

2019: 224 Students 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<sup>20</sup>

Please talk about homework, but don't copy

Maybe some SciKit Learn on Jupyter Notebook (TA problem)

Piazza help

# GPU datahub.ucsd.edu

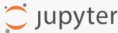
<https://datahub.ucsd.edu/hub/login>  
[Documentation](#)

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

DATA SCIENCE / MACHINE LEARNING PLATFORM

UC San Diego

Information Technology Services - Educational Technology Services    Help Options ▾

 Home    Token Logout

## Spawner Options

☐ py3torch-cuda9 (2 CPU, 8GB RAM)  
(Deprecated WI 19- use Scientific Python+ML) Python 3, PyTorch 1.0.0, TensorFlow 1.11.0

☐ Scientific Python + Machine Learning Tools (2 CPU, 8GB RAM)  
ucsdets/scipy-ml-notebook: Python 3, PyTorch 1.0.1, TensorFlow 1.12.0 (replaces ets-pytorch)

☐ Data Science: Base Notebook (2 CPU, 4GB RAM)  
ucsdets/datascience: Julia, R, Python 3 (based on jupyter/datascience-notebook)

☐ ECE228\_SP19\_A00: Scientific Python + Machine Learning Tools (2 CPU, 8GB RAM)  
ucsdets/scipy-ml-notebook: Python 3, PyTorch 1.0.1, TensorFlow 1.12.0 (replaces ets-pytorch)

☐ ECE228\_SP19\_A00: Scientific Python + Machine Learning Tools (1 GPU, 4 CPU, 16GB RAM)  
ucsdets/scipy-ml-notebook: Python 3, PyTorch 1.0.1, TensorFlow 1.12.0 (replaces ets-pytorch)

Spawn

# Projects

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

## 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

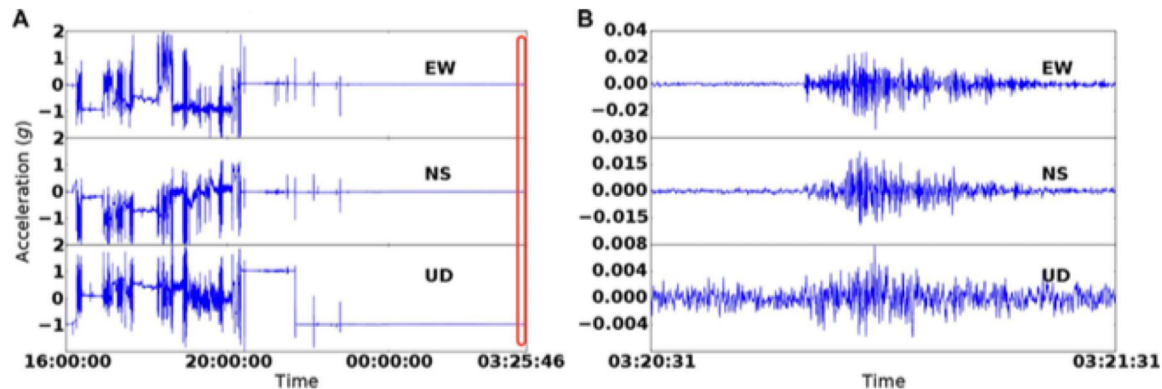
# MyShake: A smartphone seismic network for earthquake early warning and beyond

Qingkai Kong,<sup>1\*</sup> Richard M. Allen,<sup>1</sup> Louis Schreier,<sup>2</sup> Young-Woo Kwon<sup>3</sup>

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

2016 © The Authors, some rights reserved;  
exclusive licensee American Association for  
the Advancement of Science. Distributed  
under a Creative Commons Attribution  
NonCommercial License 4.0 (CC BY-NC).  
10.1126/sciadv.1501055

## 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.

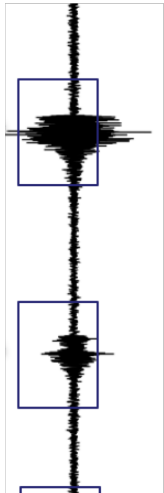
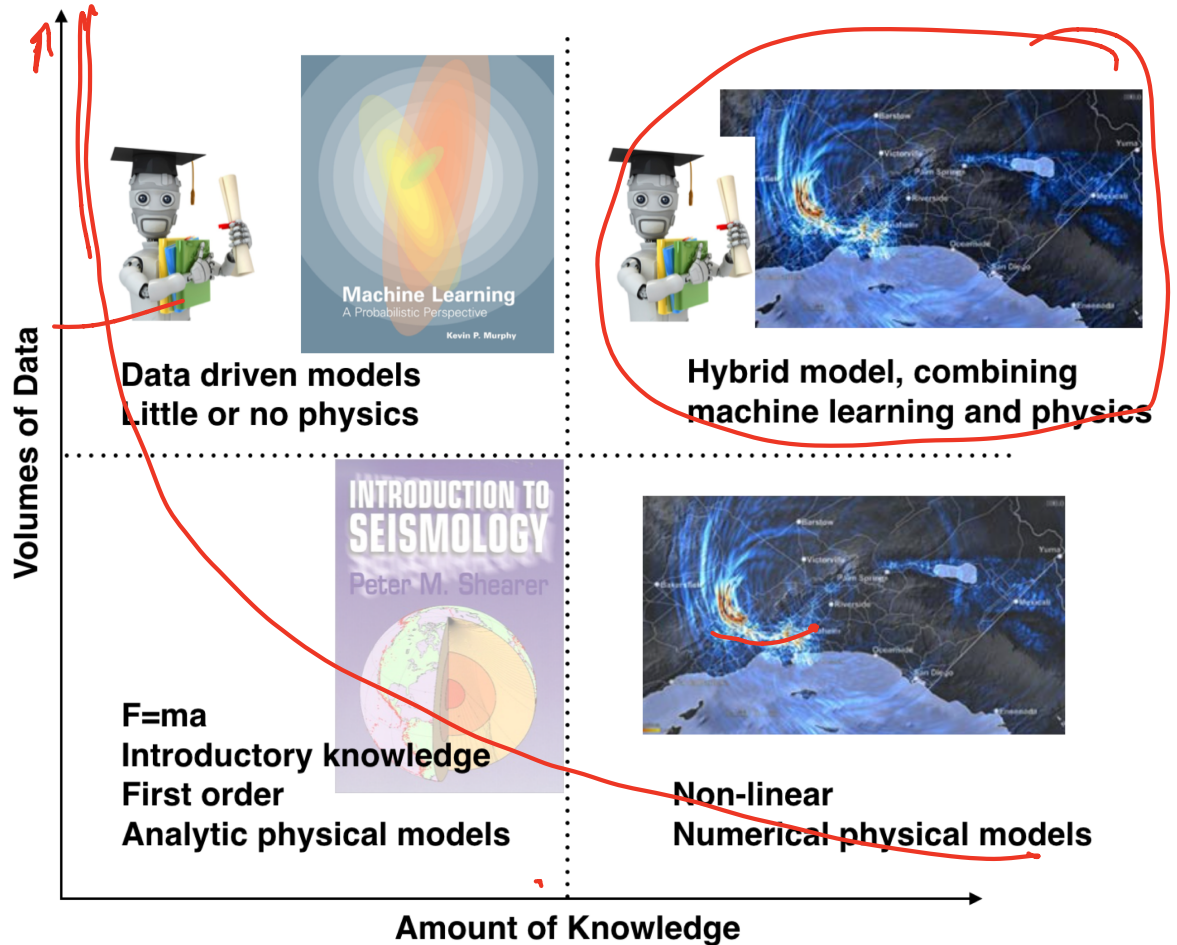
SIO

