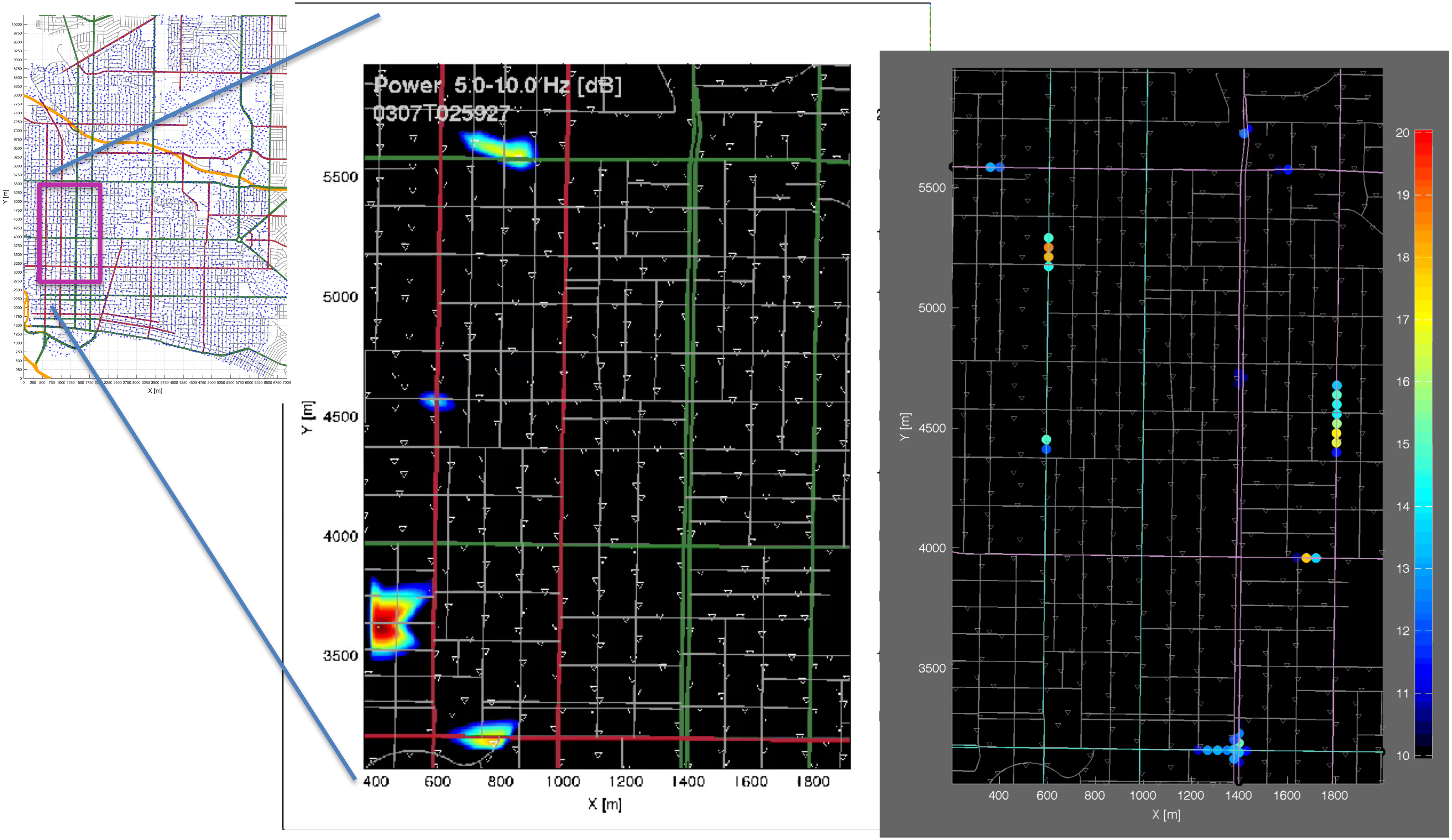
Why we got interested in traffic



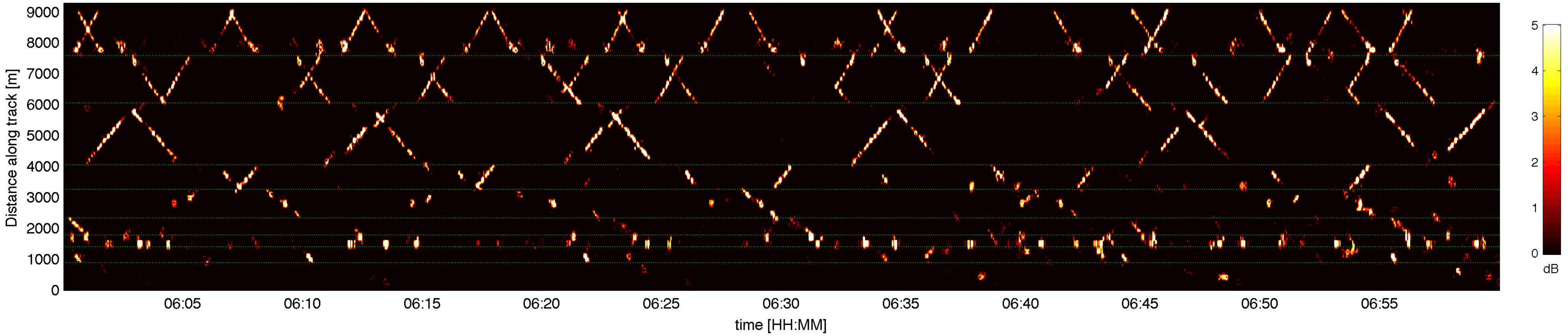
March 5—12, 2011

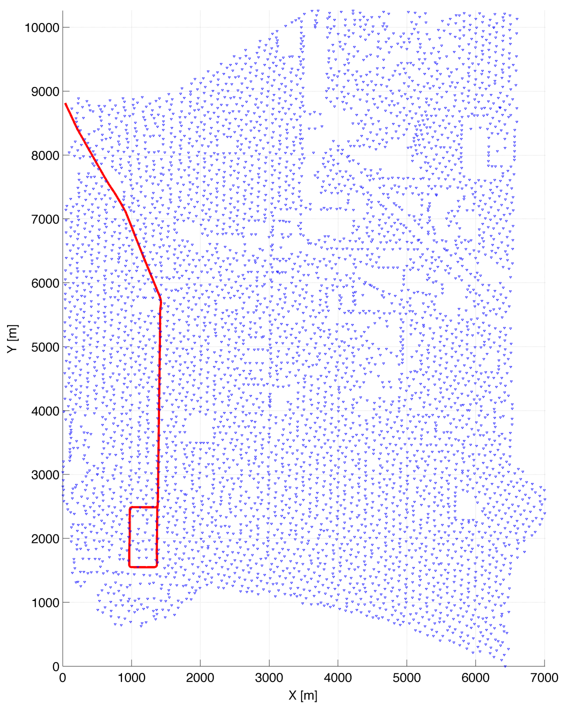
Noise Tracking of Cars/Trains/Airplanes

5200 element Long Beach array (Dan Hollis)

