Leverage based sampling for classification

Technical University of Denmark

Julian Kopka Larsen and Jesper Løve Hinrich .

Superviser Lars Kai Hansen

DTU Compute · Technical University of Denmark

Kgs. Lyngby, Denmark Copenhagen N, Denmark

DTU Compute
Department of Applied Mathematics and Computer Science

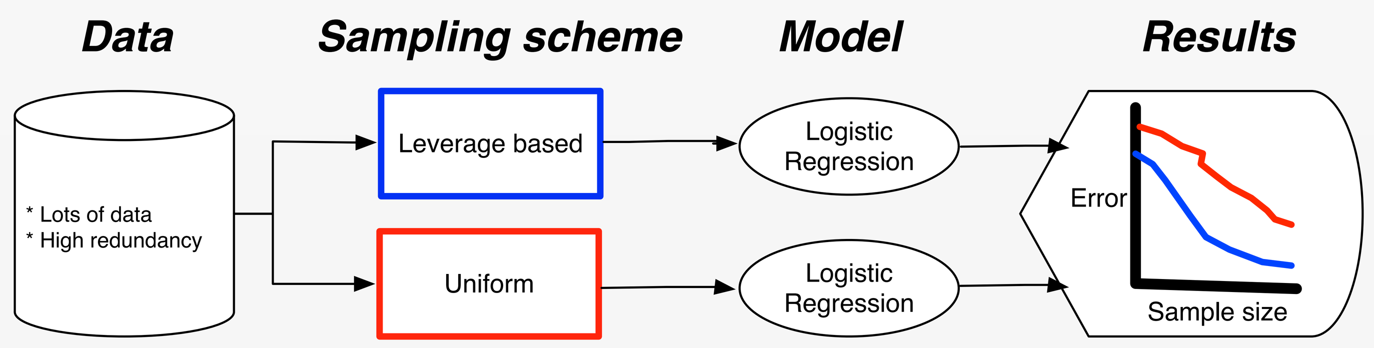
Abstract

We validate the results of leverage based sampling for LS-regression, shown by Ma et al. [1]. We explore the possibility of using the leverage based sampling distribution from LS-regression on 2 class classification, and introduce a new approach for sampling from an leverage distribution (important points).

Motivation

The importance of sampling methods is initiated by very large datasets where it is not feasible to use all of the available data. This is illustrated by the rise in online access to video data. These data contain many frames that are basically the same and therefore redundant.

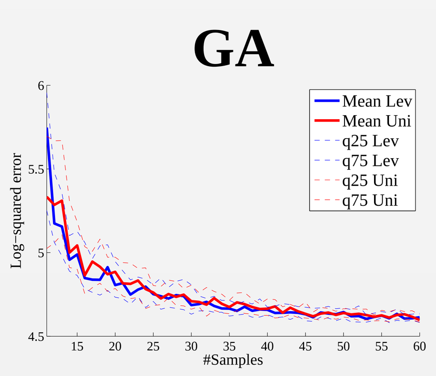
Concept



Research Questions

- Will the regression based sampling distribution improve our performance in classification?
- Can leverage based sampling be generalized to classification?

Datasets



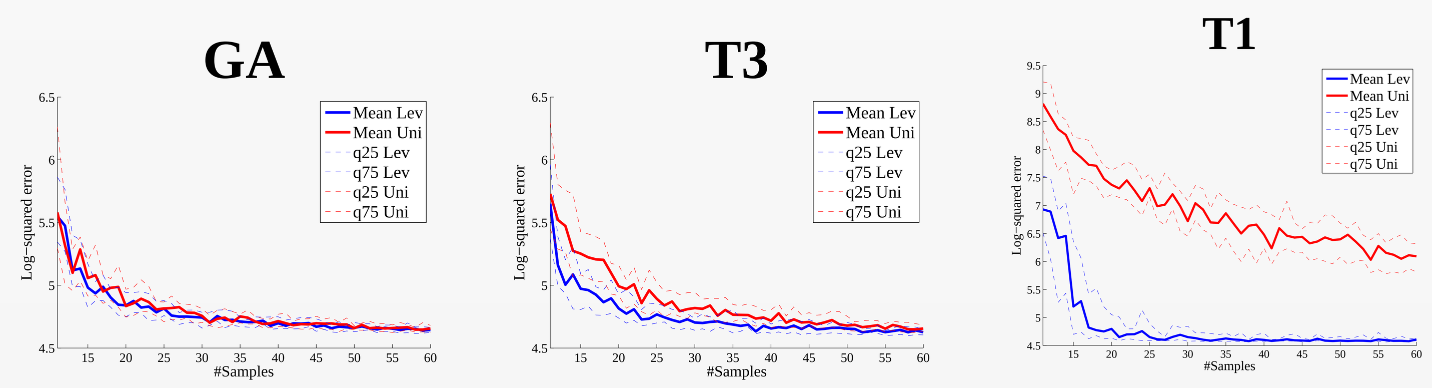
These datasets are drawn from distributions defined in Ma et al. [1] and have:

- GA: Nearly uniform leverage-scores
- T3: Mildly non-uniform leverage-scores
- T1: Very non-uniform leverage-scores

Leveraging in for least squares regression

For any leveraging we want to find for each datapoint a leverage-score representing the importance of the point. This we can normalize into a distribution to sample from.