Signal-Proc

Peter

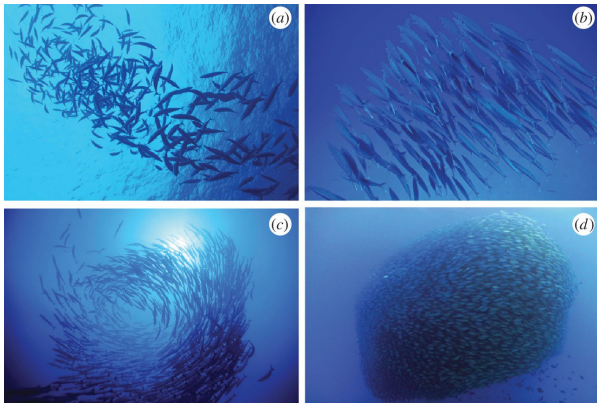
Google

# Machine learning versus knowledge based

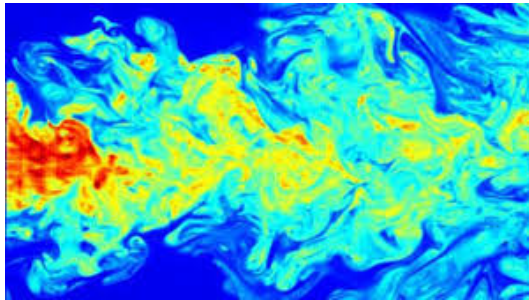


3D spectral elements

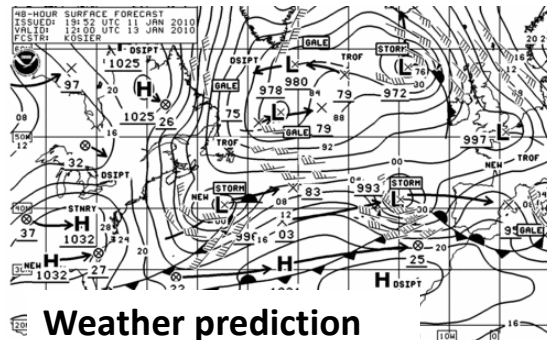




**Back scattering from fish school**

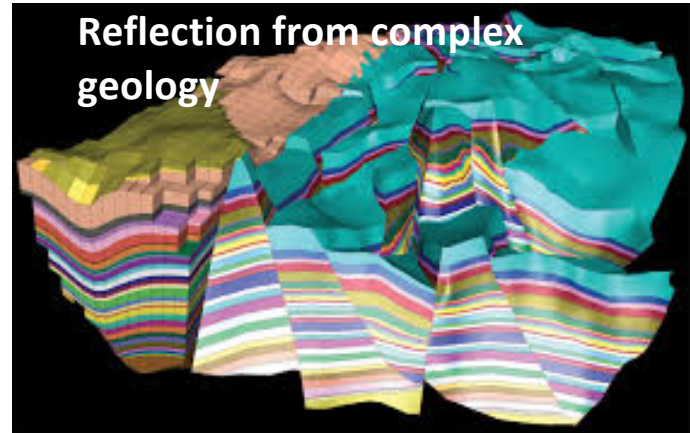


**Predict acoustic field in turbulence**



**Weather prediction**

We can't model everything...



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

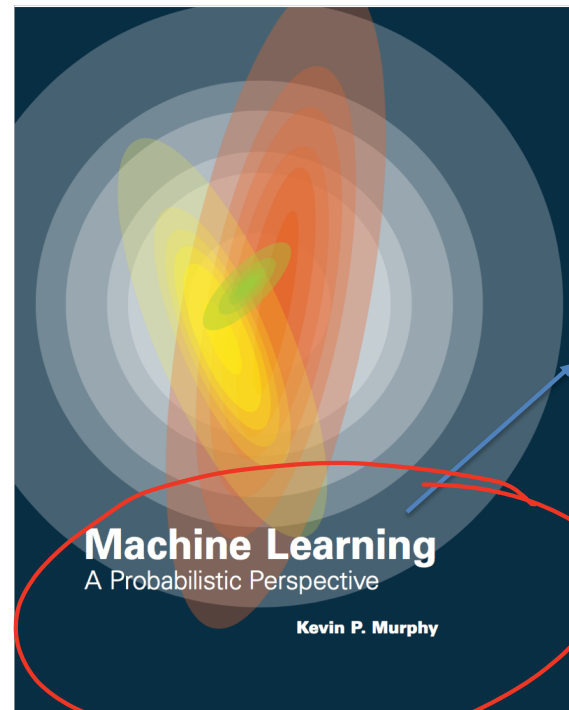
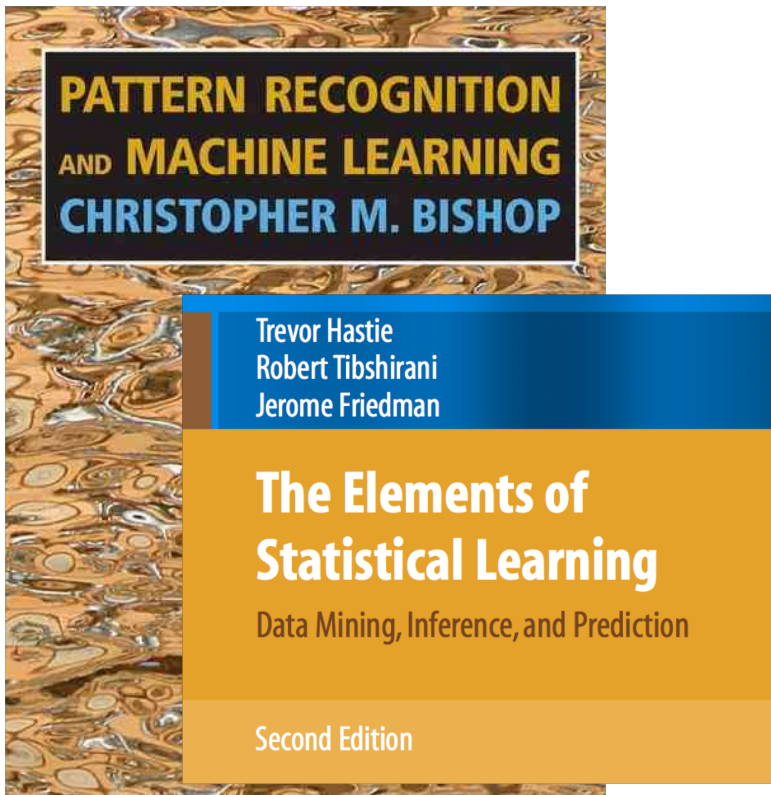
**Dolphins = real ML!**