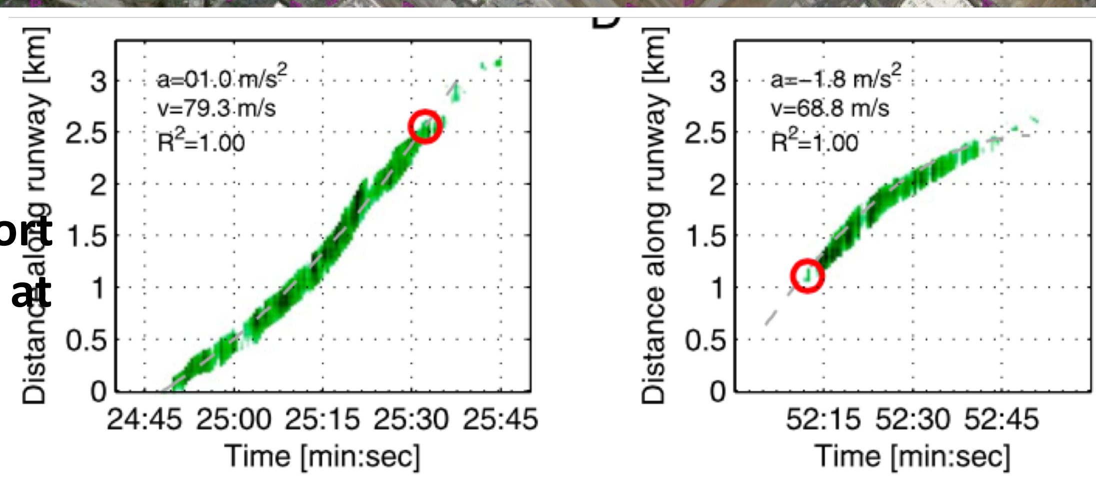


Noise Tracking of Cars/Trains/Airplanes





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

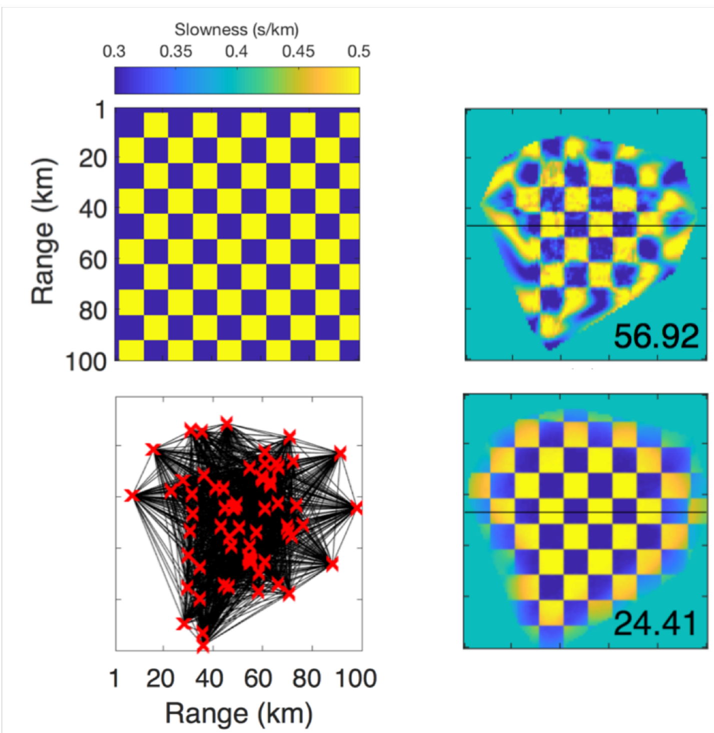


Accelerating airplane on Long Beach Airport runway, moving northwest and taking off at about 120 mi/h.

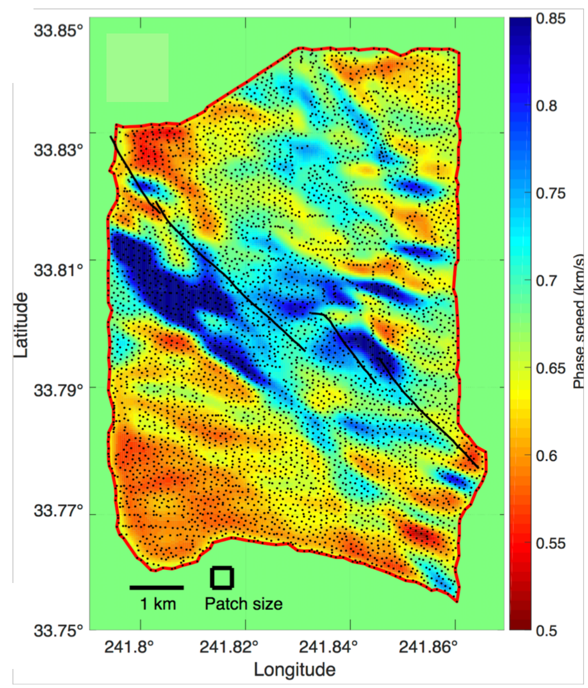
Riahi, Gerstoft, GRL 2015

"Travel time tomography with adaptive dictionaries"
Bianco and Gerstoft 2018, IEEE Transactions on Computational Imaging

- The Earth contains both smooth and discontinuous variations in slowness (e.g. Moho, faults) at multiple spatial scales
- Most existing travel time inversion methods are ad hoc: regularize inversion assuming exclusively smooth or discontinuous slownesses
- Propose locally-sparse 2D travel time tomography (LST) method with three main ingredients:
  - Sparsity constraint on slowness patches
  - Dictionary learning (unsupervised machine learning)
  - Damped least squares regularization on overall slowness map

