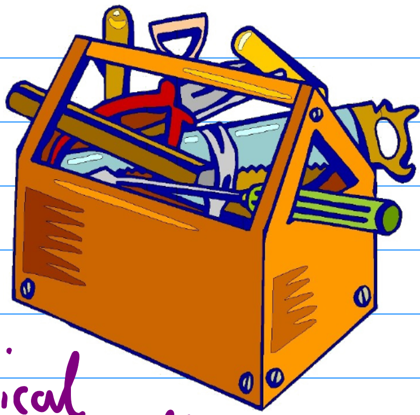
Lecture 28: Perspectives on Data Science

We saw a plethora of algorithms/techniques for extraction of useful information from data.

* your toolbox has now grown significantly!



* You are also equipped with minimum knowledge to read technical papers and books to continue your education.

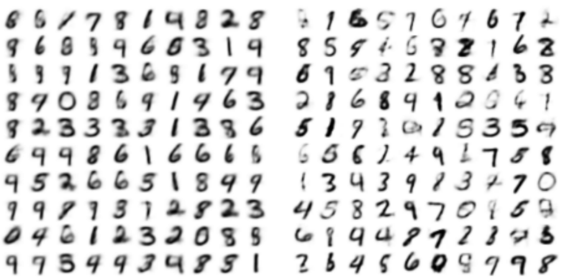
Conditional Sampling With Monotone GANs

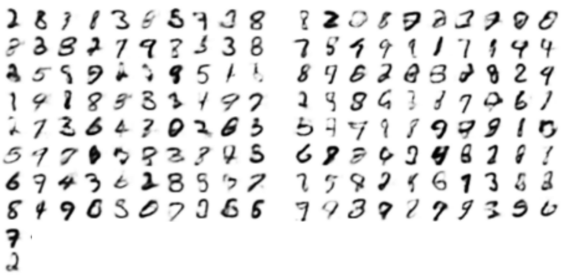
Nikola B. Kovachki[1] Ricardo Baptista[2] Bamdad Hosseini[1] Youssef Marzouk[2]

IEEE JOURNAL ON SELECTED AREAS IN INFORMATION THEORY, VOL. 1, NO. 1, MAY 2020

Deep Learning Techniques for Inverse Problems in Imaging

Gregory Ongie, Ajil Jalal, Christopher A. Metzler, Richard G. Baraniuk, Alexandros G. Dimakis, and Rebecca Willett





Auto-Encoding Variational Bayes

Diederik P. Kingma
Machine Learning Group
Universiteit van Amsterdam
dpkingma@gmail.com

Max Welling
Machine Learning Group
Universiteit van Amsterdam
welling.max@gmail.com

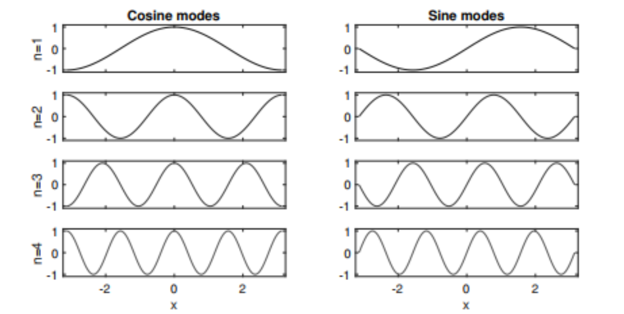
NIPS 2016 Tutorial: Generative Adversarial Networks

Ian Goodfellow

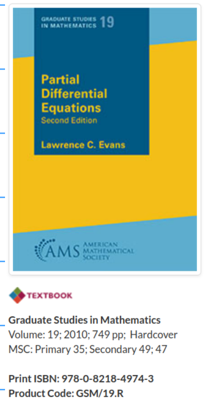
OpenAI, ian@openai.com

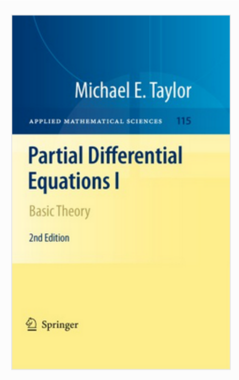
We started with classic techniques from applied mathematics