# 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

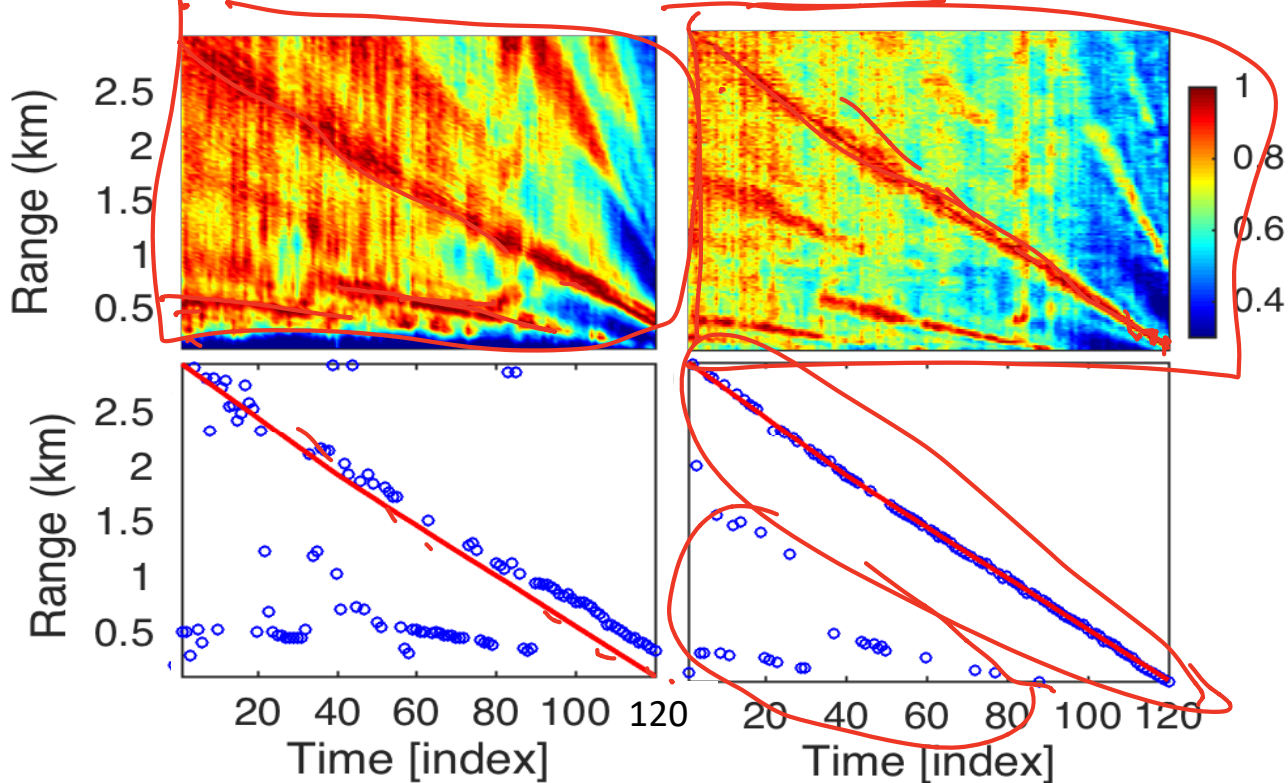
Frequencies [300:10:950]Hz

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

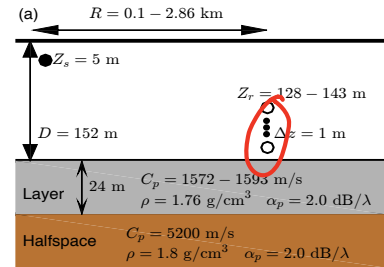
$$B = p^H C_p$$

synthetic replicas.

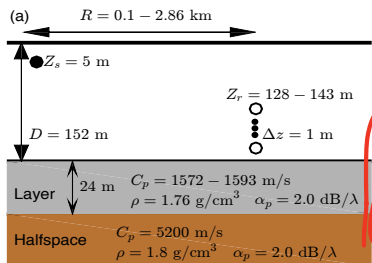
measured replicas



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

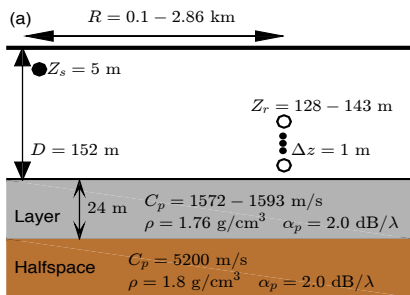


# Classification versus regression



N potential source ranges  
 $R = \{r_1, \dots, r_N\}$

**Regression:**



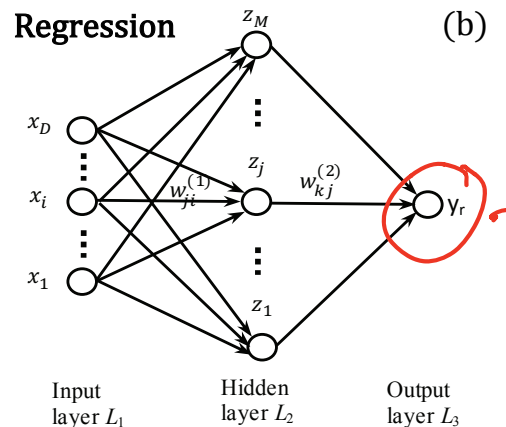
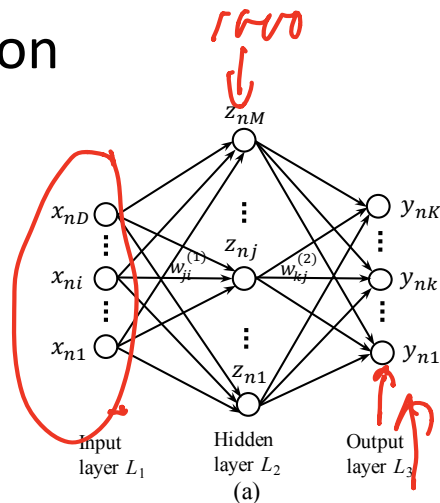
one source continuous range

**Regression is harder**

Number of parameters

MFP:  $O(10)$

ML:  $400 \cdot 1000 + 1000 \cdot 1000 + 1000 \cdot 100$   
 $= O(1000000)$



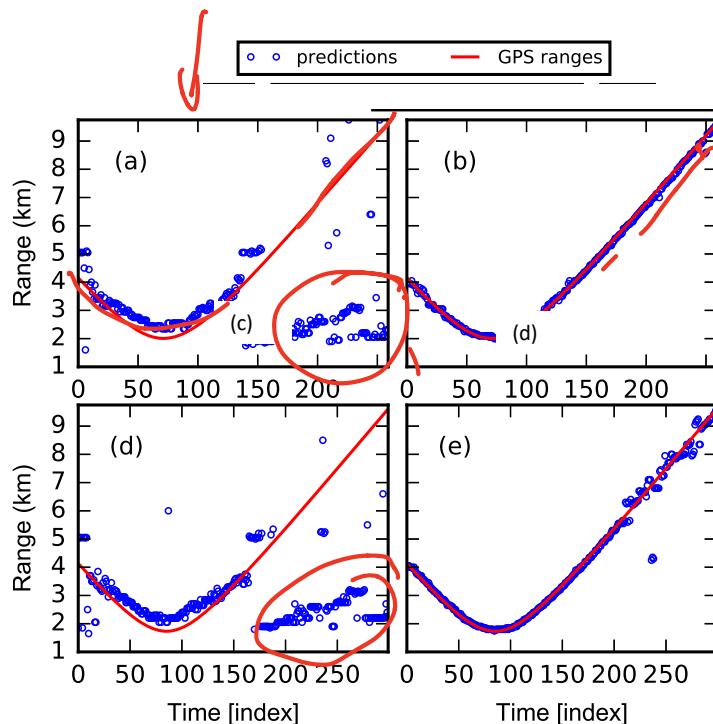
## So far...

- 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 al. 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