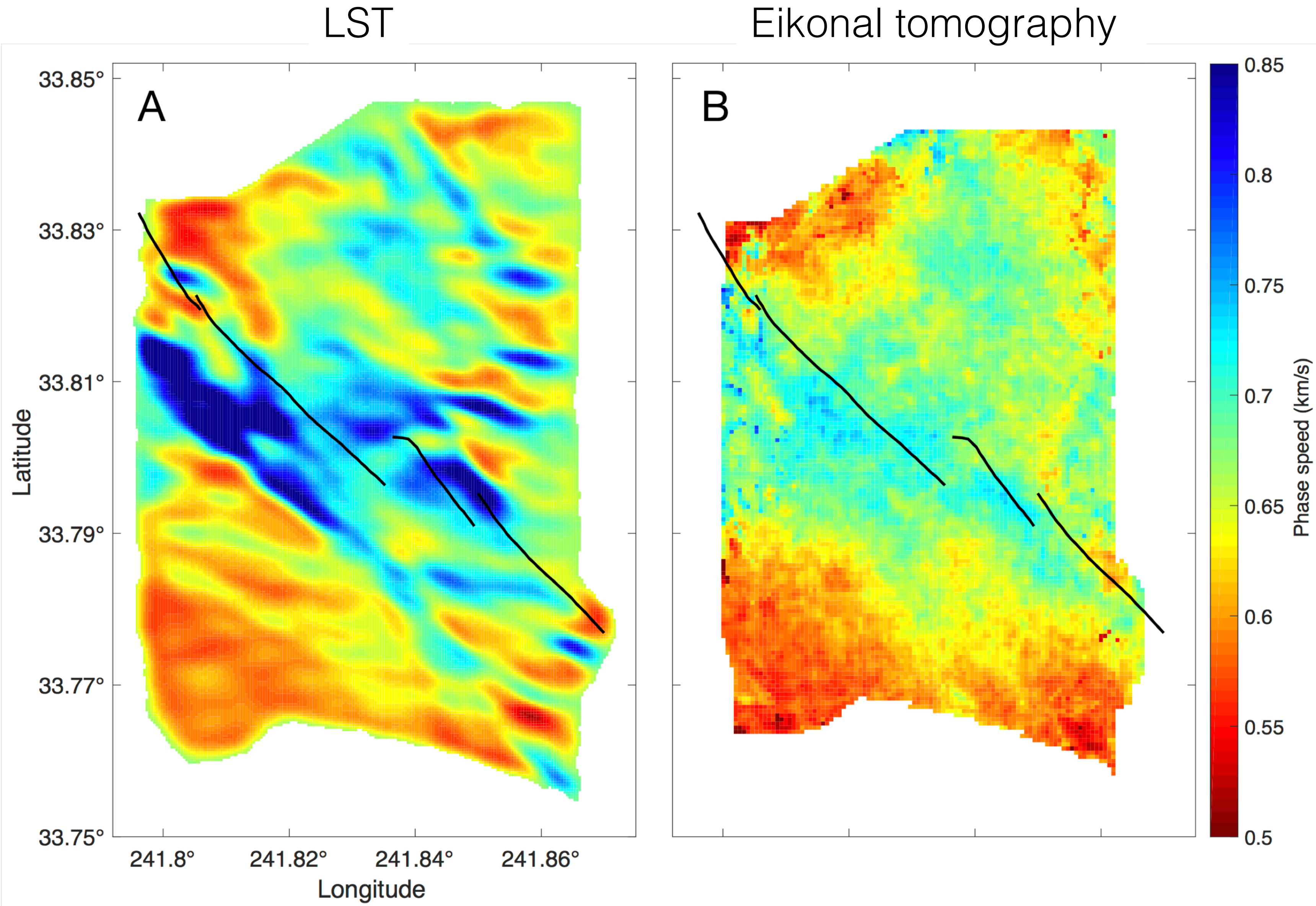


Synthetic checkerboard



LST in Long Beach, CA, USA

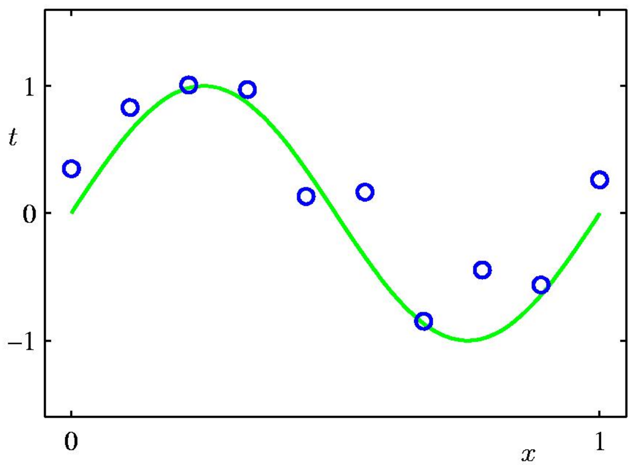
Comparison of LST with Eikonal Tomography (Lin et al. 2009)



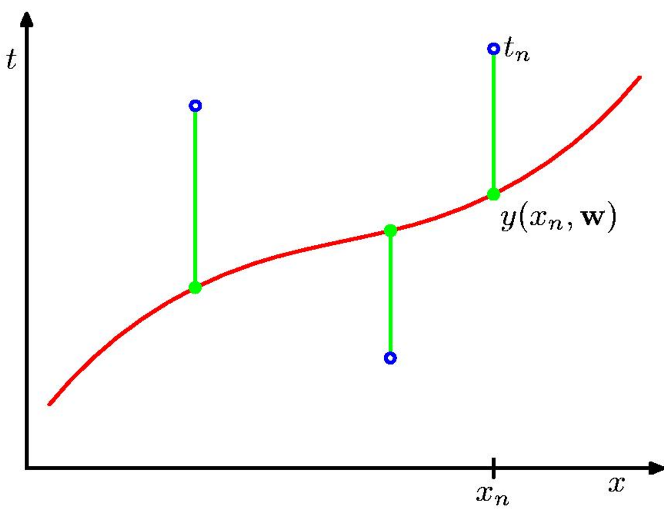
- BISHOP 1.2

Polynomial Curve Fitting

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



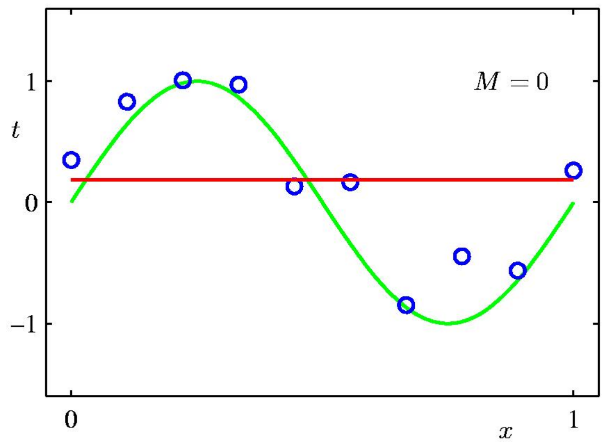
Sum-of-Squares Error Function



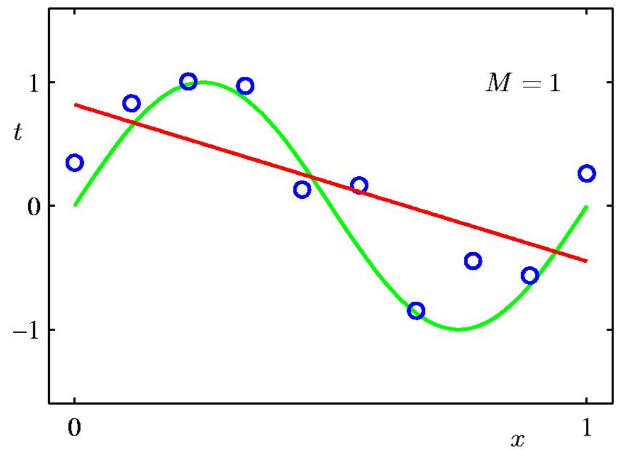
$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

