Data Driven Method From A Math Perspective

Host: *Yiping Lu(Peking University)* http://about.2prime.cn

Some Useful Links:

http://ttic.uchicago.edu/~nati/Teaching/TTIC31120/2016/
http://www2.isye.gatech.edu/~tzhao80/Lectures/8803.pdf

http://math.stanford.edu/~ryzhik/STANFORD/MEAN-FIELD-GAMES/notes-mean-field.pdf

https://bguedj.github.io/nips2017/50shadesbayesian.html

http://deepbayes.ru/

http://proceedings.mlr.press/

http://openaccess.thecvf.com

Preliminary

Linear Algebra, Calculus

Basic Probability(characteristic function!), Optimization(Reference:http://bicmr.pku.edu.cn/~wenzw/opt-2018-fall.html lecture10,11,12)

Practicle Algorithms in deep learning(Reference:CS231n http://cs231n.stanford.edu/syllabus.html lecture1,2,3,4,5)

Abstract

A machine and deep learning tutorial from both **Statistic** and **Computational Math** perspective . The introduction is separate into two parts, one is basic machine learning knowledge(un-biased) and another is advanced topics(Biased).

Outline

Basic Machine Learning

PAC Learning Framework

Chernoff Bound

- chernoff bound and concerntration inequality(Hoeffding's inequalilty, McDiarmid's Inequality, Bernstein's inequality...)
- Pan Zhou, Jiashi Feng "Understanding generalization and optimization performance of deep CNNs"ICML2018

Rademacher Framework and VC Dimension

- Double sampel trick and symmetric trick
- On Tighter Generalization Bound for Deep Neural Networks: CNNs, ResNets, and Beyond Xingguo Li, Junwei Lu, Zhaoran Wang, Jarvis Haupt and Tuo Zhao arxiv1806.05159
- Neyshabur, Behnam, et al. "Toward Understading the Role of Over-Parametrization in Generalization of Neural Networks" arXiv:1805.12076
- Approximability of Discriminators Implies Diversity in GANsY Bai, T Ma, A Risteski arXiv preprint arXiv:1806.10586
- On the Discrimination-Generalization Tradeoff in GANs arxiv:1711.02771(ICLR2018)
- Dropout Training, Data-dependent Regularization, and Generalization Bounds Wenlong Mou, Yuchen Zhou, Jun Gao, Liwei Wang ; PMLR 80:3642-3650(ICML2018)

Algorithm Stability

- Application in machine learning
- Application in optimization
    - Stability and Convergence Trade-off of Iterative Optimization AlgorithmsY Chen, C Jin, B YuarXiv preprint arXiv:1804.01619
    - Train faster, generalize better: Stability of stochastic gradient descentM Hardt, B Recht, Y Singer - arXiv preprint arXiv:1509.01240, 2015 - arxiv.org(ICML2016)

PAC Bayesian Theory

- https://bguedj.github.io/nips2017/pdf/laviolette_nips2017.pdf
- PAC Bayesian thoery and application in SVM bound(we also can talk about "SVM=lasso")
- PAC Bayesian thoery and flat minima(Neyshabur, Behnam, et al. "Exploring generalization in deep learning." NIPS2017.)
- PAC Bayesian thoery and SGLD
- A PAC-Bayesian Tutorial with A Dropout Bound(arXiv:1307.2118)
- pac-bayesian theory meets bayesian inference(TBD arxiv:1605.08636 NIPS2016)

covering number(TBD)

kernel method(TBD)

Take semi-supervised learning as an example

(Reference: A PAC-style model for learning form labeled and unlabeled data. /Semi-supervised learning by higher order regularization.)

- To Understand Deep Learning We Need to Understand Kernel Learning(ICML2018,TBD)
- Toward deeper understanding of neural networks: The power of initialization and a dual view on expressivity(NIPS16,TBD)

Other Bounds

- Stronger generalization bounds for deep nets via a compression approachS Arora, R Ge, B Neyshabur, Y Zhang - arXiv preprint arXiv:1802.05296, 2018 - arxiv.org
- Implicit regularization in deep learning B Neyshabur PhD Thesis(TBD)

Online learning

AdaBoost and online learning

Margin Theory For Boosting

Data Privacy

Differential Privacy

- Laplacian Mechanics
- BLR Mechanics

Differential Privacy And Entropy SGD(TBD)

- Entropy-SGD optimizes the prior of a PAC-Bayes bound: Generalization properties of Entropy-SGD and data-dependent priors arXiv:1712.09376(ICML2018)

Data-dependent PAC-Bayes priors via differential privacy(arXiv:1802.09583 TBD)

Bayesian Methods

http://deepbayes.ru/

Variational Inference

TBD

Generative Models

TBD,

Normalizing Flow: http://akosiorek.github.io/ml/2018/04/03/norm_flows.html

Bayesian Filtering

Optimization

Basic Optimization Method

Bottou L, Curtis F E, Nocedal J. Optimization Methods for Large-Scale Machine Learning[J]. 2016.

Differential Equation View Of Momentum Method

Deep Relaxation: partial differential equations for optimizing deep neural networks. arXiv:1704.04932v2

Stochastic modified equations and adaptive stochastic gradient algorithms. arXiv:1511.06251

A general analysis of the convergence of ADMM. arXiv:1507.06970

A Lyapunov Analysis of Momentum Methods in Optimization. arXiv:1611.02635

A Variational Perspective on Accelerated Methods in Optimization. arXiv:1603.04245

Asynchronous Coordinate Descent under More Realistic Assumptions. arXiv:1705.08494

Adaptive Methods

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

S. J. Reddi, S. Kale, and S. Kumar. On the convergence of adam and beyond. In International Conference on Learning Representations, 2018.