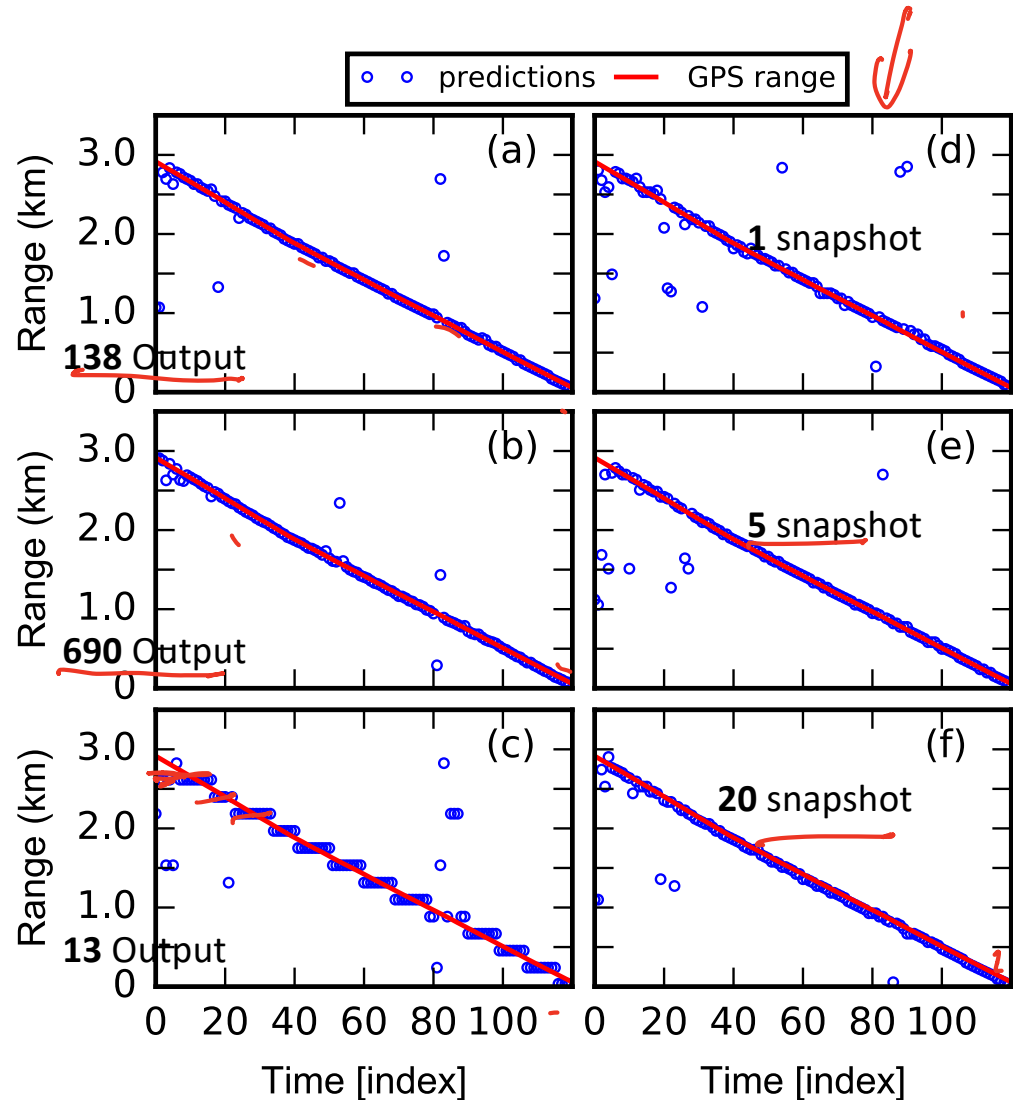
60s Science  
Scientific Am



# Other parameters: FNN

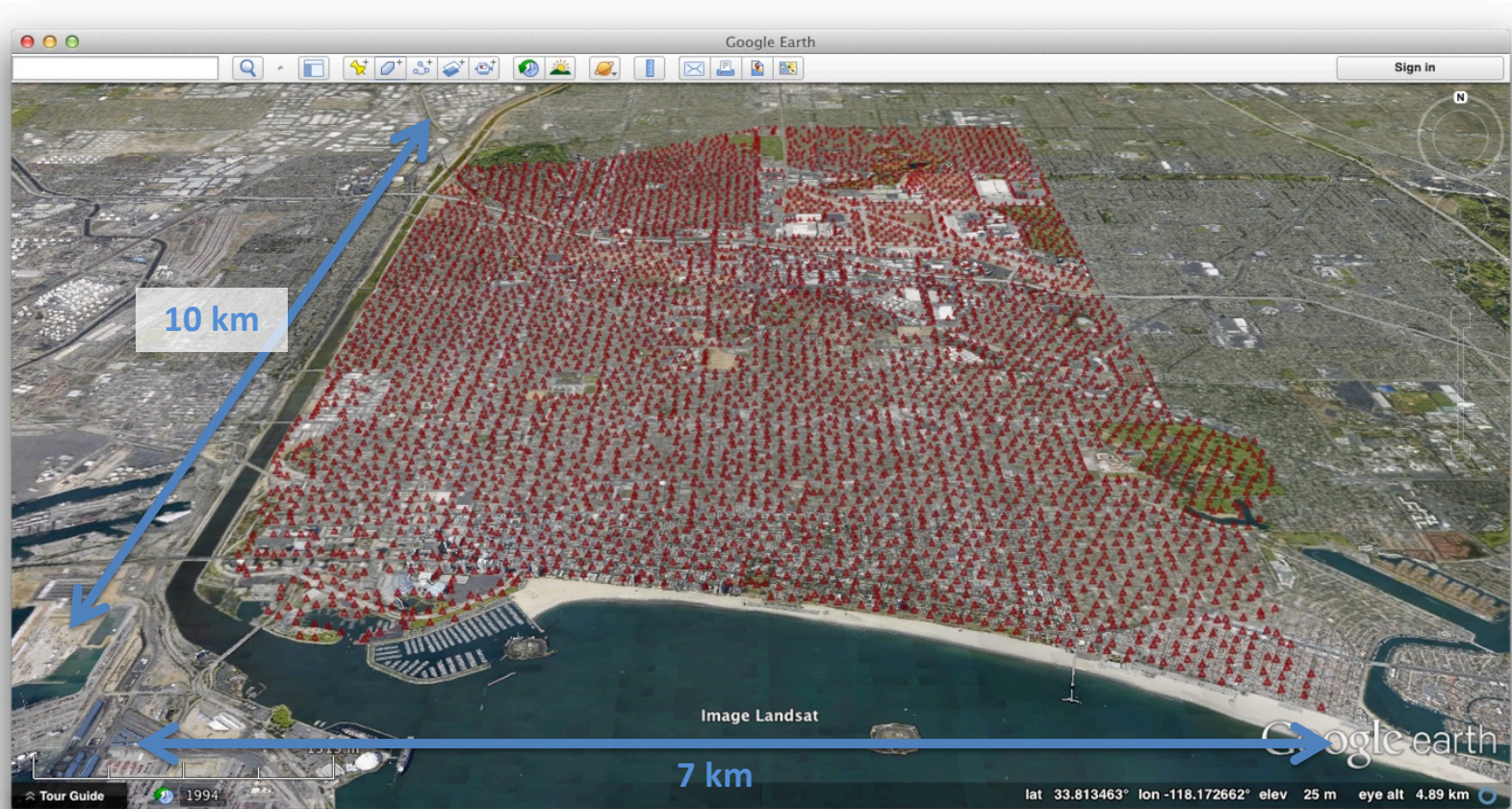
## Conclusion

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



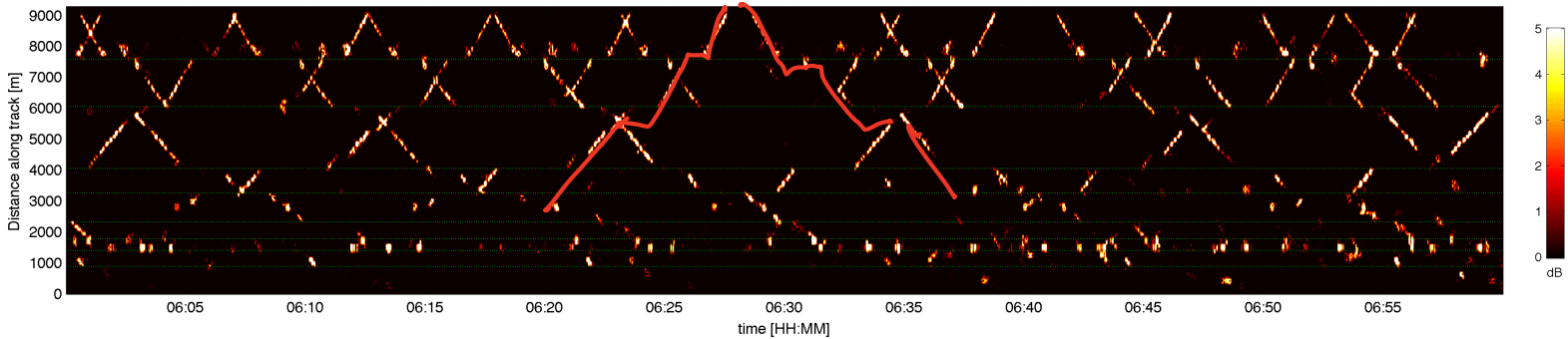


# Why we got interested in traffic

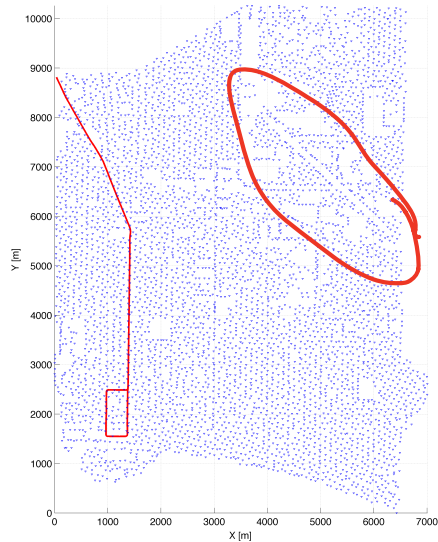


March 5—12, 2011

# Noise Tracking of Cars/Trains/Airplanes

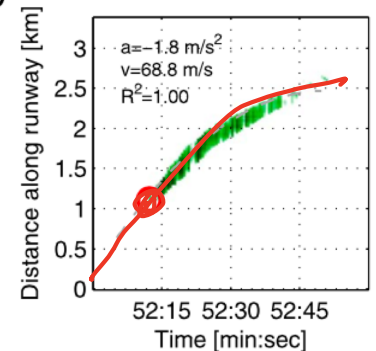
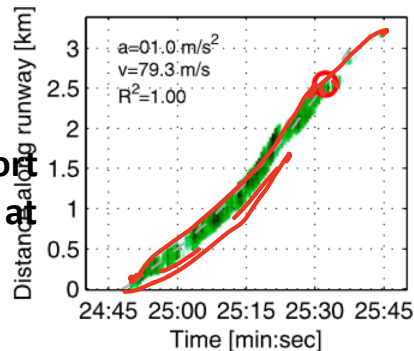


March 7<sup>th</sup>, 6-7am, rush hour, Blue Line