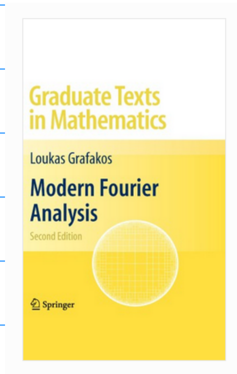
$$f(t) = \sum_k c_k \exp(i\pi k t)$$

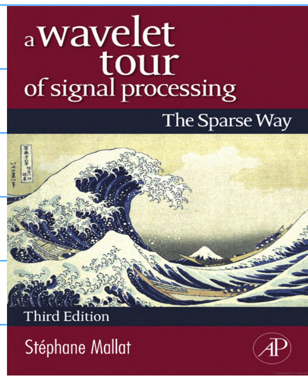


3.1 The first four Fourier sine and cosine modes.







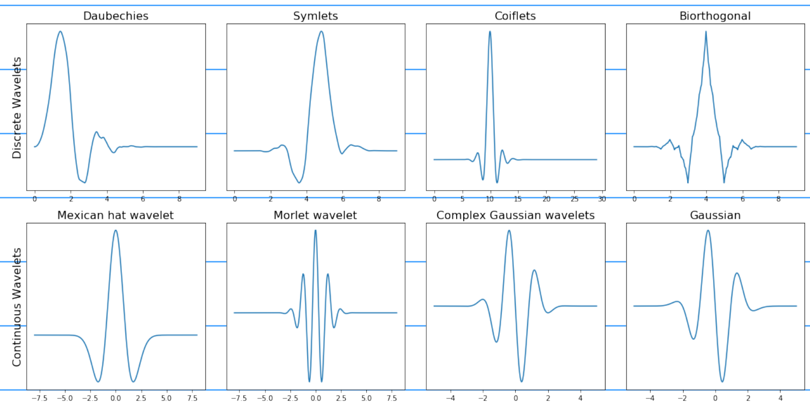


The workhorse of Fourier analysis (& much of signal & image processing) is the Fast Fourier transform (FFT) that computes the discrete Fourier transform (DFT) of a signal efficiently.

$$c_k \approx \underbrace{\frac{1}{N}\sum_{n=0}^{N-1} f(x_n)\exp\left(-2\pi i \frac{kn}{N}\right)}_{\text{DFT}}$$

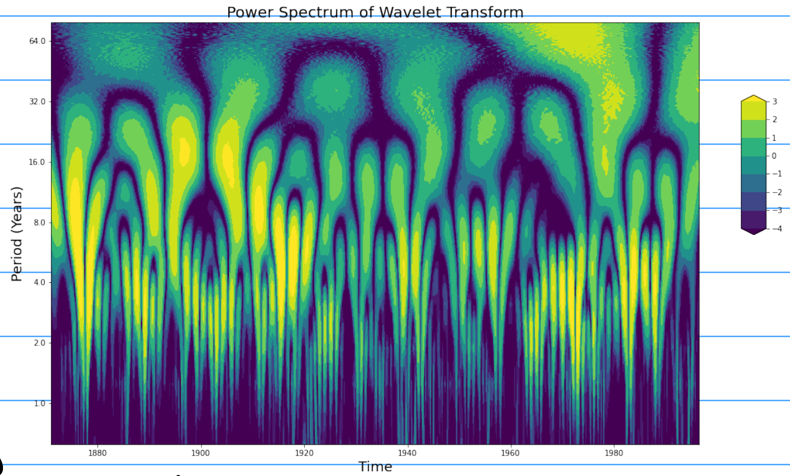
We then considered more modern decompositions.

The wavelets.



