Key information

Course number: CS 269 SEM 2

Course Catalog Title: Advanced Topics in Machine Learning

Short Title: Advanced Topics in Machine Learning

Units: 4

Grading: Letter (A-F); auditing is not allowed

Format: Seminar (instructor might give a few lectures to set the stage/make up deficiencies)

GE or Major/Minor requirement: Major for the PhD Proposal of Field “Artificial Intelligence” or “Machine Learning”

Requisites: CS 276A/Stat 231 Pattern Recognition and Machine Learning or its equivalent, or the instructor’s approval

Description: CS 269 is a graduate-level seminar course. Its main purpose is to introduce to graduate students in machine learning and related areas with a selected subset of research topics in the frontiers of machine learning. Those research topics have not made into standard textbooks for machine learning, yet they represent state-of-the-art research results, in either basic or applied research. The course is intended to be offered at least once per academic year and its primary audience is doctoral students or advanced master/undergraduate students with adequate preparation. The topics could be adjusted from year to year, reflecting the rapid expansion of machine learning research.

Format: The course will be taught in instructor-led lectures and student-led discussion and presentations of original research articles.

Enrollment limit: 15

Classroom: BH 5422

Office hours: TBD

Preferred communication: Email to feisha@cs.ucla.edu. Include CS269 in your subject line to avoid delaying in processing your email.
Supplemental Information about the course

For the first offering in the winter quarter of 2016, this course will cover the following advanced topics in machine learning: algorithms and theoretical analysis of deep learning, kernel methods (and if necessary, related topics from computational statistics). Those topics are not only under active research by the machine learning community but also are of overwhelming success and great importance to many practical application areas. Specific topics include but not are not limited to: various deep learning architectures and associated parameter estimation techniques, understanding why deep learning architectures are effective, investigating its relation to other learning paradigms such as large-scale kernel methods, etc. The overarching theme is to investigate the concept of “representation learning” – the core arguments made by deep learning architectures — and various methods and algorithms for learning good representations from data.

Student Work and Grading Policy

The students will be exposed to state-of-the-art research in machine learning through paper reading and presentation, and be assessed in their critical analysis (including regular writeup/summary and class participation), presentation of those research papers, as well as a course project.

- Critical analysis: 25%
  - Each student is required to write independently a short survey on each of 4 research topics — three will be designated by the instructor and the fourth one is up to the student (as long as the selected topic is related to the course and does not repeat the other two). The survey should provide a good coverage of up-to-date research results (by citing articles published in recent conferences or on arxiv), a clear critical analysis of the literature, and form a basis for possible research ideas. A survey that compiles a list of papers and their abstracts is deemed inadequate. The survey should be concise as it merely studies a relatively narrower set of research articles. Thus, a length of no more than 5 pages (excluding the bibliography) is expected.
  - Each student is expected to contribute to class participation. During the participation, you are expected to raise scientific questions that are particularly pertinent to the papers being presented. Thus, you need to read the papers prior to their presentation — excessively asking clarifying questions (thus interrupting the presentation) is not appreciated.

- Presentation: 30%
  Each student is required to present 2 to 3 papers (depending on the number of the final enrollments) to the whole class. On average, each presentation lasts about 40 minutes, with additional 10 minutes for Q&A. The standard of the presentation is held at the professional level such as those in technical conferences. Note that the total length of the presentation is about 50 minutes, which is more than a typical conference presentation. Thus, you are expected to
give detailed and in-depth presentation of the paper. Hence, a high-level/“advertisement for coming to my poster” style presentation is not considered sufficient.

The presentation slides (or notes, if handwritten) are to be shared with the whole class.

- Course project: 45%

Each student is required to finish a course project by preparing a two-page project proposal, turning in a mid-term project report (in the form of a short presentation) and submitting a final project report. Group projects with up to 2-members per team are permitted.

The course project is evaluated as if it were a submission to a technical conference. Each project will be presented to the class at the due time of the mid-term project progress report.

As a rough guideline, a project at the level of being a convincing submission for top conference venues (NIPS, ICML, AISTATS, UAI, or their equivalent ones) will be given a grade of A+. A project that requires some work to qualify such submissions will be given a grade of A and a project that requires substantial followup work will be given a grade of A- or below.

Note that, any of the 4 surveys can be used as a basis for the project, though you are welcome to select a totally new one.