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

Riahi, Gerstoft, GRL 2015

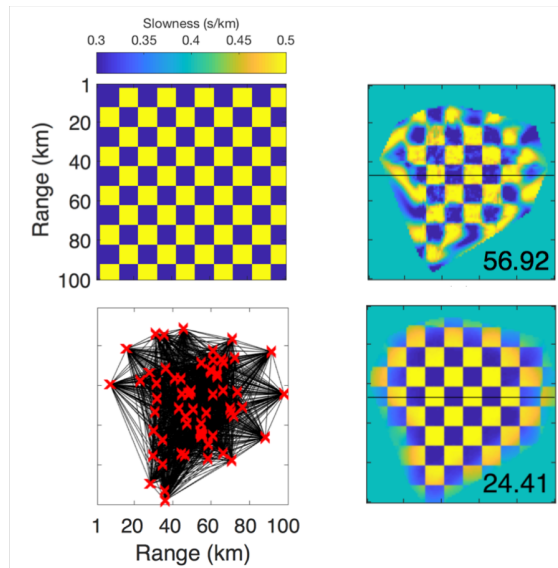




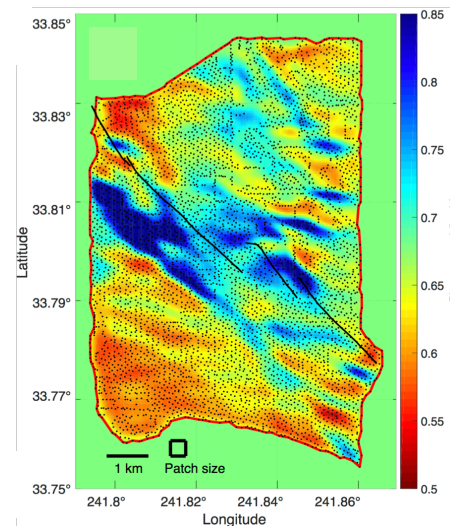
# "Travel time tomography with adaptive dictionaries"

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

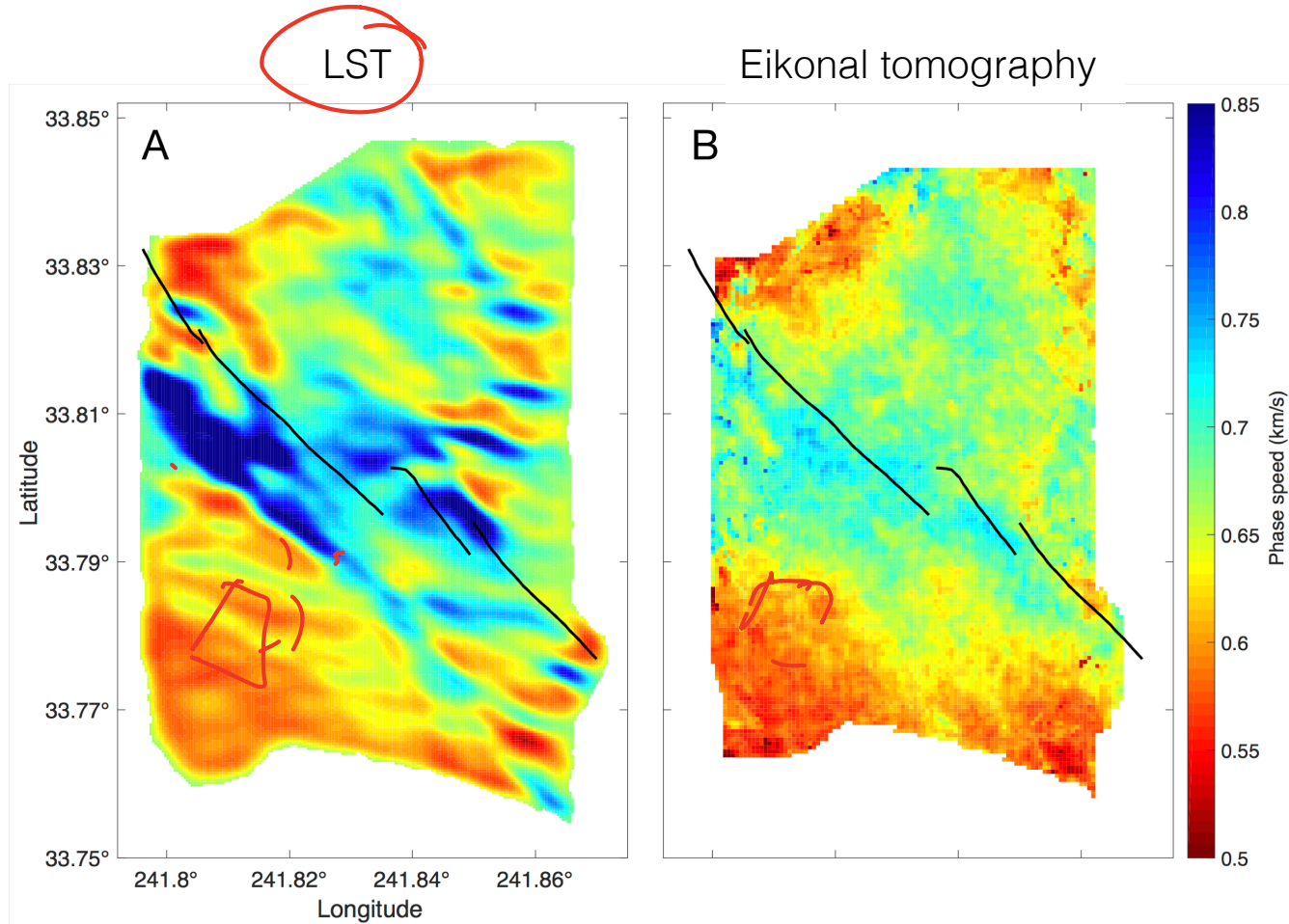


Synthetic checkerboard



LST in Long Beach, CA, USA

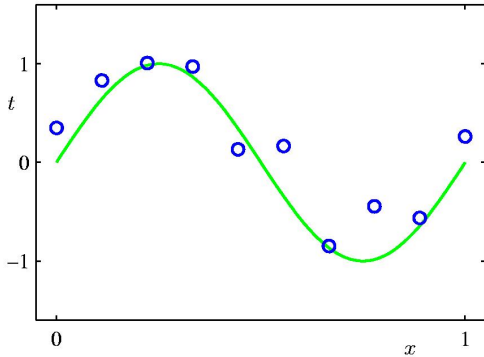
# Comparison of LST with Eikonal Tomography (Lin et al. 2009)



