ESE 546, FALL 2020

MODULE 4 SUMMARY

SHEIL SARDA [SHEIL@SEAS]

Lecture outlines and key takeaways for Module 4

Shape of the energy landscape of neural networks (Chapter 12)

(1) Introduction
(2) Deep Linear Networks
  (a) Least squares solution
  (b) Projection of a vector onto a matrix
  (c) Back to deep linear networks
  (d) Critical points of B if A is fixed
  (e) Critical points of A if B is fixed
  (f) Critical points of (A, B)
  (g) If W is a critical point then it can be written as a projection of the least squares solution on the subspace spanned by some p eigenvectors of Σ
  (h) If W is the global minimum for a two-layer network then it is a projection of the solution for a single-layer network onto the subspace spanned by the top p eigenvectors of Σ
  (i) There are exponentially many saddle points for a two-layer network
  (j) No local minima in a deep linear network, all minima are global and not unique
  (k) All the previous results are true for multi-layer linear networks
(3) Extending the picture to deep networks

Key takeaways

- Hello

Generalization performance of machine learning models (Chapter 13)

(1) The PAC-Learning Model
  (a) PAC-learnable hypothesis class
  (b) Learning Monotone Boolean Formulae
(2) Concentration of Measure
  (a) Union Bound (or Boole's Inequality)
  (b) Weak Law, Strong Law and Central Limit Theorem
  (c) Chernoff Bound
(3) Uniform Convegence

(4) Vapnik-Chernovenkis (VC) dimension
    (a) Shattering a set of inputs
    (b) Bounds on the VC-dimension of deep neural networks

Key takeaways

- Hello

Variational Inference (Chapter 14)

(1) The model
(2) Some technical basics
    (a) Variational calculus
        (i) Picking the domain and objective in variational optimization
        (ii) Choosing a functional to measure the distance between q and p
    (b) Laplace approximation
    (c) Digging deeper into KL-Divergence
(3) Evidence Lower Bound (ELBO)
    (a) Parameterizing ELBO
(4) Gradient of the ELBO
    (a) The Reparameterization Trick
    (b) Score function estimator of the gradient
    (c) Gradient of the remaining terms in ELBO
(5) Some comments

Key takeaways

- Hello

Generative Adversarial Networks (Chapter 15)

(1) Two-sample tests and Discriminators
(2) Building the Discriminator in a GAN
(3) Building the Generator of a GAN
(4) Putting the discriminator and generator together
    (a) Training a GAN
    (b) Solving min-max problems is difficult
    (c) A harsh discriminator inhibits the training of the generator
(5) How to perform valiation for a GAN
(6) The zoo of GANs

Key takeaways

- Hello

Table of lecture and recitation topics:

Lecture	Topic
21	SGD III, Background on Principal Component Analysis
Rec 12	Generalization bound using uniform convergence
22	Deep Linear networks
Rec 13	Vignette: Deep reinforcement learning
23	Background: Information theory and Variational Inference
24	Auto-Encoders, ELBO I
Rec 14	Recap of info theory, variational auto-encoders
25	ELBO II
26	Bayesian neural networks
27	Generative Adversarial Networks, Recap of post-midterm topics