M Order Polynomial Fit

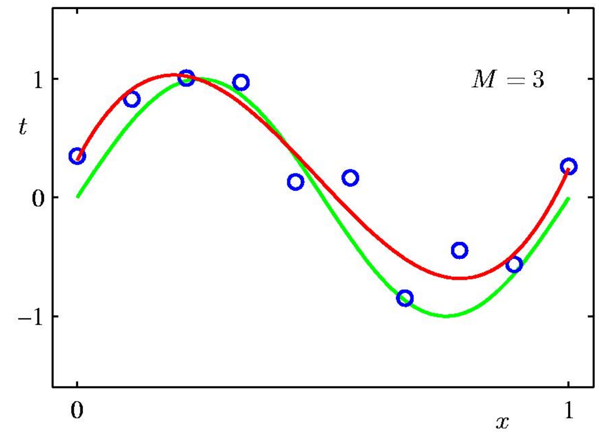
0 Order Polynomial



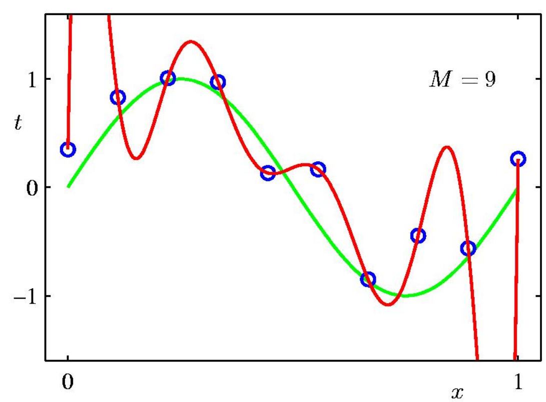
1st Order Polynomial

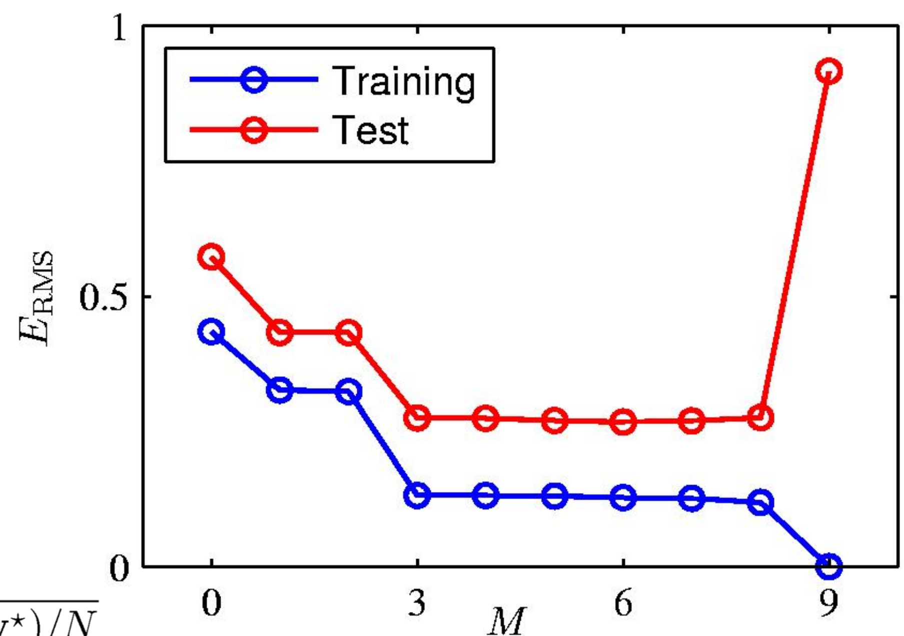


3 Order Polynomial



9 Order Polynomial



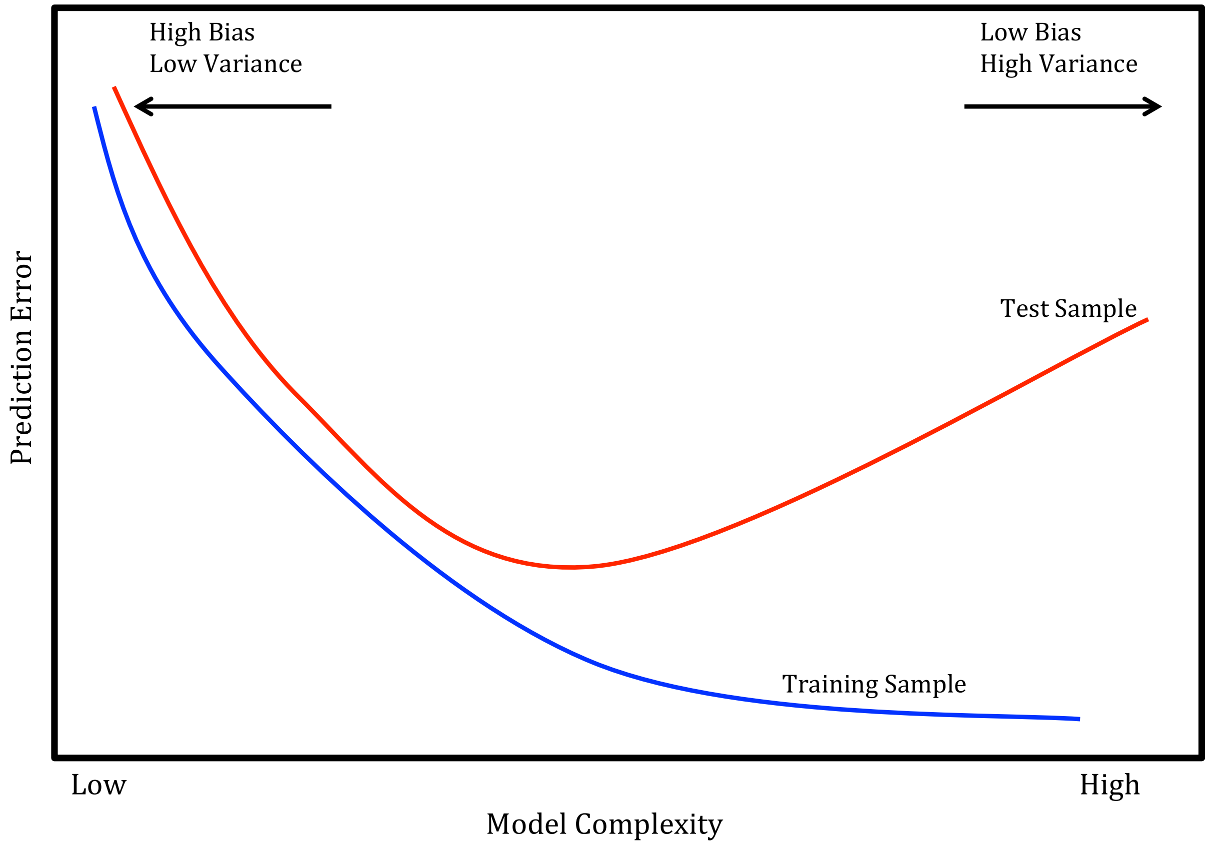


Root-Mean-Square (RMS) Error:

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

Bias-variance tradeoff

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



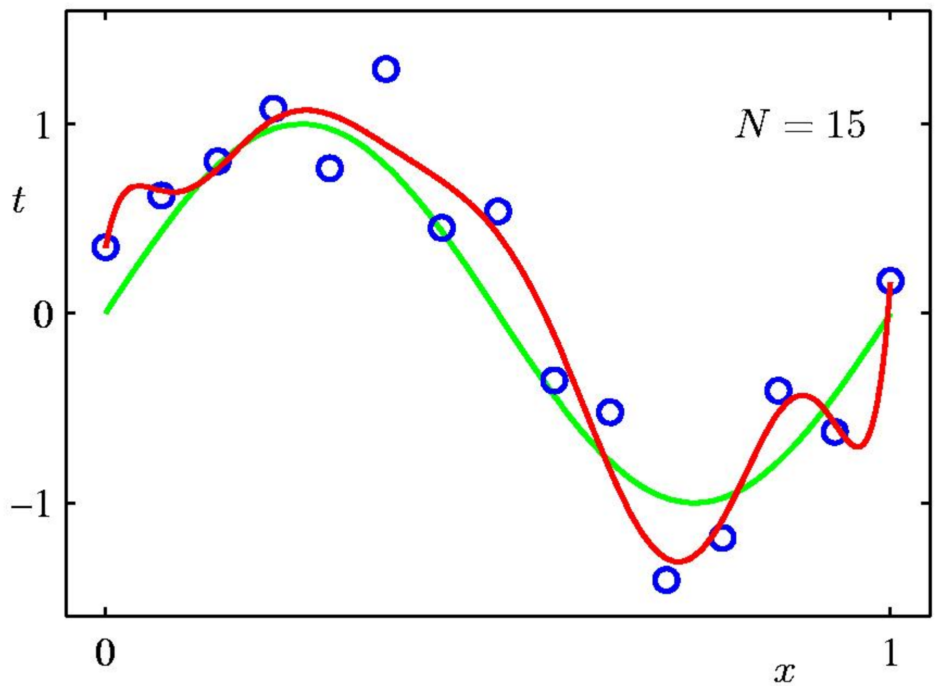
Polynomial Coefficients

	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^\star$	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

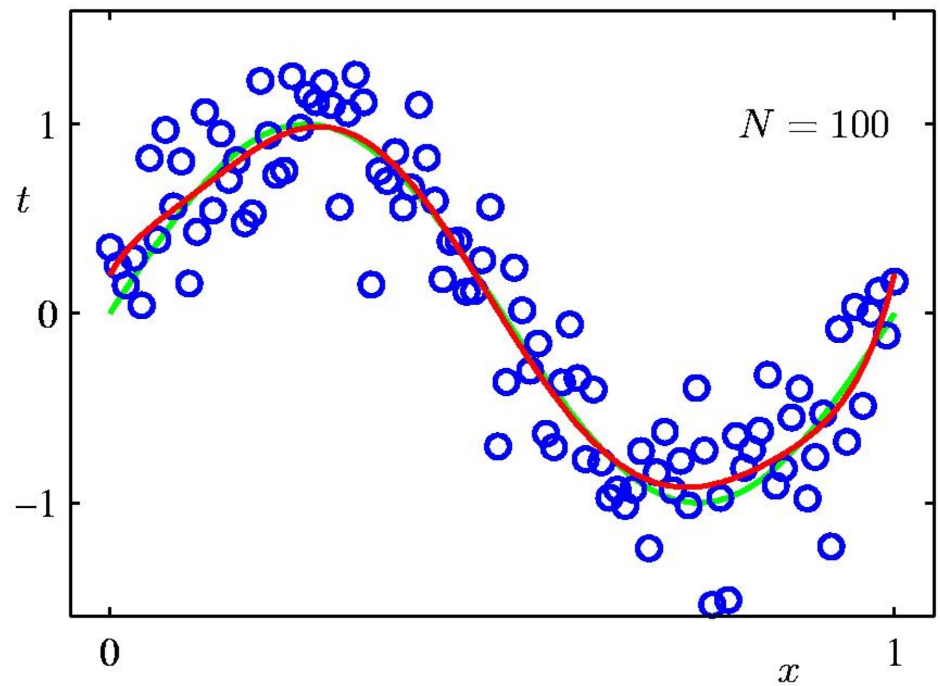
Data Set Size:

9th Order Polynomial

$N = 15$



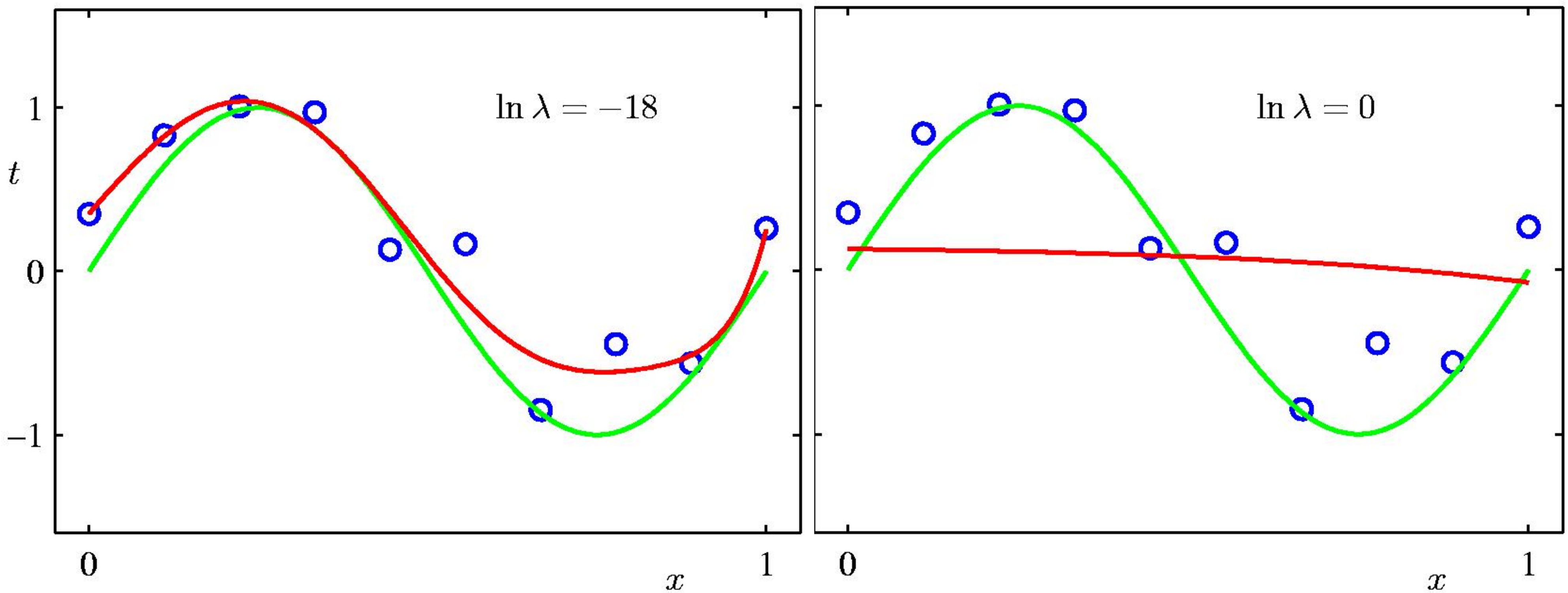
$N = 100$



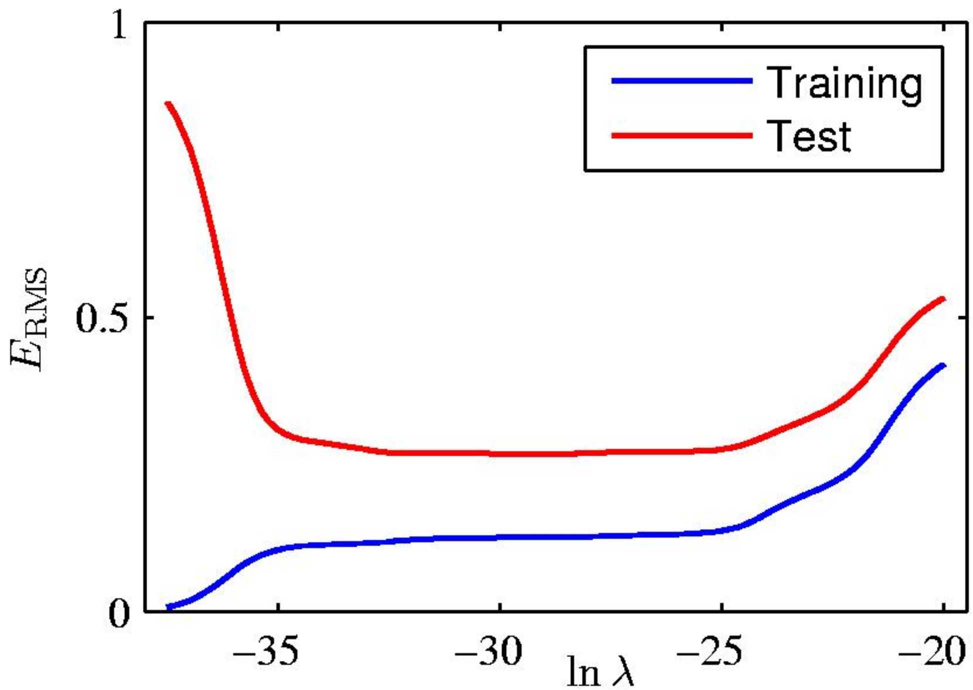
Regularization

- Penalize large coefficient values

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



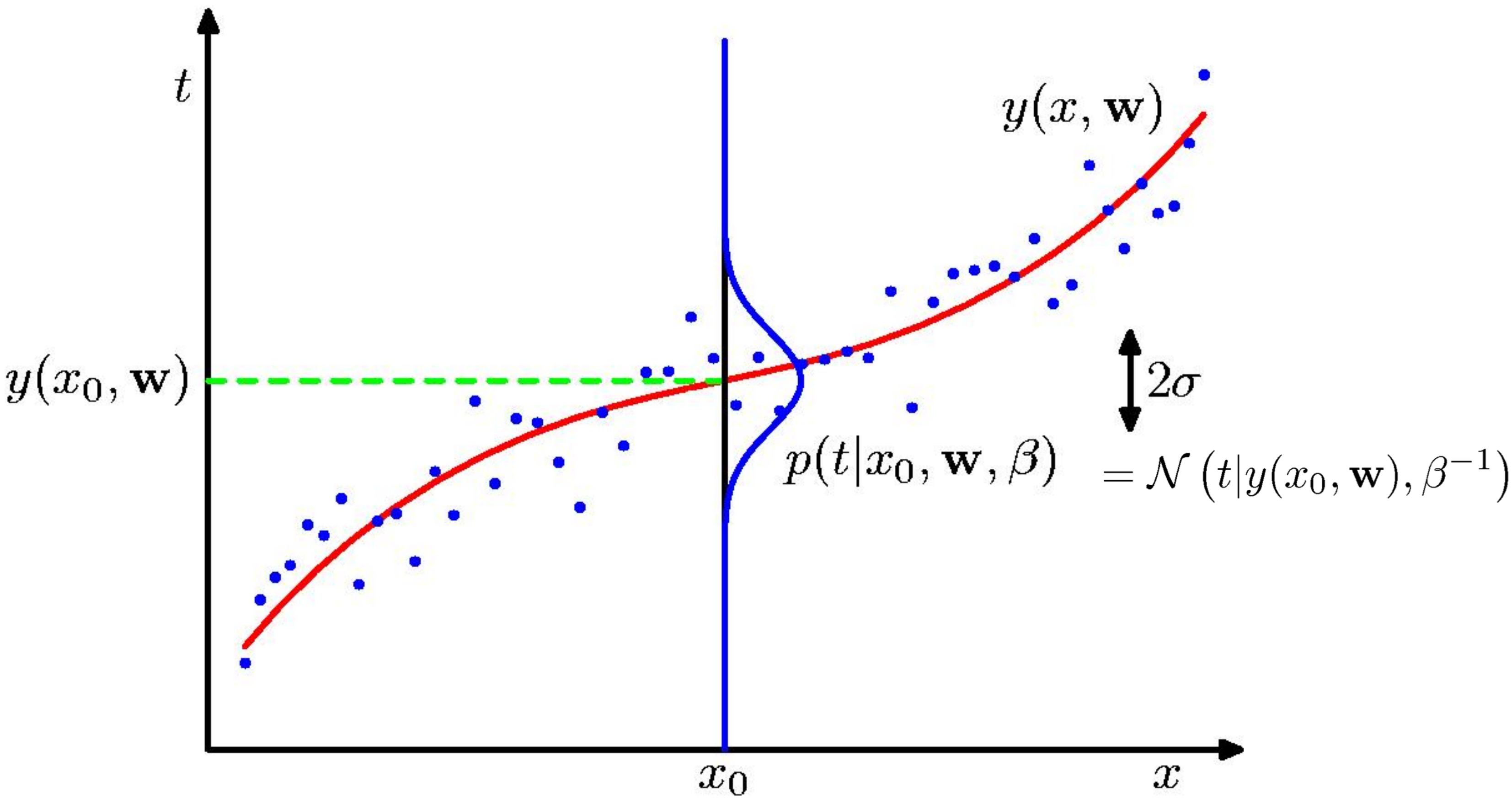
Regularization: E_{RMS} vs. $\ln \lambda$



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

Curve Fitting Re-visited, Bishop1.2.5



Maximum Likelihood Bishop 1.2.5

- Model

- Likelihood

- differentiation

Maximum Likelihood

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right). \tag{1.61}$$

As we did in the case of the simple Gaussian distribution earlier, it is convenient to maximize the logarithm of the likelihood function. Substituting for the form of the Gaussian distribution, given by (1.46), we obtain the log likelihood function in the form

