George Zhang
Computer Science 227r
Cynthia Dwork; Victor Balcer
Due: March 24, 2017

Final Project Proposal

The goal of my final project is to investigate the experimental utility of private algorithms versus their non-private counterparts, specifically in regards to private stochastic gradient descent. We have looked at two versions of private stochastic gradient descent: one where we perturb the gradient used for minimization and one where we perturb the objective function itself. I would like to see how these two methods differ from each other in terms of the minimum they find as compared to both each other and the non-private algorithm.

I will look at a basic two-class classification problem and use logistic regression in my algorithms. The reason behind this decision is that it is a simple setup and will allow me to focus on the analysis of the algorithms. The train and test data can be easily generated, and again for simplicity, we shall assume that the data is separated perfectly by a hyperplane. If time allows, I will also generate more noisy data, and see how a less perfect division of classes affects the algorithms. Because the data is generated, I will be able to not only compare the private versions to themselves and the non-private algorithm, but also compare all these algorithms to the true boundary.

Secondary goals for this project include analysis with respect to theoretical utility, visualization of algorithms, and extensions of the private algorithm beyond two classes. I imagine analysis with respect to theoretical utility would be a logical next step after getting data on the algorithms and would tie the experimental results back to the theoretical guarantees. The visualization of the algorithms would be in line with my goal of producing concrete examples that could be used when teaching the material. Most likely, this will take the form of a 2-dimensional classification problem, where the algorithm iterates through the steps. A further extension would be to generalize logistic regression to multinomial logistic regression, where there are more than two classes, and analyze the private versions of those algorithms.

My proposed timeline for this project is as follows:

- Friday March 31st – Implement foundational pieces (data generation, non-private version)
- Friday April 7th – Implement both private algorithms
- Friday April 14th – Finish data collection (algorithm utility & comparison)
- Thursday April 20th – Be prepared for presentation (visualization here if possible)
- Friday April 28th – Complete all data analysis / secondary goals (as time allows)
- Wednesday May 3rd – Write-up complete