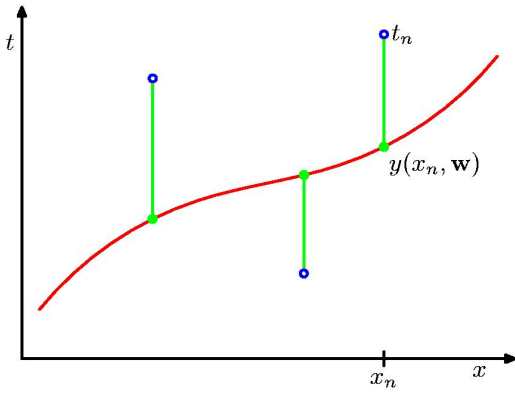
- BISHOP 1.2

# Polynomial Curve Fitting

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$



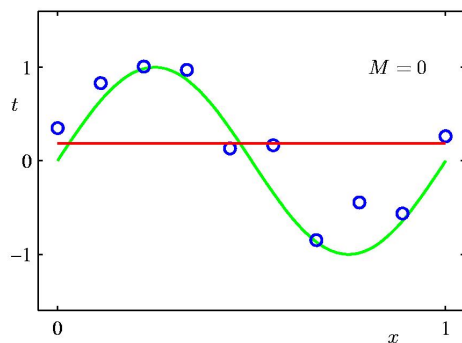
## Sum-of-Squares Error Function



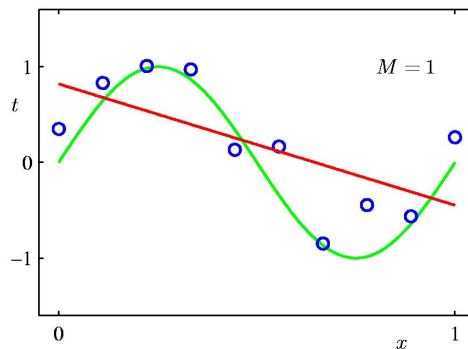
$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

# M Order Polynomial Fit

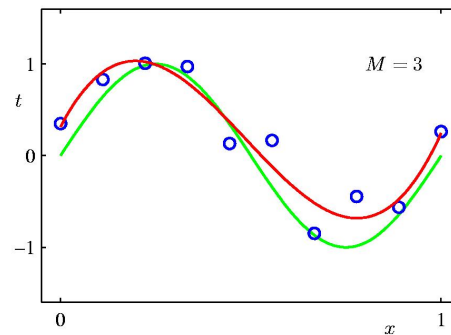
## 0 Order Polynomial



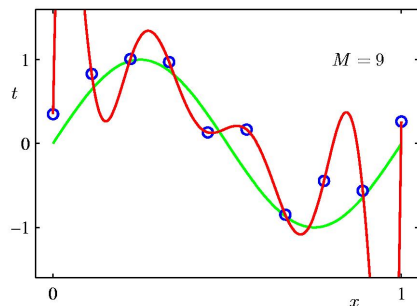
## 1<sup>st</sup> Order Polynomial



## 3 Order Polynomial

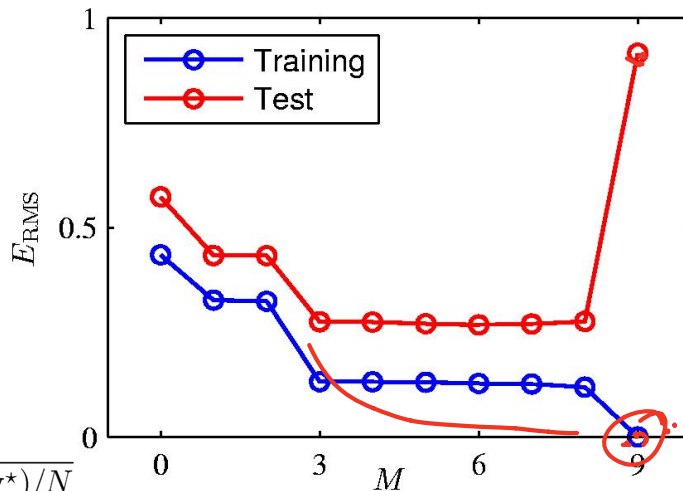


## 9 Order Polynomial



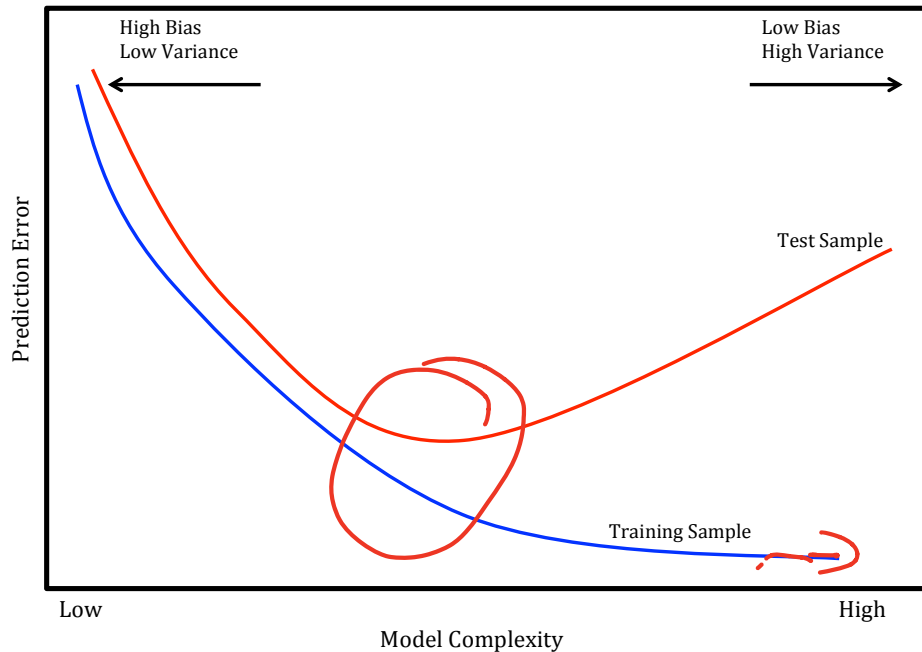
Root-Mean-Square (RMS) Error:

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



# Bias-variance tradeoff

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



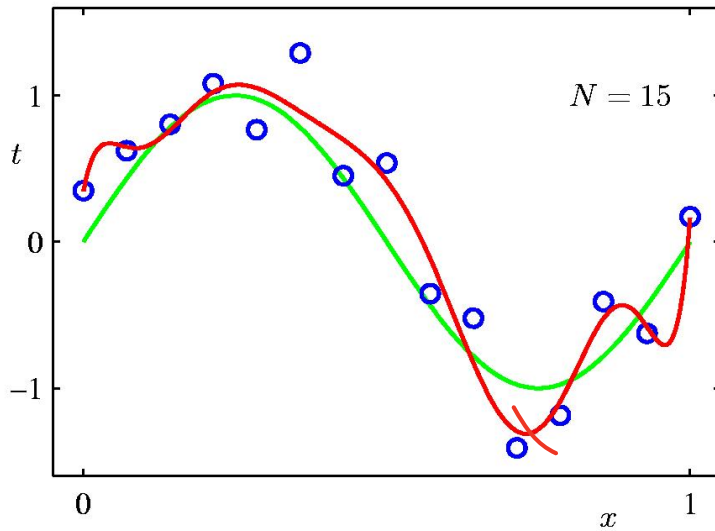
# Polynomial Coefficients

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

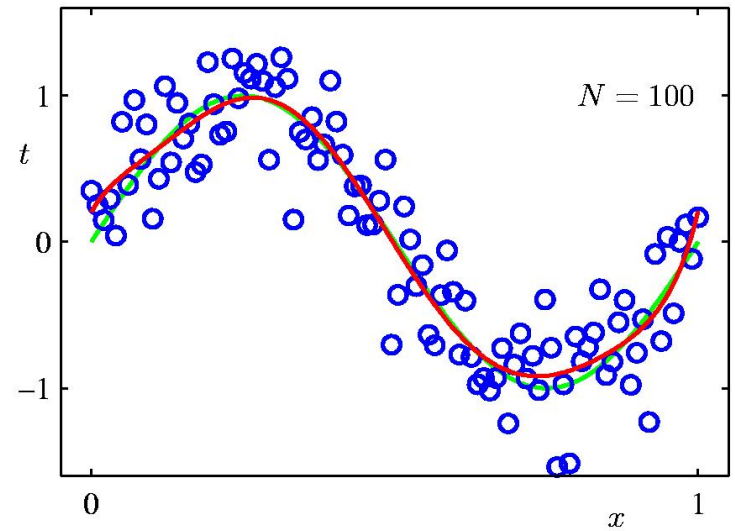
# Data Set Size:

9<sup>th</sup> Order Polynomial

$N = 15$



$N = 100$

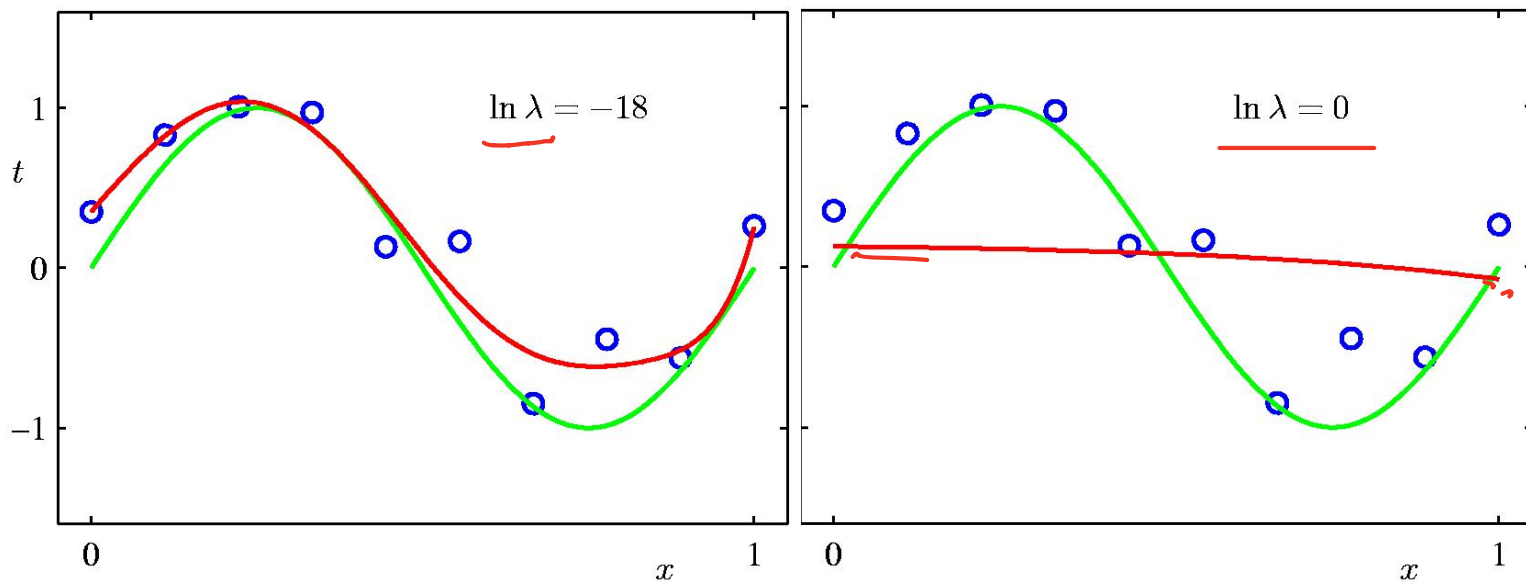




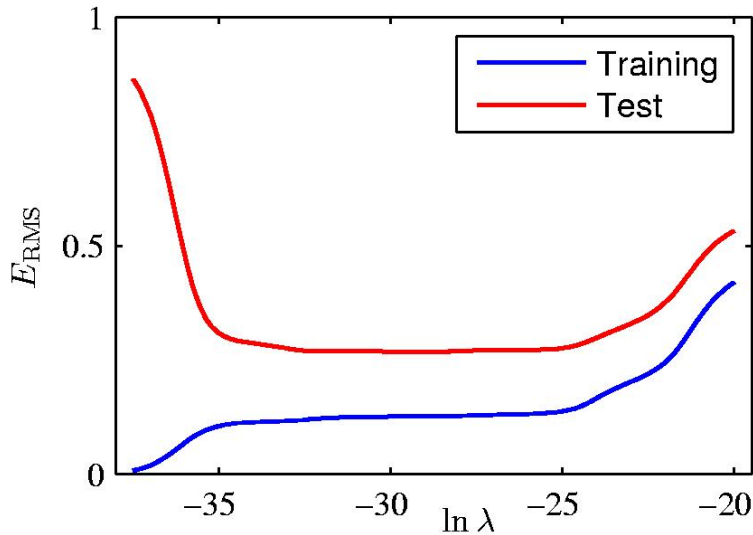
# Regularization

- Penalize large coefficient values

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$



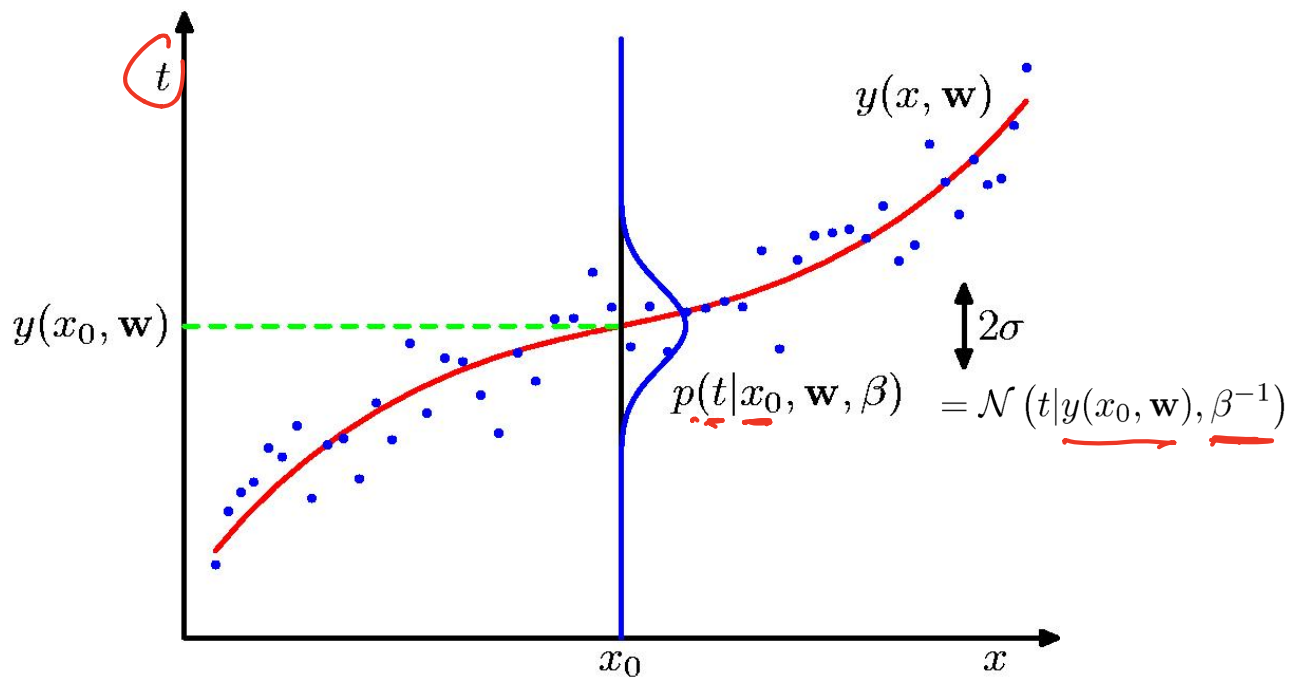
# Regularization: $E_{\text{RMS}}$ vs. $\ln \lambda$