Required textbooks

There is no required textbooks. However, you are encouraged to refresh your general knowledge about machine learning using the following ones:

- C. Bishop’s *Pattern Recognition and Machine Learning*
- K. Murphy’s *Machine Learning: a Probabilistic Perspective*
- N. Friedman and D. Koller’s *Probabilistic Graphical Models*

Mathematical concepts and tools are extensively used throughout this class. You should be familiar with multivariate calculus, vector and matrix algebra. For applied mathematics, common ideas and algorithms for optimizations are a must.

As for deep learning, a good (though still work-in-progress) textbook can be found at [http://www.deeplearningbook.org/](http://www.deeplearningbook.org/).

Required background reading material as well as original research articles are listed in detail in the planned schedule. See below.

Tentative Schedule

Known external deadlines

Those who would like to aim their projects for conference submissions should note the following deadlines (the list will be amended to include other venues):
2/6/2016 ICML
2/12/2016 COLT
3/1/2016 UAI
3/14/2016 ECCV
6/3/2016 NIPS

Calendar-in-a-glance

<table>
<thead>
<tr>
<th>Date</th>
<th>Administrative</th>
<th>Main goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4/2016</td>
<td>Sha</td>
<td>Background and quick survey; (large-scale) computing</td>
</tr>
<tr>
<td>1/6/2016</td>
<td>Sha</td>
<td>environment setup; project teaming</td>
</tr>
<tr>
<td>1/11/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/13/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/18/2016</td>
<td>MLJ Day</td>
<td></td>
</tr>
<tr>
<td>1/20/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/25/2016</td>
<td>Travel</td>
<td></td>
</tr>
<tr>
<td>1/27/2016</td>
<td>Travel</td>
<td></td>
</tr>
<tr>
<td>2/1/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/3/2016</td>
<td></td>
<td>Deep dive; narrowing down project ideas</td>
</tr>
<tr>
<td>2/8/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/10/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/15/2016</td>
<td>President’s Day</td>
<td></td>
</tr>
<tr>
<td>2/17/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/22/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/24/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/29/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/2/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/7/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/9/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/14/2016</td>
<td></td>
<td>Make up and wrap up; project presentation</td>
</tr>
<tr>
<td>3/16/2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/18/2016</td>
<td></td>
<td>Final project report due at 5pm PT</td>
</tr>
<tr>
<td>3/26/2016</td>
<td></td>
<td>Final grade submitted</td>
</tr>
</tbody>
</table>
Temporary reading assignments

**Due times:** all due items are to be submitted electronically via email to feisha@cs.ucla.edu by 7pm Pacific Time. Include CS269 in your subject line to avoid delaying in processing your email.

<table>
<thead>
<tr>
<th>Date</th>
<th>Papers to present</th>
<th>Background reading, other tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4/2016</td>
<td>-</td>
<td>[1-3]</td>
</tr>
<tr>
<td>1/6/2016</td>
<td>-</td>
<td>[4-7]</td>
</tr>
<tr>
<td>1/11/2016</td>
<td>[10-12]</td>
<td>[8, 9]</td>
</tr>
<tr>
<td>1/13/2016</td>
<td>[17-18]</td>
<td>[13-16]</td>
</tr>
<tr>
<td>1/18/2016</td>
<td>MLJ Day</td>
<td></td>
</tr>
<tr>
<td>1/20/2016</td>
<td>[19-20]</td>
<td>[13-16], wrap-up</td>
</tr>
<tr>
<td>1/25/2016</td>
<td>Travel</td>
<td></td>
</tr>
<tr>
<td>1/27/2016</td>
<td>Travel</td>
<td>Teaming, catching up reading, look-ahead reading</td>
</tr>
<tr>
<td>2/1/2016</td>
<td>[21,23]</td>
<td>[2,22]</td>
</tr>
<tr>
<td>2/8/2016</td>
<td>[27-28]</td>
<td>[26]</td>
</tr>
<tr>
<td>2/10/2016</td>
<td>[30,31]</td>
<td>[29]</td>
</tr>
<tr>
<td>2/15/2016</td>
<td>President’s Day</td>
<td></td>
</tr>
<tr>
<td>2/17/2016</td>
<td>[32,33]</td>
<td>Discussion and 2nd survey topic: “regularization”; 1st survey due</td>
</tr>
<tr>
<td>2/22/2016</td>
<td>[34-36]</td>
<td></td>
</tr>
<tr>
<td>2/24/2016</td>
<td>[37-38]</td>
<td></td>
</tr>
<tr>
<td>2/29/2016</td>
<td>[39-40]</td>
<td></td>
</tr>
<tr>
<td>3/2/2016</td>
<td>[41-42]</td>
<td>Discussion and 3rd survey topic: “optimization”; 2nd survey due</td>
</tr>
<tr>
<td>3/7/2016</td>
<td>[43-44]</td>
<td></td>
</tr>
<tr>
<td>3/9/2016</td>
<td>[45-47]</td>
<td></td>
</tr>
<tr>
<td>3/14/2016</td>
<td>[48, X]</td>
<td>Discussion and 4th (optional) survey topic: “knowledge”; 3rd survey due</td>
</tr>
<tr>
<td>3/16/2016</td>
<td>-</td>
<td>project presentation</td>
</tr>
<tr>
<td>3/18/2016</td>
<td>Final project report and 4th survey due</td>
<td></td>
</tr>
<tr>
<td>3/26/2016</td>
<td>Final grade submitted</td>
<td></td>
</tr>
</tbody>
</table>