Validating the results of Ma et al.



Figur 2: Lal di la

LS-based Distribution

We seek to find the effect that a datapoint's class has on the predicted class for that datapoint.

$$\frac{\delta \hat{y}_n}{\delta y_n} \tag{1}$$

There is a closed form solution which is linear in y

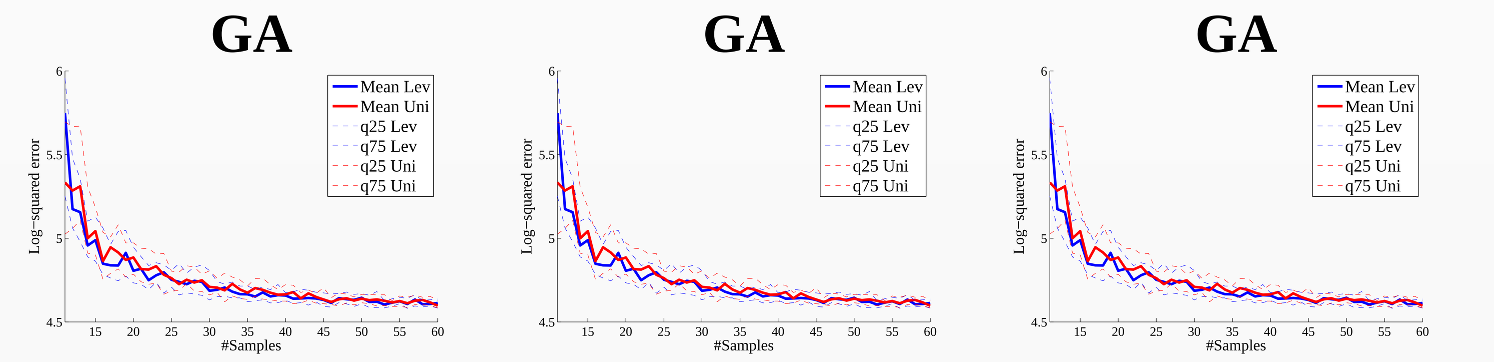
$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y \quad \text{where} \quad \hat{y}_n = X_n * \hat{\beta}$$

Therefore the leverage-score (1) is the coefficient

$$\frac{\delta \hat{y}_n}{\delta y_n} = X\left(X^T X\right)^{-1} X^T$$

Test Results

We compared logistic regression when sampling from a LS-distribution (blue) vs. a uniform-distribution (red). The mean, 25th and 75th quantile are plotted.



It clearly shows that a LS-distribution sample scheme, does not outperform a uniform-distribution for classification. The results shown are for dimension $p = 10$ and $N = 1000$ datapoints, but it is consistent when varying p and N.

Sensitivity Based Distribution

Vi generalisere til noget andet, og derfor kan WLS ikke bruges, s**r**a vi laver bare LS. The target is again (1)

$$\hat{y}_n = p(y|\bar{x}, \bar{w}) \quad \bar{w} \text{ s.t. } \frac{\delta L}{\delta \bar{w}} = 0$$

Since depends both directly and indirectly on y we see that

$$\frac{\delta}{\delta y}\frac{\delta L}{\delta w} = 0$$

$$\Downarrow$$

$$\frac{\delta^2 L}{\delta y \delta \bar{w}} + \frac{\delta^2 L}{\delta \bar{w} \delta \bar{w}^T}\frac{\delta \bar{w}}{\delta y} = 0$$

and from this we can get our leverage-score (1)

$$\frac{\delta \hat{y}_n}{\delta y_n} = \frac{\delta p(y|\bar{x}_n, \bar{w})}{\delta \bar{w}^T}\frac{\delta \bar{w}}{\delta y} = -\frac{\delta p(y|\bar{x}_n, \bar{w})}{\delta \bar{w}^T}\left[\frac{\delta^2 L}{\delta \bar{w} \delta \bar{w}^T}\right]^{-1}\frac{\delta^2 L}{\delta y \delta \bar{w}}$$

When evaluating this model, weights trained on a small training-set is used. This is expected to be better than LS-based sampling since it introduces dependence on class information.

Conclusion

The LS-based leverage sampling gives no advantage over uniform sampling and generally performs worse. LS-distribution is based on what is important for linear regression, it does not have an advantage in finding important points for classification.

References

Litteratur

[1] Ma et al. A statistical perspective on algorithmic leveraging. *arXiv:1306.5362v1 [stat.ME]*, June 2013.