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi). \tag{1.62}$$

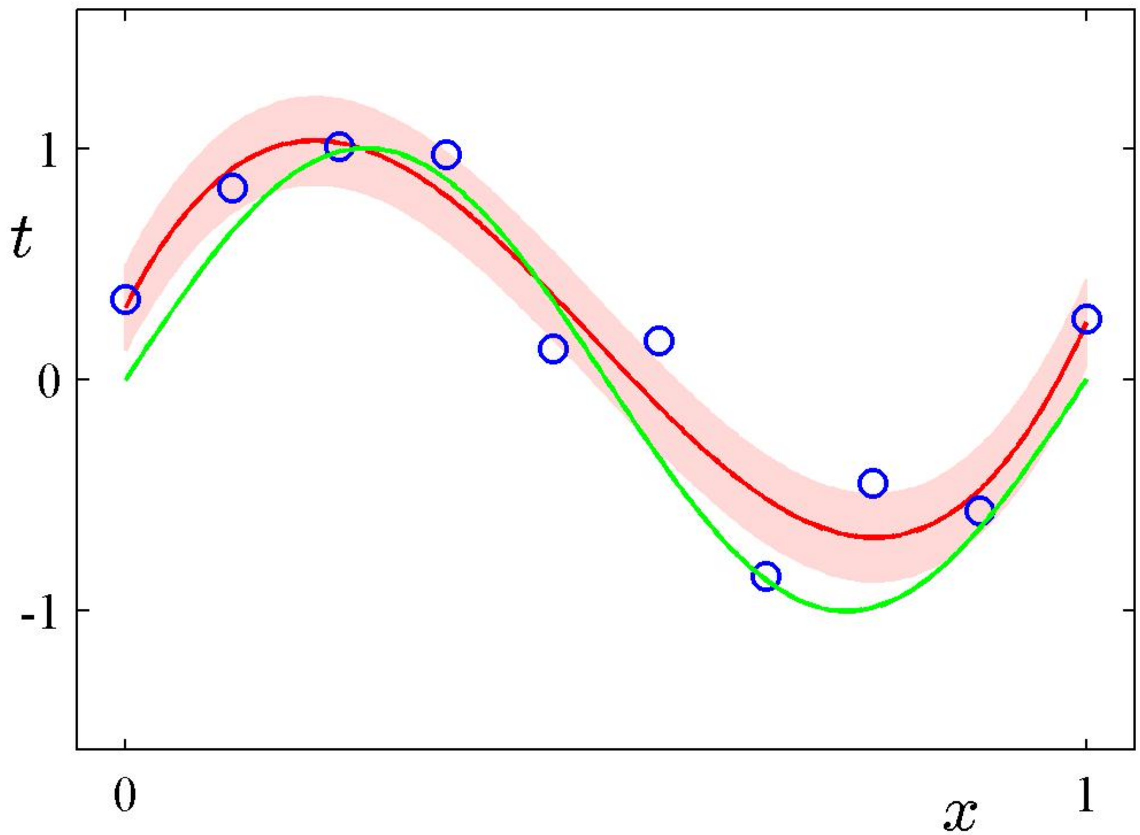
$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}_{\mathrm{ML}}) - t_n\}^2. \tag{1.63}$$

Giving estimates of W and beta, we can predict

$$p(t|x, \mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}) = \mathcal{N}\left(t | y(x, \mathbf{w}_{\mathrm{ML}}), \beta_{\mathrm{ML}}^{-1}\right). \tag{1.64}$$

Predictive Distribution

$$p(t|x, \mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}) = \mathcal{N}\left(t|y(x, \mathbf{w}_{\mathrm{ML}}), \beta_{\mathrm{ML}}^{-1}\right)$$



MAP: A Step towards Bayes 1.2.5

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left\{-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}$$

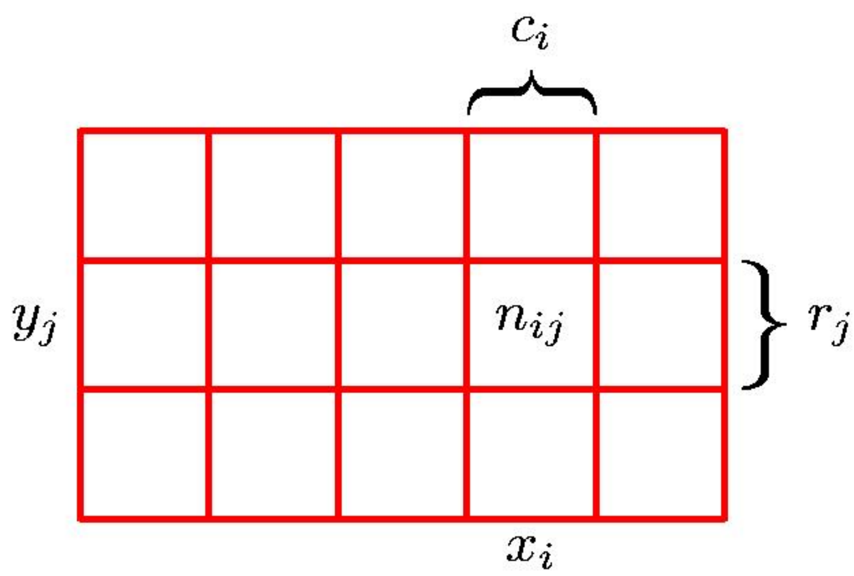
$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

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

Determine $\mathbf{w}_{\mathrm{MAP}}$ by minimizing regularized sum-of-squares error, $\widetilde{E}(\mathbf{w})$.