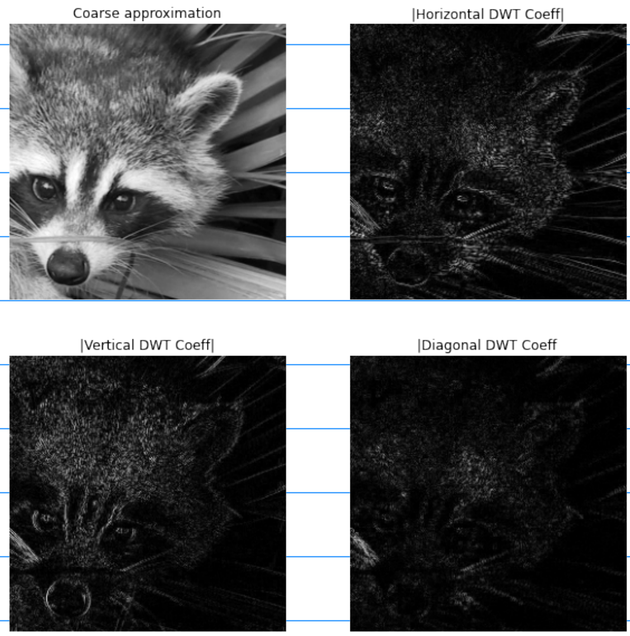
$$f(t) = \sum_{m,n} c_{m,n} \psi_{m,n}(t)$$

* localization in both scale(frequency) & location(time).

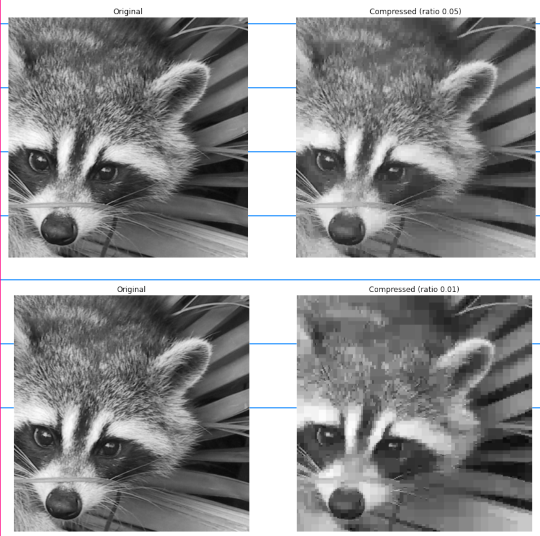


Wavelets afford us a lot of flexibility in analysis of signals/images.

The transform itself is very insightful.



* we also saw applications in compression & filtering.



In light of kernel methods we then realized that most signal processing techniques such as Fourier, Cosine transform, Wavelets etc. are essentially explicit constructions of feature maps

$$f(\underline{x}) = \sum_{j=0}^{J-1} c_j F_j(\underline{x})$$

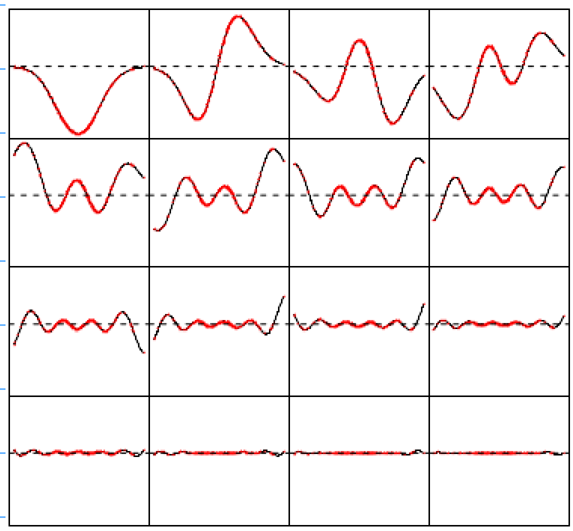
Mercers' Thm

If $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is an NDS kernel then

$$K(\underline{x}, \underline{x}') = \sum_{j=0}^{\infty} F_j(\underline{x}) F_j(\underline{x}')$$

for appropriate features $F_j : \mathbb{R}^d \to \mathbb{R}$.