Polynomial Coefficients

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

## Curve Fitting Re-visited, Bishop1.2.5



# Maximum Likelihood Bishop 1.2.5

- Model

$$t = y + n = \underline{w}^T \underline{x} + n$$

$$x = [x_0 \ x_1^2 \ x_2^3 \ \dots]$$

$$n \sim N(0, \beta^{-1})$$

$$N(0, \sigma^2)$$

$$t \sim N(w^T x, \sigma^2)$$

- Likelihood

$$l(x) = \prod_{n=1}^N N(t_n | w^T x, \beta^{-1})$$

- differentiation

$$L = \frac{1}{2\sigma^2} \sum (t_n - w_n^T x)^2 + \frac{N}{2} \log(\sigma^2 \cdot 2\pi)$$

$$\frac{\partial L}{\partial \sigma} = 0$$

$$\frac{\partial L}{\partial w} = 0$$

# Maximum Likelihood

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

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

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

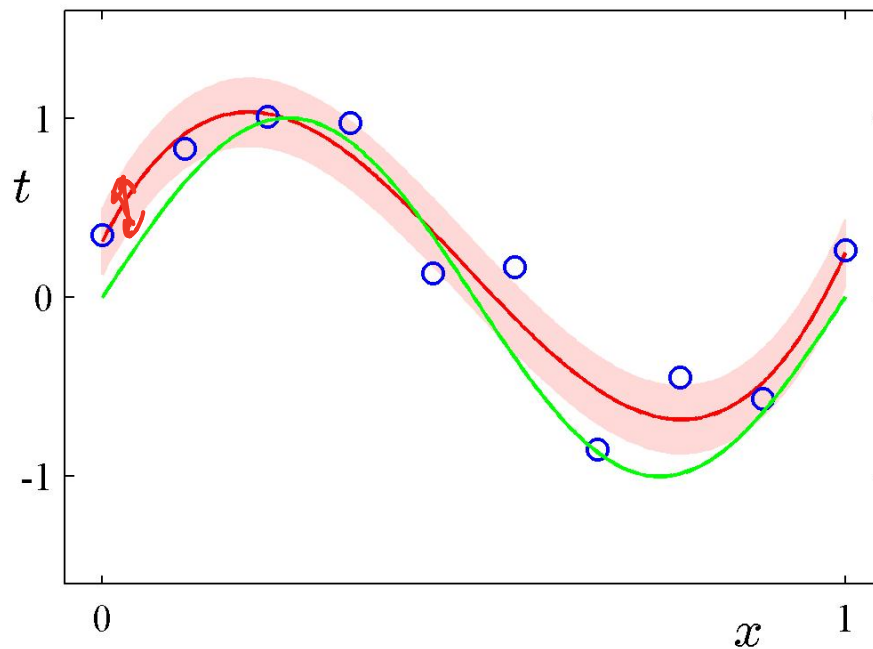
$$\frac{1}{\beta_{\text{ML}}} = \frac{1}{N} \sum_{n=1}^N \{y(x_n, \mathbf{w}_{\text{ML}}) - t_n\}^2. \quad (1.63)$$

Giving estimates of  $\mathbf{w}$  and  $\beta$ , we can predict

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

# Predictive Distribution

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



## MAP: A Step towards Bayes 1.2.5

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

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

*posterior Likelihood*

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

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

Regularized sum of squares

# Probability Theory

## Joint Probability

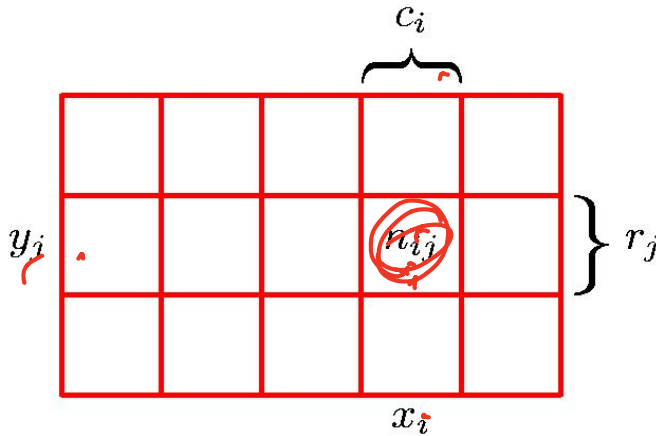
$$p(\underline{X} = x_i, Y = \underline{y_j}) = \frac{\underline{n_{ij}}}{N}$$

## Marginal Probability

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

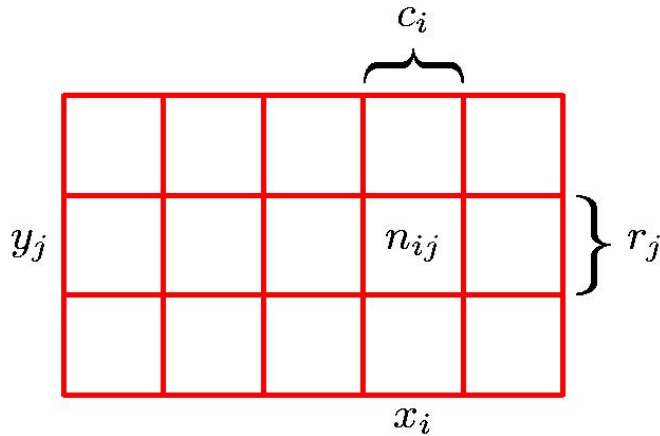
## Conditional Probability

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





# Probability Theory



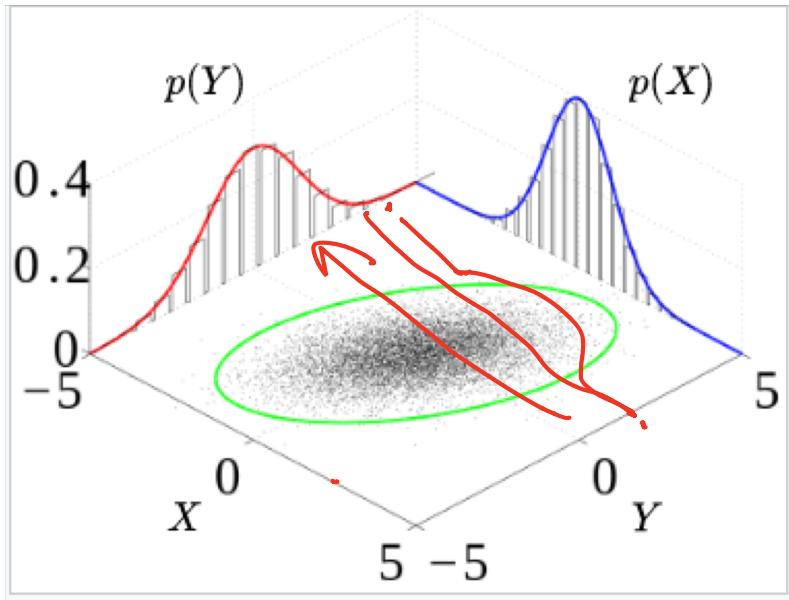
## •Sum Rule

$$\begin{aligned} p(X = x_i) &= \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^L n_{ij} \\ &= \sum_{j=1}^L p(X = x_i, Y = y_j) \end{aligned}$$

## 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

$$p(x, y)$$

Marginal Probability

$$p(y) = \int p(x, y) dx$$

Conditional Probability

$$p(x|y)$$

# 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

$$\begin{aligned} p(x, y) &= p(x|y) p(y) \\ &= p(y|x) p(x) \end{aligned}$$

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

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

posterior  $\propto$  likelihood  $\times$  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.