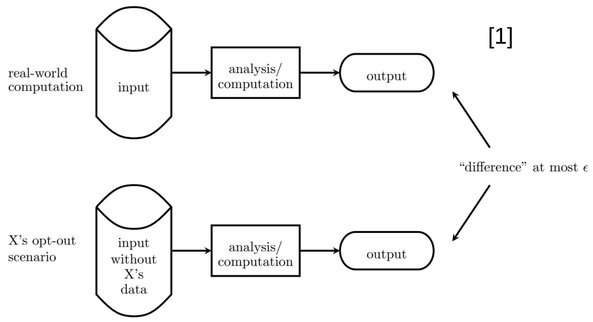
Reading Between the Lines: Surveying Differential Privacy in Different Regression Methods

Eric Shen
Harvard College
Email: ericshen@college.harvard.edu

Nishant Mishra
Harvard College
Email: nmishra@college.harvard.edu

Introduction

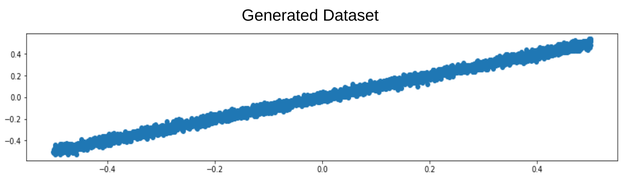
The field of differential privacy (DP) offers a framework where potentially-sensitive data can be analyzed in aggregate while limiting the information that can be known about individual data entries. Work in the field has focused on how DP techniques can be applied to a variety of regression paradigms.



Here, we visit three commonly-used and different methods of linear regression applied to different datasets — ridge regression, lasso regression, and Bayesian regression—and provide a review of how each of these techniques has been modified to satisfy DP in academic literature. We also attempt code implementations for these regression techniques and evaluate their performance on simple datasets. Based on our efforts, we discuss practical considerations, challenges, and recommendations of the DP techniques.

Background and Datasets

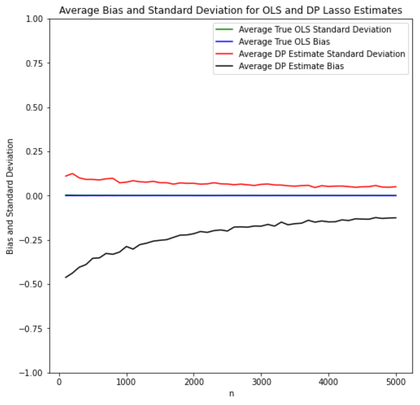
In our project, we make several references to the standard DP definitions [1] as well as some basic regression techniques already covered, namely linear regression.



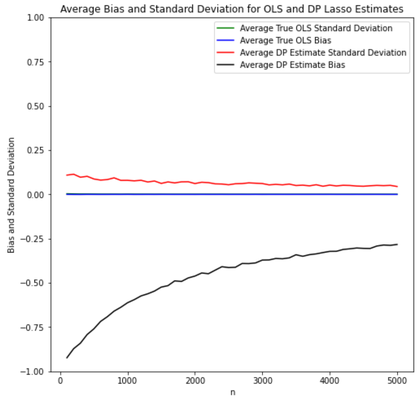
Our code implementations of our DP techniques generates simple one-dimensional datasets to use as a common point of comparison for testing. In particular, we restrict X to N uniformly-generated points in the interval [−1, 1], and generate y according the equation y = X + e, where e ∼ N (0, σ^2). Furthermore, y points are also winsorized to the interval [−1, 1]. To evaluate algorithm performance with variations in this data, we vary N and σ^2.

Lasso Regression

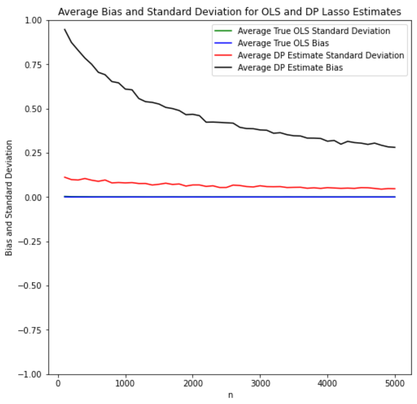
True Slope = 0.5



True Slope = 1.0



True Slope = -1.0



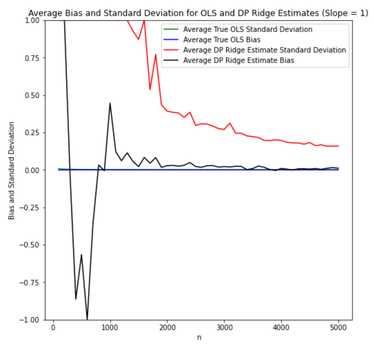
First, we implemented differentially private lasso regression with the simple dataset, basing our update steps off the Frank-Wolfe algorithm [2], specifically with T = 100 update steps, starting with an initial estimate of 0. We fixed σ^2 = 0.02 and considered increments of N from 1000 to 5000 in steps of 100. For each value of N , 100 trials were performed. In each trial, the OLS and DP Lasso standard deviation and bias were calculated (recalling that the slope from the data- generating process was set to be 1), with the results displayed above.