features of the Gaussian kernel



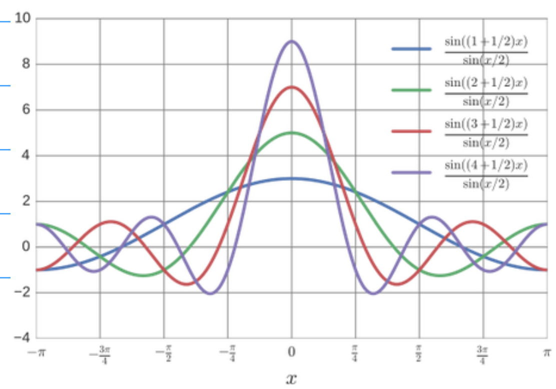
In this light we defined the Dirichlet (Fourier kernel)

$$K_N^F(x, x') = \frac{\sin((N+1/2)(x-x'))}{2\pi \sin((x-x')/2)} = \sum_{k=-N/2}^{N/2} \exp(ikx)\exp(-ikx')$$

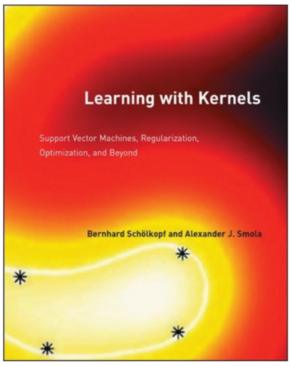
& the implicit wavelet kernel

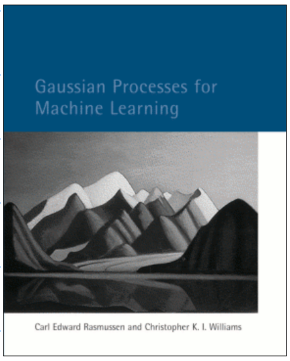
$$K_N^W(x, x') = \sum_{j=0}^{N-1} \psi_j(x) \psi_j(x')$$

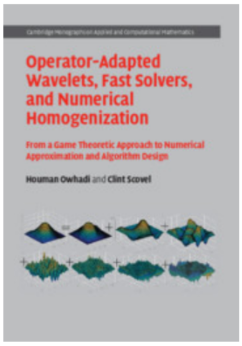
$$\psi_{m,n} \xrightarrow{\text{reorder}} \psi_j$$

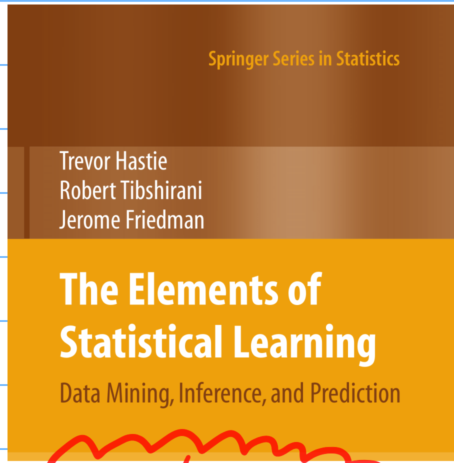


The Kernel methods were a unifying idea for us, allowing us to view classic methods of applied mathematics, statistics & machine learning from the same perspective.