Regularized sum of squares

Probability Theory



Joint Probability

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

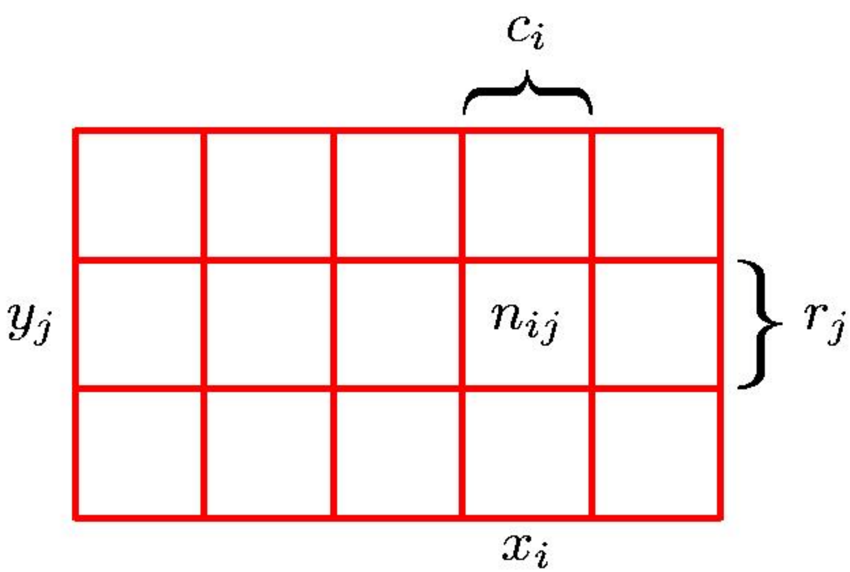
Marginal Probability

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

Conditional Probability

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

Probability Theory



- **Sum Rule**

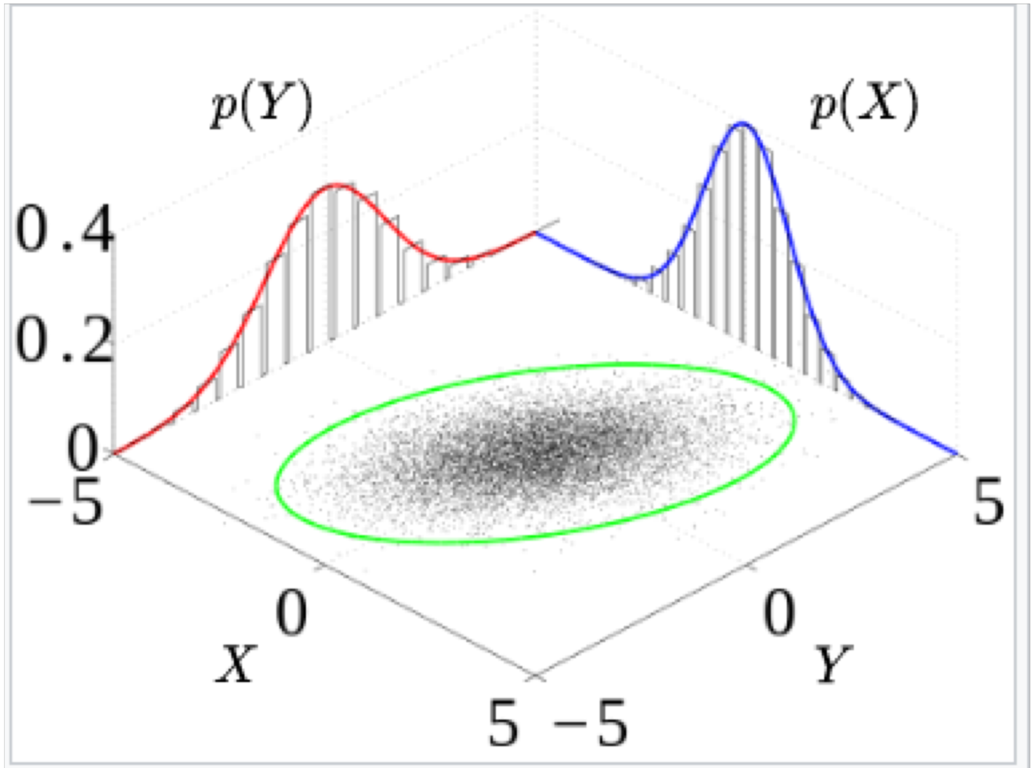
$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

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

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

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

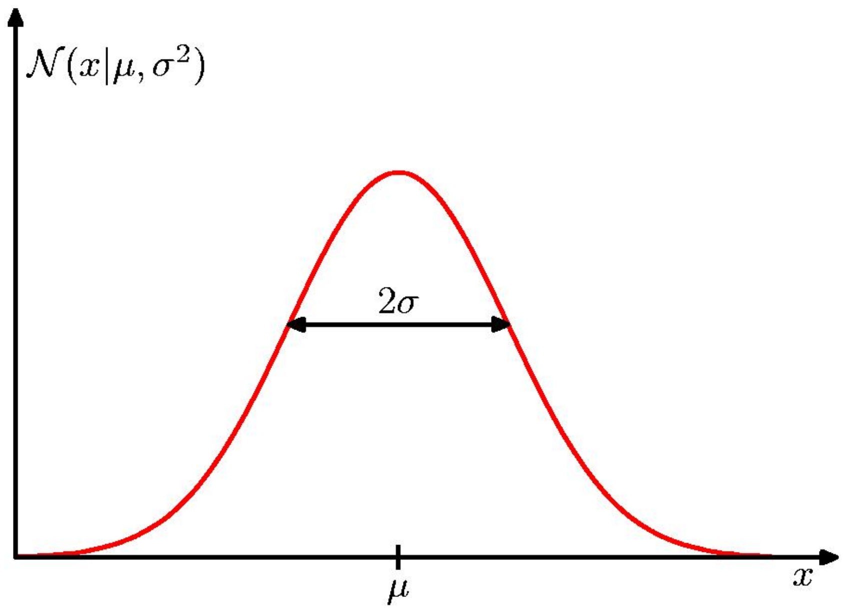
Rev'd Thomas Bayes (1702–1761)

- 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

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



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

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

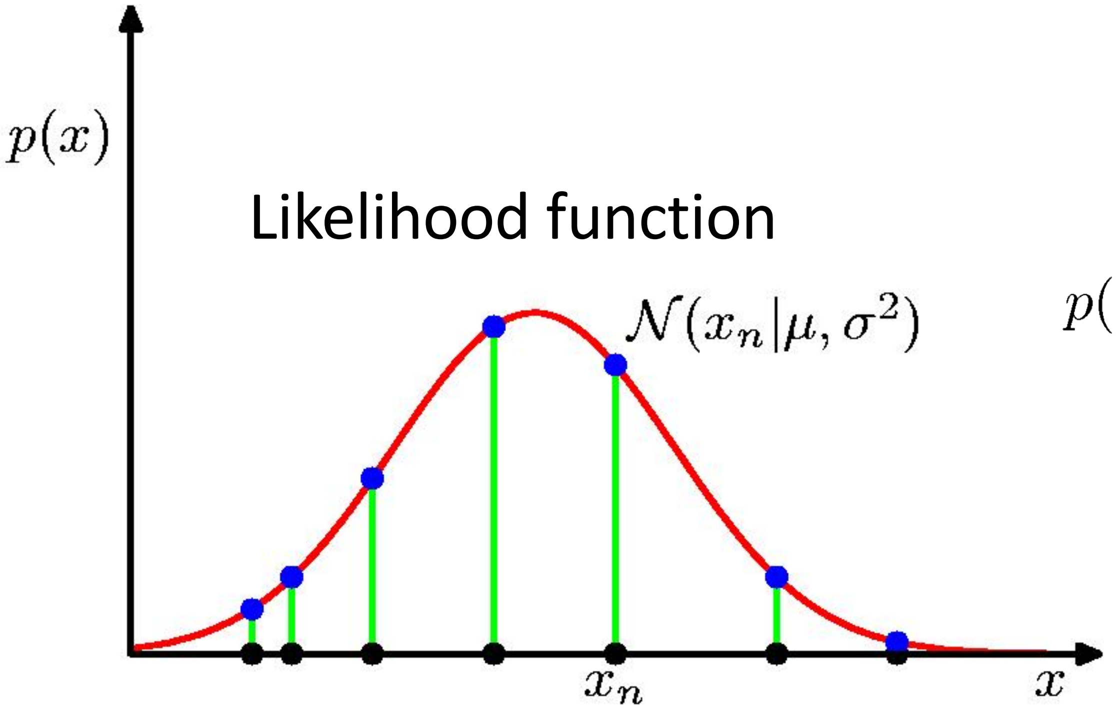
Gaussian Mean and Variance

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

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

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

Gaussian Parameter Estimation



Likelihood function

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

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.