Machine Learning
(Quick Reference)

Parameswaran Raman • 2018 • UC Santa Cruz

Last Revision: September 17, 2018

Table of Contents

Abstract

I prepared these notes as a resource to help myself; and anyone else interested, in quickly reviewing concepts in this course. My goal is to provide a very concise, end-to-end resource that covers all the important material discussed in a course on this topic. If you spot any errors or have any suggestions, please contact me directly at params@ucsc.edu.

Thanks for reading!

1 Probability basics

Refer to probability notes.

2 Matrix Algebra basics

- various types of vector and matrix norms
- sparse representation of vectors and matrices
- gradient, hessian, jacobian of matrices
- positive definite, negative definite properties and eigen-values, eigen-vectors, rank
- Why low-rank matters? Or why is it important?
- condition # of matrices and degeneracy conditions
- review types of factorization ($LU, PLU, LDU, LL^T, QR, SVD$) and their computational complexity
- how QR helps in least-squares problems?
- quadratic problem form (gradient, hessian for it)

(source: skim through appendix of griva-sofer-nash, nocedal or PRML)

3 Optimization

- intuition behind lagrange multipliers?
- Taylor series, power series, infinite series sums
- constrained and unconstrained optimization
- non-linear optimization methods (gradient descent and first-order optimality conditions, conjugate gradient, newton methods, quasi-newton and L-BFGS, augmented lagrangian, penalty methods, SGD)
- wolfe/armijo rules for line search
- strong convexity, lipschitz conditions
- KKT conditions
- gradient descent, mirror descent, projected gradient descent, sub-gradients
- types of convergence (linear, quadratic, super-linear, etc), convergence rates of various optimization methods

(source: skim through griva-sofer-nash, keerthi's slides/writeup, nocedal or gleich notes)

4 Bayesian Inference / Graphical Models

- AMS 203 notes, exp family, conjugate distributions, frequently used probability distributions, stats-cookbook
- generative vs discriminative models (naive bayes vs logistic regression, HMM vs CRFs), reference (LINK)
- mixture models (readup some basic version like LDA), variational inference brush up (read your presentation slides, notes of derivations, etc), distributed-gibbs-for-LDA (LINK), exchangeability, de-finetti?s theorem
- approximate inference: mcmc vs variational inference, Metropolis-Hastings, Gibbs
- non-parametric bayesian models (skim through your course notes)

5 General Machine Learning

- linear regression (objective function, closed form solution, pseudo-inverse, various ways of computing inverse of a matrix)
- linear regression vs PCA
- logistic regression (objective function for both binary and multi-class, derivations)
- svm?s (primal and dual formulations, kernel tricks, scaling issues, optimization, smo) - watch keerthi?s video, look at: connection b/w soft-margin formulation and hinge-loss based regularized risk minimization formulation.
- k-means
- EM algorithm (short intro)
- evaluation metrics (precision, recall, f-score (macro and micro), accuracy, AUC, NDCG)
- bias variance decomposition, bias-variance tradeoff, relationships to underfitting and overfitting
- recommender systems (common ways to build recommender systems? see andrew ng ml class notes)
- matrix factorization stuff (svd, low rank approximations, nmf, how to formulate recommender problems as matrix factorization, how to solve them using als or sgd, or others? parallelization?)
- learning to rank formulations
- information theory basics (mutual information, entropy, relative entropy (ie: KL divergence), bregman divergence, its diagram and its properties). see wiki for bregman divergence intro and inderjit's talk for details.
- types of updates: GD vs EG, implicit vs explicit, batch vs stochastic
- role of bias term? https://www.quora.com/Why-do-we-need-the-bias-term-in-ML-algorithms-such-as-linear-regression-and-neural-networks
- deep learning (brush up the diff types of deep learning models)
- what are decision trees, random forests, gradient boosted decision trees (differences b/w these and advantages of GBDT?)
- bagging and boosting methods - various types of doing bagging, AdaBoost

- pros and cons of various ml models (eg: advantages of decision trees, naive bayes, etc in practical applications)

(source: skim through PRML book)

6 Random Projections / Learning Theory

- VC dimensions, etc (look up a course / textbook and read important topics in Theoretical ML)
- Important inequalities and theorems (Markovs, Chebychev, Hoeffding, Johnson Lindenstass)
- Sketching, Leverage Scores and see notes from Randomized Numerical Lin Algebra course (mahoney or dreinas)
- Similarity Search stuff (LSH, distance measures, hashing) - read papers / talks on this topic
- Page Rank, Random Walks
- How a lot of these techniques in approximation algorithms are used in efficient distance computations / NN search / graph algorithms (see Sesh?s work), these are especially useful for Twitter like companies

7 Deep Learning

- Brush up types of DNN architectures
  - Auto-encoder decoders, Deep Auto-encoders
  - Probabilistic Deep Learning models (Variational Auto-Encoder, GANs)
  - ResNets
  - Boltzman Machines
  - CNN
  - RNN, LSTMs, GRU, Attention Mechanism
- How to do the following tasks using DNNs?
  - Deep Recommender Systems
  - Multiclass classification
  - Sequence to sequence (translation)
  - Named Entity Recognition
  - Simple Computer Vision (Image Detection / Video Detection) application
  - Dimensionality Reduction / creating embeddings
- Training
  - Tips and Tricks (momentum, adagrad, adam, early stopping, batch normalization, different validation techniques, types of activation functions and heuristics on when to use what)
  - Data, Model parallelism techniques
  - How to run on GPUs?
  - Familiarize with some libraries (TensorFlow, PyTorch, MXNet)

(source: skim through Deep Learning Stanford courses, DNN books)

8 Your Research Projects

bla bla

9 Appendix: Other random stuff

9.1 bla bla

10 Acknowledgements