Chapter 5.8

arXiv > stat > arXiv:1605.09522

Statistics > Machine Learning

[Submitted on 31 May 2016 (v1), last revised 13 Dec 2020 (this version, v4)]

Kernel Mean Embedding of Distributions: A Review and Beyond

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, Bernhard Schölkopf

arXiv > cs > arXiv:2012.00152

Computer Science > Machine Learning

[Submitted on 30 Nov 2020]

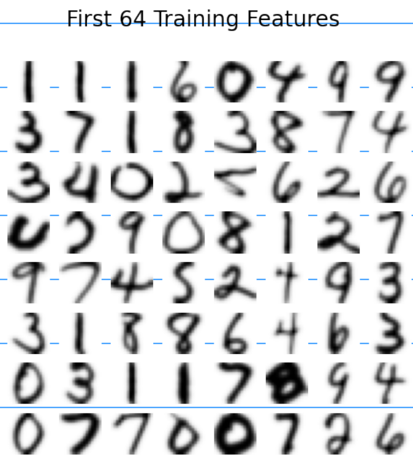
Every Model Learned by Gradient Descent Is Approximately a Kernel Machine

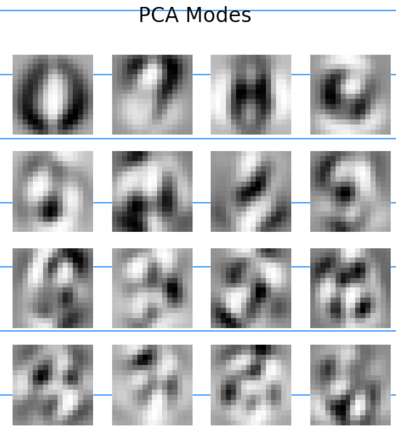
Pedro Domingos

egs

- PCA / DMD etc
- Most Ridge regression models
- Graph Laplacian embedding.
- Neural networks.

Principal Component analysis constructs the feature maps F_j from the data itself!





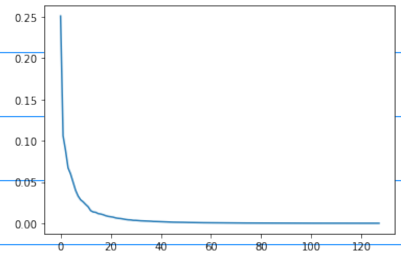
treat as kernel matrix

$$C_X = \frac{1}{N-1} X X^T$$

$$= Q \Lambda Q^T$$

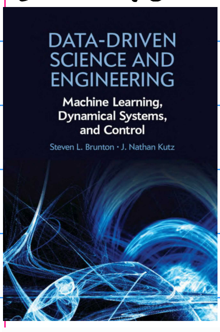
$$= \sum_j \lambda_j q_j q_j^T \equiv \sum_j F_j F_j^T$$

where $F_j = \sqrt{\lambda_j}\, q_j$ are the features.



* in the case of MNIST think of F_j as a function assigning a number to each pixel in image!

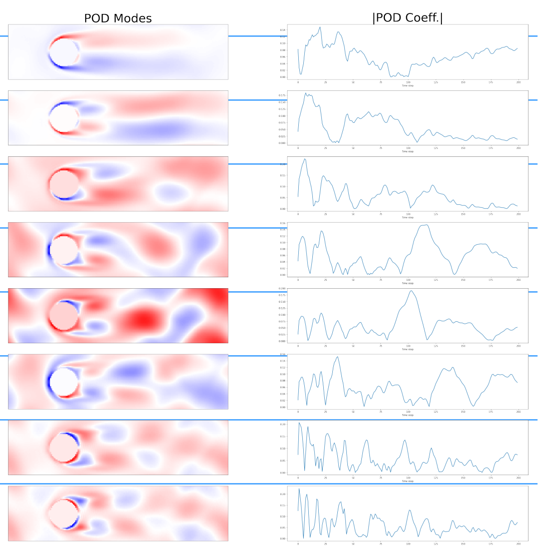
Often PCA/SVD based methods have similar interpretations ex POD.



SIAM Review
Vol. 49, No. 2, pp. 277–299

Error Estimation for Reduced-Order Models of Dynamical Systems*

Chris Homescu[†]
Linda R. Petzold[‡]
Radu Serban[§]



Published: 30 March 2006

Statistical properties of kernel principal component analysis

Gilles Blanchard, Olivier Bousquet & Laurent Zwald

Machine Learning **66**, 259–294 (2007) | Cite this article

The Theoretical Foundation of Reduced Basis Methods

Ronald DeVore *

January 9, 2014

Graph Laplacian embeddings are another example of data-driven features that are particularly useful for clustering/semi-supervised learning.