On the Convergence of A Class of Adam-Type Algorithms for Non-Convex Optimization, Xiangyi Chen, Sijia Liu, Ruoyu Sun, Mingyi Hong.

On the Convergence of AdaGrad with Momentum for Training Deep Neural Networks arXiv:1808.03408

Manifold Learning

Point Integral Method for Solving Poisson-type Equations on Manifolds from Point Clouds with Convergence Guarantees https://arxiv.org/pdf/1409.2623.pdf

A METHOD BASED ON TOTAL VARIATION FOR NETWORK MODULARITY OPTIMIZATION USING THE MBO SCHEME https://arxiv.org/pdf/1304.4679.pdf

Low dimensional manifold model for image processing SIAM Journal on Imaging Sciences 2017

An MBO scheme on graphs for classification and image processing SIAM Journal on Imaging Sciences, 2013 - SIAM

Zhou, Xueyuan, and Mikhail Belkin. "Semi-supervised learning by higher order regularization." Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 2011.

A. Bertozzi and A. Flenner, Diffuse interface models of graphs for classification of high dimensional data, Multiscale Model. Simul., 10 (2012), pp. 1090–1118.

Natural Boundary Conditions for Smoothing in Geometry Processing ACM Transaction on Graphics 2018 http://www.cs.columbia.edu/cg/hessians/

Stochastic Block Models are a Discrete Surface Tension https://arxiv.org/pdf/1806.02485

Bayesian Semi-supervised Learning with Graph Gaussian Processes https://arxiv.org/pdf/1809.04379v1.pdf

Local High-order Regularization on Data Manifolds https://arxiv.org/abs/1602.03805

Jacobs, Matt, Ekaterina Merkurjev, and Selim Esedoḡlu. "Auction dynamics: A volume constrained MBO scheme." Journal of Computational Physics 354 (2018): 288-310.

Balcan, Maria-Florina, and Avrim Blum. "A PAC-style model for learning from labeled and unlabeled data." International Conference on Computational Learning Theory. Springer, Berlin, Heidelberg, 2005.

Application In Computer Visioin And Recent Challenges

- **Low level:** Image Processing(Restoration),Image Inpainting / Image Completion, Image matting.
- **Middel Level:** Image Segementation, Optical Flow(Image registration).
- **High Level**: (Fine-grained) Classification, Object Detection.
- **Challenges**: Weak Supervision, Adversarial Examples, Video Understanding, Domain Adaptation, Irregular data structure.

Application In Nature Language Processing

TBD, Invited Speakers

Data Mining

Large-Scale Image Search, Transfer learning etc.

Adanced Topics

Optimal Transport

Villani, Cédric. Topics in optimal transportation. No. 58. American Mathematical Soc., 2003.

Jordan, Richard, David Kinderlehrer, and Felix Otto. "The variational formulation of the Fokker--Planck equation." SIAM journal on mathematical analysis 29.1 (1998): 1-17.

Peyré, Gabriel, and Marco Cuturi. Computational optimal transport. No. 2017-86. 2017. https://optimaltransport.github.io/

Solomon, Justin. "Optimal transport on discrete domains." AMS Short Course on Discrete Differential Geometry (2018).

Carlen, Eric A., and Wilfrid Gangbo. "Constrained steepest descent in the 2-Wasserstein metric." Annals of mathematics(2003): 807-846.

Chen, Yifan, and Wuchen Li. "Natural gradient in Wasserstein statistical manifold." *arXiv preprint arXiv:1805.08380* (2018).

Differential Equation And Deep Learning

Differential Equation View Of Deep Learning

slide: http://about.2prime.cn/slide/dynamicNN_slide.pdf

Yiping Lu, Aoxiao Zhong, Quanzheng Li, Bin Dong. "Beyond Finite Layer Neural Network:Bridging Deep Architects and Numerical Differential Equations" Thirty-fifth International Conference on Machine Learning (ICML), 2018

Theory：

A mean-field optimal control formulation of deep learning",

Weinan E, Jiequn Han, Qianxiao Li,

*Solving PDEs using Neural Networks

"Solving high-dimensional partial differential equations using deep learning", Jiequn Han, Arnulf Jentzen, Weinan E,

Zichao long*, Yiping Lu*, Xianzhong Ma*, Bin Dong. "PDE-Net:Learning PDEs From Data",Thirty-fifth International Conference on Machine Learning (ICML), 2018(*equal contribution)

http://www.ajentzen.de/

Weinan E, Yu B. The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems[J]. Communications in Mathematics & Statistics, 2018, 6(1):1-12.

Ma C, Wang J, Weinan E ,. Model Reduction with Memory and the Machine Learning of Dynamical Systems[J]. 2018.

Control Theory

Control and mean field control

- PMP and value function(relation with reinforcement learning)
- Hamilton-Jacobi-Bellman Equation
- A Mean-Field Optimal Control Formulation of Deep LearningW E, J Han, Q LiarXiv preprint arXiv:1807.01083
- Li Q, Chen L, Tai C, et al. Maximum Principle Based Algorithms for Deep Learning[J]. Journal of Machine Learning Research, 2018, 18.

  Li Q, Hao S. An Optimal Control Approach to Deep Learning and Applications to Discrete-Weight Neural Networks. ICML2018.

Mean field game

3D Data Processing

Attention: Preliminary: Differential Geometry!!!

Shape Analysis

Geometry Procssing

Manifold CNN

Random Matrixs

https://arxiv.org/pdf/1004.0861.pdf Universality of Wigner random matrices: a survey of recent results

An Introduction to Matrix Concentration Inequalities. J. A. Tropp. arXiv:1501.01571

https://zhenyu-liao.github.io/

Reinforcement Learning