Sketching Methods for Analysis of Matrices and Data: Take Home Assignment

Prof. Haim Avron

Fall 2019/2020

In the take home assignment you will use the background learned in class to independently study a recent research paper. You will summarize the paper in a report, empirically evaluate the algorithm suggested in the paper, and finally meet me to explain what you have learned.

1 Requirements

1. The project should be executed either individually or in pairs. If you chose to work in pairs, you must choose from the list of paper suitable for pairs (individuals can chose from the list of paper for pairs as well).
2. A paper can be chosen by multiple individuals/pairs. However, **you are strictly forbidden to discuss the papers between teams**.
3. Submit the report by **April 19, 2020**. Submission is by email.
4. Submit both the report and code reproducing all numerical results and figures in the report. No need to attach the code to the report; submit the code using a zip file. Include (in the zip file) detailed instructions on how to reproduce the results in the report.
5. You are free to use use MATLAB/Python/C/R or any other language of your choice.
6. Some of the papers below have the author's code on the web. **Do not download/use/inspect the author's code!** The only exception to this rule is if you have built an example for which their algorithm's supposedly fails, and you want to demonstrate it using their code.

2 Instructions

1. Choose one of the papers below. Read and understand the paper.
2. Empirically evaluate the algorithm. This can be done either by reproducing the results in the paper, or alternatively, designing your own set of experiments. If you chose the latter, make sure your experiments are comprehensive enough to support/disprove the conclusions of the paper. It is important that your experiments demonstrate the correctness of your implementation (e.g. by reproducing at least some of the results of the paper).
3. Write a summary of the paper (up to 5 pages), including your numerical results. You do not need to repeat proofs, but do explain the main proof technique/idea.

4. Prepare to meet me to explain what you have learned, and answer questions regarding the paper and/or your report. It is recommended, but not required, that you prepare a few slides (~ 5 slides) of executive summary for us to discuss.

3 Grading

- 50% on report (structure, clarity, accuracy, etc.)
- 20% on implementation (correctness and efficiency).
- 30% after-report discussion of the paper.

4 Papers

Papers for individual work:

1. *Random Projections for Linear Support Vector Machines*,
   Paul, Boutsidis, Magdon-Ismail and Drineas (AISTATS 2013)
   Remark: the journal version of the paper contains more experiments. You may reproduce only the experiments of the conference version.
2. *Randomized Dimensionality Reduction for k-Means Clustering*
   Boutsidis, Zouzias, Mahoney, and Drineas (IEEE Transactions on Information Theory, 2015).
3. *Sketched Ridge Regression: Optimization Perspective, Statistical Perspective, and Model Averaging*
   Wang, Gittens and Mahoney (ICML 2017)
4. *Fast and Accurate Least-Mean-Squares Solvers*
   Maalouf, Jubran and Feldman (NIPS 2019)
5. *Low-Rank Tucker Decomposition of Large Tensors Using TensorSketch*
   Osman Asif Malik and Stephen Becker (NIPS 2018)
6. *Nystrom Method vs Random Fourier Features: A Theoretical and Empirical Comparison*
   Yang, Li, Mahdavi, Jin and Zhou (NIPS 2012)
7. *Fast DPP Sampling for Nystrom with Application to Kernel Methods*
   Ji, Jegelka and Sra (ICML 2016)
8. *Efficient Frequent Directions Algorithm for Sparse Matrices*
   Ghashami, Liberty and Phillips (KDD 2016)
9. *Streaming Kernel Principal Component Analysis*
   Ghashami, Perry and Phillips (AISTATS 2016)
10. *SPALS: Fast Alternating Least Squares via Implicit Leverage Scores Sampling*
    Cheng, Peng, Liu and Perros (NIPS 2016)
11. *Sub-sampled Newton Methods with Non-uniform Sampling*
    Xu, Yang, Roosta-Khorasani, Re and Mahoney (NIPS 2016)

Papers for pair work:

10. I*terative Hessian Sketch: Fast and Accurate Solution Approximation for Constrained Least-Squares*
Pilanci and Wainwright (JMLR 2016)

11. *A Randomized Approximate Nearest Neighbors Algorithm*
Jones, Osipov and Rokhlin (ACHA 2013)