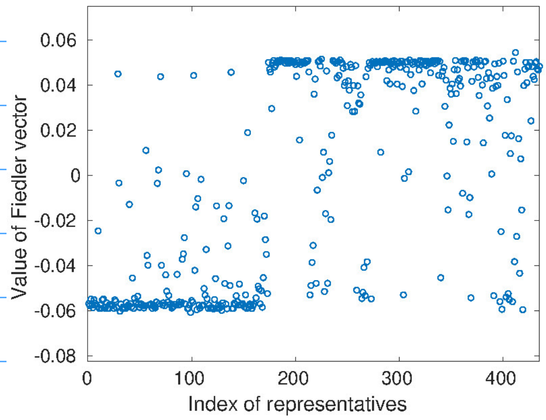
$G = \{X, W\}$, $L = D - W$

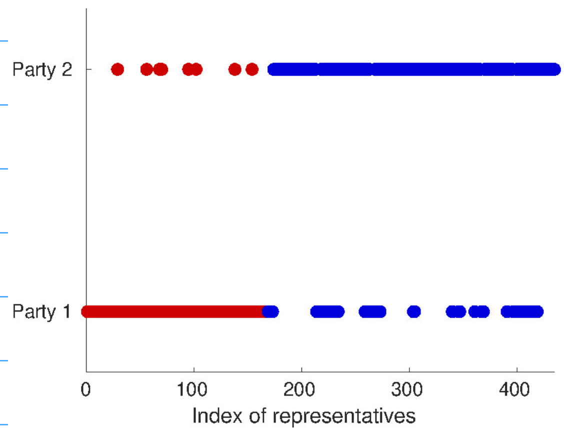
Solve eig. val. prob.

$$L = Q \Lambda Q^T$$

treat L^{-1} as kernel matrix

$$= \sum_j \lambda_j \underline{q}_j \underline{q}_j^T$$





Take feature maps

$\underline{x}_i \in X$ $\quad F_j(\underline{x}_i) = q_{ji} \quad j = 0, \dots, J-1$

with corresponding kernel

$$K(\underline{x}_i, \underline{x}_k) = \sum_{j=0}^{J-1} F_j(\underline{x}_i) F_j(\underline{x}_k) = \sum_{j=0}^{J-1} q_{ji} \cdot q_{jk}$$

$K: X \times X \to \mathbb{R}$

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Appl. Comput. Harmon. Anal. 21 (2006) 5–30

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Diffusion maps

Ronald R. Coifman *, Stéphane Lafon [1]

Mathematics Department, Yale University, New Haven, CT 06520, USA

Received 29 October 2004; revised 19 March 2006; accepted 2 April 2006

Available online 19 June 2006

Communicated by the Editors

THE LAPLACIAN SPECTRUM OF GRAPHS †

Bojan Mohar ‡

Department of Mathematics

University of Ljubljana

Jadranska 19, 61111 Ljubljana

Yugoslavia

Algorithms for manifold learning

Lawrence Cayton

lcayton@cs.ucsd.edu

June 15, 2005

48th Annual IEEE Symposium on Foundations of Computer Science

Spectral Graph Theory and its Applications

Daniel A. Spielman

Department of Computer Science and

Program in Applied Mathematics

Yale University

spielman@cs.yale.edu

Feed Forward Neural Nets (FNNs) can be viewed as parameterized / Compositional kernel methods

$$f(\underline{x}) = F^{(L)}(F^{(L-1)}(\cdots(F^{(1)}(\underline{x}))\cdots))$$

write $F^{(L)}(\underline{z}) = \sum_j \beta_j^{(L)} F_j^{(L)}(\underline{z})$ & $\phi^{(L-1)}(\underline{x}) := F^{(L-1)}(\cdots(F^{(1)}(\underline{x}))$

