

Теоретический анализ подходов глубокого обучения

к.ф.-м.н. Кретов М.К. (МФТИ)

Голиков Е.А. (МФТИ)

Аннотация

Современные подходы глубокого обучения радикально увеличили эффективность систем по распознаванию изображений, звука, текстов. В то же время, не существует общепризнанной теории для объяснения этого успеха. Также отсутствует теория, которая позволяла бы предсказывать “удачные” архитектуры нейронных систем. В значительной мере разработка таких архитектур в глубоком обучении основывается на численных экспериментах и подборе методов оптимизации. Целью курса является знакомство слушателей с основными теоретическими подходами к анализу глубоких нейронных сетей и подготовка к исследовательской работе в этой области.

Программа курса

1. Вводная лекция.

Что такое глубокое обучение. Компоненты глубокого обучения. Парадоксы глубокого обучения. Основные направления анализа.

Литература:

Rene Vidal, Joan Bruna, Raja Giryes and Stefano Soatto. Mathematics of Deep Learning, 2017.

2а. Статистическая теория обучения.

Формализация задачи поиска алгоритма обучения. PAC-подход. Эмпирический риск. Оценки на обобщающую способность моделей. Конечное и бесконечное пространство гипотез. VC-размерность. PAC-Bayes фреймворк.

Литература:

Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar. Foundations of Machine Learning. MIT Press, 2012.

2б. Оценки на обобщающую способность для глубоких нейронных сетей.

Новые техники для доказательств оценок (covering numbers, неравенство McDiarmid'a). Применение PAC-Bayes фреймворка: оценки через нормы матриц весов. Compression фреймворк и соответствующие оценки. Примеры расчёты оценок на обобщающую способность.

Литература:

Behnam Neyshabur, Ryota Tomioka and Nathan Srebro. In Search of the Real Inductive Bias: On the Role of Implicit Regularization in Deep Learning, 2014.

Behnam Neyshabur, Ryota Tomioka, Ruslan Salakhutdinov and Nathan Srebro. Geometry of Optimization and Implicit Regularization in Deep Learning, 2017.

Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester and Nathan Srebro. Exploring Generalization in Deep Learning, 2017.

Behnam Neyshabur, Srinadh Bhojanapalli and Nathan Srebro. A PAC-Bayesian Approach to Spectrally-Normalized Margin Bounds for Neural Networks, 2017.

Sanjeev Arora, Rong Ge, Behnam Neyshabur and Yi Zhang. Stronger generalization bounds for deep nets via a compression approach, 2018.

Behnam Neyshabur, Zhiyuan Li, Srinadh Bhojanapalli, Yann LeCun and Nathan Srebro. Towards Understanding the Role of Over-Parametrization in Generalization of Neural Networks, 2018.

2в. Связь обобщающей способности с характеристиками поверхности функции потерь. Характеристики найденного минимума и связь с обобщающей способностью. “Широкие” минимумы и репараметризация.

Литература:

Sepp Hochreiter and Juergen Schmidhuber. Simplifying Neural Nets by Discovering Flat Minima, 1999.

3а. Ландшафт функции потерь нейронных сетей.

Глобальность локальных минимумов в линейных сетях. Отсутствие седловых точек в окрестности нуля в линейных ResNet'ах. Почти наверное отсутствие локальных минимумов в двухслойных ReLU-сетях. Модель спинового стекла. Удаление локальных минимумов из поверхности лосс-функции.

Литература:

Pierre Baldi and Kurt Hornik. Neural Networks and Principal Component Analysis, 1989.

Kenji Kawaguchi. Deep Learning without Poor Local Minima, 2016.

Haihao Lu and Kenji Kawaguchi. Depth Creates No Bad Local Minima, 2017.

Moritz Hardt and Tengyu Ma. Identity Matters in Deep Learning, 2016.

Daniel Soudry and Yair Carmon. No bad local minima: Data independent training error guarantees for multilayer neural networks, 2016.

Anna Choromanska, Mikael Henaff, Michael Mathieu, Gérard Ben Arous and Yann LeCun. The Loss Surfaces of Multilayer Networks, 2014.

