1 Locality Sensitive Hashing

- P. Indyk, R. Motwani, "Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality," In STOC, pages 604–613, 1998.
- M. Datar, N. Immorlica, P. Indyk, V.S. Mirrokn, "Locality-sensitive hashing scheme based on p-stable distributions," In SCG, pages 253–262, 2004.
- B. Neyshabur, N. Srebro, "On Symmetric and Asymmetric LSHs for Inner Product Search," ICML 2015.

2 Hashing for Model Compression

- A. Karatzoglou, A. Smola, M.Weimer, "Collaborative filtering on a Budget," AISTATS 2010.
- W. Chen, J.T. Wilson, S. Tyree, K.Q. Winberger, Y.Chen, "Compressing Neural networks with the Hashing Trick," ICML 2015.

3 Scalable Bayesian Learning

- M.D. Hoffman, D.M. Blei, F. Bach, "Online Learning for Latent Dirichlet Allocation," NIPS 2010.
- K. Zhai, J. Boyd-Graber, M. Asadi, M. Alkhouja, "Mr. LDA: A Flexible Large Scale Topic Modling Package using Variational inference in MapReduce," WWW 2012.
- M. Welling, Y.W. Teh, "Bayesian Learning via Stochastic Gradient Langevin Dynamics," ICML 2011.

4 Stochastic Gradient

- Zinkevich et al., "Parallelized Stochastic Gradient Descent."

5 Kernel Approximation

- Rahimi and Recht, "Weighted Sum of Random Kitchen Sinks."
- Rahimi and Recht, "Random Features for Large-Scale Kernel Machines."
- Le, Sarlos, Smola, "Fastfood - Approximating Kernel Expansion in Loglinear Time."

6 Parallel computing

- Li et al., "Scaling Distributed ML with the Parameter Server."
- Lin, Tsai, Lee, Lin, "Large-Scale Logistic Regression and linear SVM using Spark."
- Zinkevich, Weimer, Smola, Li, "Parallelized Stochastic Gradient Descent."

7 Boosting

- Bühlmann, Yu, "Boosting with L2-Loss."
- Friedman, "Greedy Function Approximation."
- Rätsch, Onoda, Müller, "Soft Margins for AdaBoost."

8 Dimensionality reduction/Visualization

- van der Maaten, Hinton, "Visualizing Data using t-SNE."
- van der Maaten, "Accelerating t-SNE using Tree-Based Algorithms."