Then $$f(\underline{x}) = \sum_j \beta_j^{(L)} \underbrace{F_j^{(L)}(\phi^{(L-1)}(\underline{x}))}_{\text{features}}$$

Then we can view f as a function in the RKHS of $K^{(L)}(\phi^{(L-1)}(\underline{x}), \phi^{(L-1)}(\underline{x}'))$ where $K^{(L)}(\underline{z}, \underline{z}') = \sum_j F_j^{(L)}(\underline{z}) F_j^{(L)}(\underline{z}')$.

we can repeat the same argument for $F^{(L-1)}$ & realize a FNN as compositional model where each layer $F^{(\ell)}$ is a function in an RKHS that parameterizes the kernel of the next layer.

arXiv > stat > arXiv:2008.03920

Statistics > Machine Learning

[Submitted on 10 Aug 2020 (v1), last revised 2 Nov 2020 (this version, v2)]

Do ideas have shape? Plato's theory of forms as the continuous limit of artificial neural networks

Houman Owhadi

Journal of Machine Learning Research 20 (2019) 1-32 Submitted 10/17; Revised 6/18; Published 4/19

A Representer Theorem for Deep Kernel Learning

Bastian Bohn BOHN@INS.UNI-BONN.DE
Christian Rieger RIEGER@INS.UNI-BONN.DE
Institute for Numerical Simulation, University of Bonn
Endenicher Allee 19b, 53115 Bonn, Germany
Michael Griebel GRIEBEL@INS.UNI-BONN.DE
Institute for Numerical Simulation, University of Bonn
Endenicher Allee 19b, 53115 Bonn, Germany,
Fraunhofer Center for Machine Learning
Fraunhofer Institute for Algorithms and Scientific Computing SCAI
Schloss Birlinghoven, 53754 Sankt Augustin, Germany

Deep Gaussian Processes

Andreas Damianou, Neil D. Lawrence *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, PMLR 31:207-215, 2013.*

Kernels for Vector-Valued Functions: A Review

Publisher: Now Foundations and Trends Cite This PDF

Mauricio A. Álvarez; Lorenzo Rosasco; Neil D. Lawrence **All Authors**

Neural Ordinary Differential Equations

Part of Advances in Neural Information Processing Systems 31 (NeurIPS 2018)

Bibtex Metadata Paper Reviews Supplemental

Authors

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David K. Duvenaud

Neural Tangent Kernel: Convergence and Generalization in Neural Networks

Part of Advances in Neural Information Processing Systems 31 (NeurIPS 2018)

Bibtex Metadata Paper Reviews Supplemental

Authors

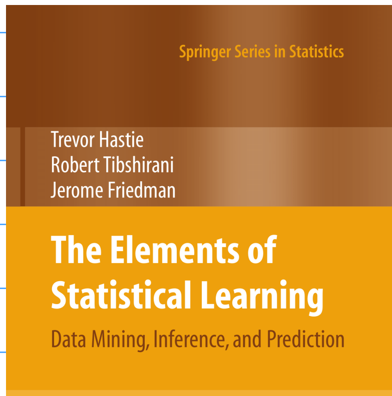
Arthur Jacot, Franck Gabriel, Clement Hongler

- More Hot topics / keywords to Search for:
  - Representation learning
  - Autoencoders
  - Normalizing flows
  - Reinforcement learning
  - Generative Adversarial Networks
  - Physics Informed Learning
  - Active learning
  - Meta learning
- Be responsible

A Survey on Bias and Fairness in Machine Learning

NINAREH MEHRABI, FRED MORSTATTER, NRIPSUTA SAXENA,
KRISTINA LERMAN, and ARAM GALSTYAN, USC-ISI

- If you want one book on your shelf



- If you are a math student thinking about research in data Science take:
  - functional analysis
  - Linear analysis
  - Probability theory
  - Real analysis