Syllabus for CS6787

Advanced Machine Learning Systems

Term	Fall 2017	**Instructor**	Christopher De Sa
Course website	`cs.cornell.edu/courses/cs6787/`	**E-mail**	`cdesa@cs.cornell.edu`
Schedule	MW 7:30pm - 8:45pm	**Office hours**	W 2:00pm 3:00pm or by appointment
Room	Bill and Melinda Gates Hall G01	**Office**	Bill and Melinda Gates Hall 450

Description. So you've taken a machine learning class. You know the models people use to solve their problems. You know the algorithms they use for learning. You know how to evaluate the quality of their solutions.

But when we look at a large-scale machine learning application that is deployed in practice, it's not always exactly what you learned in class. Sure, the basic models, the basic algorithms are all there. But they're modified a bit, in a bunch of different ways, to run faster and more efficiently. And these modifications are really important—they often are what make the system tractable to run on the data it needs to process.

CS6787 is a graduate-level introduction to these system-focused aspects of machine learning, covering guiding principles and commonly used techniques for scaling up to large data sets. Informally, we will cover the techniques that lie between a standard machine learning course and an efficient systems implementation. Topics will include stochastic gradient descent, acceleration, variance reduction, methods for choosing metaparameters, parallelization within a chip and across a cluster, popular ML frameworks, and innovations in hardware architectures. An open-ended project in which students apply these techniques is a major part of the course.

Prerequisites. Knowledge of machine learning at the level of CS4780. Optionally, knowledge of computer systems and hardware on the level of CS 3410.

Format. One traditionally formatted lecture per week, and one presentation by a group of students of a selected paper, followed by discussion.

Grading. Students will be evaluated on the following basis.

20%	Paper presentation — each student will present, in a group, one of the listed papers
10%	In-class quizzes — there will be a quiz before each paper presentation on that paper's content
10%	Discussion participation
30%	Paper reviews — students must submit a review of every paper we discuss
30%	Final project

Final project parameters. The final project can be done in groups of up to three (although more work will be expected from groups with more people). The subject of the project is open-ended, but it must include the implementation of a machine learning system to solve a problem, using one or more of the techniques discussed in the course (or similar techniques) to achieve a speedup over some baseline method. The project will culminate in a project report of at least four-pages. Project proposals (at most one page) are due on **Monday, November 13**. An abstract for the report is due on **Monday, November 27**, and we will discuss the abstracts in class on that day. The final project report is due on **Wednesday, December 6**.

Course Calendar.

Wednesday, August 23	**No in-person lecture. I am traveling this week. Do not go to the lecture room. No one will be there.**
Monday, August 28	**Lecture 1.** Topics: • Overview • Course outline and syllabus • Gradient descent • Stochastic gradient descent: the workhorse of machine learning • Theory of SGD for convex objectives
Wednesday, August 30	**Due: Sign-up for paper presentations.**

Lecture 2. Topics:
- The effect of choosing the step size/learning rate
- Mini-batching and batch size
- Overfitting
- Generalization error
- Regularization
- Early stopping

Monday, September 4 — **Labor day. No lecture.**

Wednesday, September 6 — **Paper Discussion 1.** Rich Caruana, Steve Lawrence, and C Lee Giles. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Advances in neural information processing systems*, pages 402–408, 2001

Monday, September 11 — **Due: Review of Paper 1.**

Lecture 3. Topics:
- The condition number
- Momentum and acceleration
- Momentum for quadratic optimization
- Momentum for principle component analysis (PCA)

Wednesday, September 13 — **Paper Discussion 2.** Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013

Monday, September 18 — **Due: Review of Paper 2.**

Lecture 4. Topics:
- The kernel trick
- Gram matrix versus feature extraction: systems tradeoffs
- Adaptive/data-dependent feature mappings

Wednesday, September 20 — **Paper Discussion 3.** Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in neural information processing systems*, pages 1177–1184, 2007

Monday, September 25 — **Due: Review of Paper 3.**

Lecture 5. Topics:
- Online versus offline learning
- Variance reduction
- SVRG
- Fast linear rates for convex objectives

Wednesday, September 27 — **Paper Discussion 4.** Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in neural information processing systems*, pages 315–323, 2013

Monday, October 2 — **Due: Review of Paper 4.**

Lecture 6. Topics:
- Metaparameter optimization
- Assigning parameters from folklore
- Random search over parameters

Wednesday, October 4	**Paper Discussion 5.** James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb): 281–305, 2012
Monday, October 9	**Fall break. No lecture.**
Wednesday, October 11	**Due: Review of Paper 5.**
	Paper Discussion 6. Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012
Monday, October 16	**Due: Review of Paper 6.**
	Lecture 7. Topics: • Non-convex stochastic gradient descent • Weakness of theoretical guarantees • One case where we can say something: stochastic power iteration • Deep learning as non-convex optimization
Wednesday, October 18	**Paper Discussion 7.** Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014
Monday, October 23	**Due: Review of Paper 7.**
	Lecture 8. Topics: • Major bottleneck for ML systems: parallelism • Asynchronous execution • Hogwild!
Wednesday, October 25	**Paper Discussion 8.** Feng Niu, Benjamin Recht, Christopher Re, and Stephen Wright. Hogwild: A lock-free approach to parallelizing stochastic gradient descent. In *NIPS*, pages 693–701, 2011
Monday, October 30	**Due: Review of Paper 8.**
	Lecture 9. Topics: • Major bottleneck for ML systems: memory bandwidth and locality • Low precision computation • Vector computation • Scan orders
Wednesday, November 1	**Paper Discussion 9.** Suyog Gupta, Ankur Agrawal, Kailash Gopalakrishnan, and Pritish Narayanan. Deep learning with limited numerical precision. *ICML*, 2015
Monday, November 6	**Due: Review of Paper 9.**
	Lecture 10. Topics: • Algorithms other than SGD • What happens on the inference side? • Stochastic coordinate descent • Markov chain Monte Carlo and Gibbs sampling • Contrastive divergence • Derivative free optimization

Wednesday, November 8	**Paper Discussion 10.** Song Han, Huizi Mao, and William J Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *ICLR*, 2016
Monday, November 13	**Due: Review of Paper 10.** **Due: Final Project Proposal.** **Lecture 11.** Topics: • Hardware for machine learning • The dominance of GPUs • Accelerators for machine learning • Will all computation become matrix multiply?
Wednesday, November 15	**Paper Discussion 11.** Norman P Jouppi, Cliff Young, Nishant Patil, David Patterson, Gaurav Agrawal, Raminder Bajwa, Sarah Bates, Suresh Bhatia, Nan Boden, Al Borchers, et al. In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th Annual International Symposium on Computer Architecture*, pages 1–12. ACM, 2017
Monday, November 20	**Due: Review of Paper 11.** **Lecture 12.** Topics: • Machine learning frameworks and cluster parallelism • TensorFlow • SciKit-Learn • PyTorch • Is Python the ML language of the future?
Wednesday, November 22	**Thanksgiving break. No lecture.**
Monday, November 27	**Abstract for Final Project.** **Abstract swap and discussion.**
Wednesday, November 29	**Abstract swap continued and/or final lecture.** Depending on time and number of project groups. Topic for lecture will depend on student interest.
Wednesday, December 6	**Final Project report due.**