10702 Statistical Machine Learning
Project Proposal
Disagreement Based Active Learning

Team Members: Varun Joshi (vjoshi1), Fabricio Flores (wflores)

Problem Description: Active learning is a general framework in machine learning where the algorithm has access to a small set of labeled examples and a lot of unlabeled examples. The algorithm can also access an expert which gives it the correct label for any example when queried. The goal of Active Learning algorithms is to ask for the labels of those examples which would create a highly accurate classifier with much less data as compared to passive learning. There are various techniques to do Active Learning. Disagreement Based Active Learning is one such technique in which multiple classifiers are built using labeled data and the algorithm asks for labels of those examples where at least two of the different classifiers disagree. We plan to present important results and proof sketches for this technique since it is well understood and has a rich variety of established results.

Scope of the project: Summarize the key theoretical results in Disagreement Based Active Learning and provide proof sketches of important results.

Reading List:

[1]: Hanneke, S. (2014). Theory of Disagreement-Based Active Learning. Foundations and Trends in Machine Learning, Vol. 7 (2-3), pp. 131-309

[2]: Burr Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison. 2009.

[3]: M.-F. Balcan, A. Beygelzimer, and J. Langford. Agnostic active learning. In Proceedings of the 23rd International Conference on Machine Learning, 2006.

[4]: A. Beygelzimer, D. Hsu, J. Langford, and T. Zhang. Agnostic active learning without constraints. In Advances in Neural Information Processing Systems (NIPS), Vancouver, 2010.

[5]: S. Dasgupta, D.J. Hsu, and C. Monteleoni. A general agnostic active learning algorithm. In Advances in Neural Information Processing Systems (NIPS), Vancouver, 2007

[6]: M.-F. Balcan, S. Hanneke, and J. Wortman. The true sample complexity of active learning. In Proceedings of the 21st Conference on Learning Theory, 2008.

[7]: S. Hanneke. A bound on the label complexity of agnostic active learning. In *Proceedings of the 24th International Conference on Machine Learning*, 2007b