

ESE 546, FALL 2020
MODULE 3 SUMMARY

SHEIL SARDA [SHEIL@SEAS]

Lecture outlines and key takeaways for Module 3

- (1) Stochastic Gradient Descent (Chapter 11)
 - (a) SGD for least-squares regression
 - (b) Convergence of SGD
 - (i) Strongly convex functions
 - (ii) What is the appropriate notion of convergence?
 - (iii) Descent Lemma for SGD
 - (iv) Typical assumptions in the analysis of SGD
 - (A) Stochastic gradients are unbiased
 - (B) Second moment of gradient norm does not grow too quickly
 - (v) Descent Lemma for SGD with additional assumptions
 - (vi) Convergence rate of SGD for strongly convex functions
 - (vii) Optimality gap for SGD (Theorem)
 - (viii) Heuristic for training neural networks
 - (ix) Convergence rate of SGD for decaying step-size (Theorem)
 - (x) Convergence rate for mini-batch SGD
 - (xi) When should one use SGD in place of Gradient Descent?
 - (c) Accelerating SGD using momentum
 - (i) Polyak-Ruppert averaging
 - (ii) Momentum methods do not accelerate SGD
 - (iii) Why do we use Nesterov's method to train neural networks?
 - (d) Understanding SGD as a Markov Chain
 - (i) Gradient flow
 - (ii) Markov chains
 - (iii) Invariant distribution of a Markov chain
 - (iv) Time spent at a particular state by the Markov chain
 - (v) A Markov chain model of SGD
 - (A) Transition probability of SGD
 - (B) Variance of SGD weight updates
 - (C) SGD is like GD with Gaussian noise
 - (vi) The Gibbs distribution

- (vii) Convergence of a Markov chain to its invariant distribution
- (viii) KL Divergence monotonically decreases (Lemma)

Key takeaways:

- A

- (2) Shape of the energy landscape of neural networks (Chapter 12)
 - (a) Introduction
 - (b) Deep Linear Networks
 - (c) Extending the picture to deep networks

Key takeaways:

- A

- (3) Object Detection & Image Segmentation (Recitation 11)
 - (a) Object Detection Framework
 - (b) R-CNN
 - (c) Fast R-CNN
 - (d) Faster R-CNN
 - (e) You Only Look Once (YOLO)
 - (f) Image Segmentation
 - (g) Mask R-CNN

Key takeaways:

- A

- (4) Generalization Bounds (Recitation 10)
 - (a) PAC learning
 - (b) Union and Chernoff Bounds
 - (i) Union Bound (or Boole's Inequality)
 - (ii) Measure Concentration Inequalities
 - (c) Generalization Bounds and Uniform convergence
 - (d) VC-dimension

Key takeaways:

- A

Table of lecture and recitation topics:

Lecture	Topic
18	Stochastic Gradient Descent I
19	Stochastic Gradient Descent II
Rec 11	Vignette: Object Detection
20	Markov Chains
Rec 12	Generalization Bounds