Kenji Kawaguchi and Leslie Pack Kaelbling. Elimination of All Bad Local Minima in Deep Learning, 2019.

3б. Оптимизация нейронных сетей.

Градиентный спуск почти наверное не попадает в строгие седла. Зашумлённый градиентный спуск с высокой вероятностью сходится в локальный минимум из любой инициализации. Динамика градиентного спуска в линейных сетях. Градиентный спуск с высокой вероятностью сходится в глобальный минимум в избыточно параметризованных нейронных сетях с одним скрытым слоем. Обобщение результата на глубокие сети.

Литература:

Jason D. Lee, Max Simchowitz, Michael I. Jordan and Benjamin Recht. Gradient Descent Converges to Minimizers, 2016.

Rong Ge, Furong Huang, Chi Jin and Yang Yuan. Escaping From Saddle Points - Online Stochastic Gradient for Tensor Decomposition, 2015.

Andrew Saxe, James L. McClelland and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks, 2013.

Simon S. Du, Xiyu Zhai, Barnabas Poczos and Aarti Singh. Gradient Descent Provably Optimizes Over-parameterized Neural Networks, 2018.

Simon S. Du, Jason D. Lee, Haochuan Li, Liwei Wang and Xiyu Zhai. Gradient Descent Finds Global Minima of Deep Neural Networks, 2018.

Zeyuan Allen-Zhu, Yuanzhi Li and Zhao Song. A Convergence Theory for Deep Learning via Over-Parameterization, 2018.

4а. IB-принцип (“information bottleneck”).

Определение. Различные фазы обучения сети с точки зрения IB. Критика подхода.

Литература:

Naftali Tishby, Fernando C. Pereira and William Bialek. The Information Bottleneck Method, 2000.

Naftali Tishby and Noga Zaslavsky. Deep Learning and the Information Bottleneck Principle, 2015.

Ravid Shwartz-Ziv and Naftali Tishby. Opening the Black Box of Deep Neural Networks via Information, 2017.

Andrew Michael Saxe, Yamini Bansal, Joel Dapello, Madhu Advani, Artemy Kolchinsky, Brendan Daniel Tracey and David Daniel Cox. On the Information Bottleneck Theory of Deep Learning, 2018.

4б. Обучение представлений.

Разложение функции потерь в несколько слагаемых. “Распутанные” представления. Эксперимент с обучением на случайных метках. Критические периоды обучения нейронной сети. Расстояние между задачами.

Литература:

Alessandro Achille and Stefano Soatto. Emergence of Invariance and Disentanglement in Deep Representations, 2017.

Alessandro Achille, Matteo Rovere and Stefano Soatto. Critical Learning Periods in Deep Neural Networks, 2017.

Alessandro Achille, Glen Mbeng and Stefano Soatto. The Dynamics of Differential Learning I: Information-Dynamics and Task Reachability, 2018.

5. Прохождение сигнала через глубокие сети.

Необходимое условие возможности обучения в бесконечно широких бесконечно глубоких сетях. Нейронные сети с точки зрения случайных матриц.

Литература:

Ben Poole, Subhaneil Lahiri, Maithra Raghu, Jascha Sohl-Dickstein and Surya Ganguli. Exponential expressivity in deep neural networks through transient chaos, 2016.

Samuel S. Schoenholz, Justin Gilmer, Surya Ganguli and Jascha Sohl-Dickstein. Deep Information Propagation, 2016.

Samuel S. Schoenholz, Jeffrey Pennington and Jascha Sohl-Dickstein. A Correspondence Between Random Neural Networks and Statistical Field Theory, 2017.

Lechao Xiao, Yasaman Bahri, Jascha Sohl-Dickstein, Samuel S. Schoenholz and Jeffrey Pennington. Dynamical Isometry and a Mean Field Theory of CNNs: How to Train 10,000-Layer Vanilla Convolutional Neural Networks, 2018.

Minmin Chen, Jeffrey Pennington and Samuel S. Schoenholz. Dynamical Isometry and a Mean Field Theory of RNNs: Gating Enables Signal Propagation in Recurrent Neural Networks, 2018.