Conclusions

- The variance of Lasso estimates decreases as the number of data points increases, which may hint at convergence of the Lasso estimator under this DP technique.
- The main conclusion of note, however, is that the bias is consistently negative (i.e. values of $\hat{\theta}$ are below 1.0), but decreases in magnitude as N increases.

Ridge Regression

Next, we incorporated differential privacy mechanisms with ridge regression. When creating models to fit to data, ridge regression has proven especially useful in machine learning by adopting most of the elements of linear regression, while penalizing model coefficients that are relatively large in magnitude, ensuring that models aren't over fit to the training data that they're given.



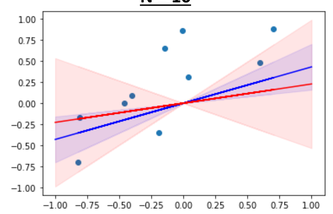
After mimicking most of the same parameters as in our lasso regression tests, this time with a 0.02 penalizing parameter, we find that the results are not particularly different from the standard OLS results, indicating that either 1) our N is not large enough, or 2) our penalizing coefficient's magnitude was too low to see a difference.

Bayesian Regression

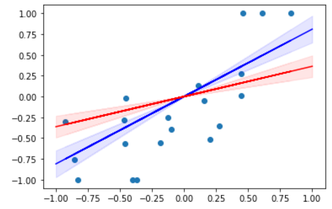
The two previous techniques covered have operated under the frequentist interpretation of probability, so we decided to implement differentially private Bayesian regression to produce point estimates of θ.

We set σ^2 = 0.5, and generate four datasets, varying values of N ∈ {10, 20, 50, 100}. The priors for the mean and variance on the slope is 0 and 1, respectively. In the plots displayed below, each dataset is displayed as a scatter plot with the Bayesian linear regression approximations overlaid.

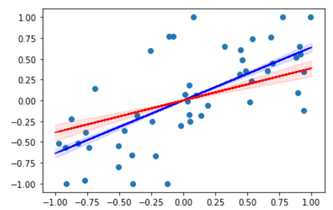
N = 10



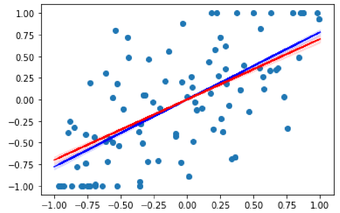
N = 20



N = 50



N = 100



The line representing the slope calculated using the raw data is displayed in blue, while the line representing the slope calculated using the "noise-naive" DP method is displayed in red. Shaded regions indicate plots for the 95% confidence intervals given the posterior variance for the slope values (i.e. the upper bound of the shaded region denotes the smallest slope value in the confidence interval, and similarly for the lower bound).

Conclusions

- The width of the 95% confidence intervals in slope space decreases with N
- The overall accuracy of the DP estimate approximates the raw estimate as N increases
- small N values are an issue - in the cases of N = 20 and N = 50, the noise added to small values of the posterior variance can result in slope ranges which do not capture the true slope of the data-generating process.

Conclusions and Future Directions

Conclusion

- When working with 1D datasets, the lasso regression technique leads to noticeably more stable plots than the OLS results uncovered in class. Ridge regression, however, was not particularly helpful in this dataset, and bayesian regression struggled relatively more at lower N.

Future Directions

- We should expand to multi-dimensional datasets, to further explore the advantages of various regression techniques with DP.
- For ridge regression specifically, we should edit the specific parameters when generating 1D dataset to see if there *is* a scenario where a noticeable difference exists.
- We could also try to implement other regression techniques, such as logistic and polynomial regression.

References

[1]: Kobbi Nissim, et al. Differential Privacy: A Primer for a Non-technical Audience. February 14, 2018.
[2]: Jaggi, Martin (2013). "Revisiting Frank–Wolfe: Projection-Free Sparse Convex Optimization". Journal of Machine Learning Research: Workshop and Conference Proceedings. 28 (1): 427–435. (Overview paper)

Acknowledgements

Special thanks to Professor Salil Vadhan and the rest of the CS 208 course staff for their support and guidance for the duration of this course and this project.