Reading list

The following reading list follows the schedule of the reading assignments in the previous section. The papers are organized according to sub-topics. They are by no means a comprehensive coverage of deep learning community. Rather, they are selected with a bias towards enhancing our scientific understanding of the learning paradigm.

   The slides for the vision part of the talk is at

January 4, 2016


http://www.mitpressjournals.org/doi/abs/10.1162/neco.2006.18.7.1527#.VojH0ZMrI

http://www.sciencemag.org/content/313/5786/504.short


http://www.deeplearningbook.org/contents/convnets.html


http://www.deeplearningbook.org/contents/rnn.html


   http://deeplearning.cs.cmu.edu/pdfs/Hochreiter97_lstm.pdf


    http://arxiv.org/abs/1506.02078

    http://arxiv.org/abs/1312.6098


25. N. Srivastava et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. JMLR 2014

26. C. Bishop. Training with noise is equivalent to Tikhonov regularization. Neural Compu-
    tation, 7(1), 108–116. 1995

27. S. Wang, C. Manning. Fast dropout training. ICML. 2013


    Layer Dropout. NIPS 2014


31. D. Kingma, T, Salimans and Max Welling. Variational Dropout and the Local Reparameter-
    ization Trick. NIPS 2015
   http://arxiv.org/abs/1506.02557

32. Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. What is the best multi-stage archi-

    AISTATS. 2011

34. Razvan Pascanu, Yann N. Dauphin, Surya Ganguli, Yoshua Bengio. On the saddle point
    problem for non-convex optimization. 2014
   http://arxiv.org/abs/1405.4604

35. Dauphin, Y., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., and Bengio, Y. Identifying and
    attacking the saddle point problem in high-dimensional non-convex optimization. NIPS. 2014.
   http://arxiv.org/abs/1406.2572
   \( \text{http://arxiv.org/abs/1412.6544} \)

37. Anna Choromanska, Mikael Henaff, Michael Mathieu, Grard Ben Arous, Yann LeCun. The Loss Surface of Multilayer Networks  
   \( \text{http://arxiv.org/abs/1412.0233} \)

38. Pratik Chaudhari, Stefano Soatto. Trivializing The Energy Landscape Of Deep Networks. 2015  
   \( \text{http://arxiv.org/abs/1511.06485} \)

   \( \text{http://arxiv.org/abs/1506.08473} \)

40. Place Holder. Topic in alternative methods for training DNN.  
   \( \text{http://www.google.com} \)

41. Hinton, G. Vinyals, O. and Dean, J. Distilling knowledge in a neural network. 2015  
   \( \text{http://arxiv.org/abs/1503.02531} \)

   \( \text{http://arxiv.org/abs/1412.6550} \)

43. Anoop Korattikara, Vivek Rathod, Kevin Murphy, Max Welling. Bayesian Dark Knowledge. NIPS. 2015.  
   \( \text{http://arxiv.org/abs/1506.04416} \)

44. David Lopez-Paz et al. Unifying distillation and privileged information.  
   \( \text{http://arxiv.org/pdf/1511.03643v2.pdf} \)

   \( \text{http://arxiv.org/abs/1312.6199} \)

   \( \text{http://arxiv.org/abs/1412.6572} \)
   http://papers.nips.cc/paper/5423-generative-adversarial-nets

   http://arxiv.org/abs/1411.1792

49. Relation to shallow kernel methods (placeholder, subject to change)

50. Deep generative models (placeholder, subject to change)
   - Salakhutdinov, R. and Hinton, G. Deep Boltzmann machines. AISTATS. 2009  
   - Kingma, D. P. and Welling, M. Auto-encoding variational Bayes. ICLR. 2014  
   - Bengio, Y., Yao, L., Alain, G., and Vincent, P. Generalized denoising auto-encoders as generative models. NIPS. 2013

51. Deep reinforcement learning (placeholder, subject to change)