Overview

- Scientists, engineers & social scientists are often interested in the relationship between a large set of features and a response.
- For example, a biologist may wish to understand the effect of natural variations of certain genes on the presence of a disease.
- Bayesian generalized linear models (GLMs) provide coherent uncertainty quantification but can be slow to learn.
- We propose a low rank approximation of data -- as a form of likelihood approximation.
- We show improved dimension dependence in time and memory scaling of inference.
- We provide theoretical guarantees and experiments providing a 10x speed-up with minimal approximation error.

Background

Generalized Linear Models (GLMs)

- Consider a regression of $N, D$-dimensional covariates, $X$ on $N$ responses, $Y$.
- GLMs are a widely used class of interpretable models w. parameter $\beta \in \mathbb{R}^D$.
- Accommodate different response types (counts, binary, heavy-tailed)
- Characterized by likelihoods of the form: $y_i|x_i, \beta \sim p(y_i|x_i^T \beta)$

Conjugate Gaussian Bayesian Regression

Generative Model

$\beta \sim N(0, \sigma \beta I)$

$p(\beta|Y, X) = N(\mu_N, \Sigma_N)$

for $i = 1, 2, \ldots, N :$

$y_i \sim N(x_i^T \beta, \tau^{-1})$

$\mu_N := \tau \Sigma_N X^T Y$

***Posterior has an analytic form, but inference takes $O(ND^2+D^3)$ time***

Conjugate Regression when $X$ is Rank $M<D$

- We can write the SVD of $X$ as: $X = U \text{diag}(\lambda) V^T$
- for some $U \in \mathbb{R}^{D,M}$, $V \in \mathbb{R}^{N,M}$ with $M < D, N$
- And then:

$\Sigma_N = \sigma^2 \left(I - U \text{diag} \left( \frac{\tau \lambda^2}{\sigma^2 + \tau \lambda^2} \right) U^T \right)$ and $\mu_N = U \frac{\tau \lambda}{\sigma^2 + \tau \lambda} V^T Y$

***Exact inference takes $O(NDM)$ time***

Inference Method

<table>
<thead>
<tr>
<th>Method</th>
<th>Naive</th>
<th>Our Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC (per iteration)</td>
<td>$O(DN)$</td>
<td>$O([D+N]M)$</td>
</tr>
</tbody>
</table>

When data are exactly low rank, our approach is exact; otherwise it is an approximation.

Key Theoretical Results

Theorem: In conjugate linear regression, if each $|y_i| < \bar{b}$ our approximation $\tilde{p}(\beta|X, Y) = N(\tilde{\mu}_N, \tilde{\Sigma}_N)$, satisfies:

$\|\tilde{\mu}_N - \mu_N\|_2 \leq \sigma^2 \sqrt{\lambda_{M+1}^2 + \lambda_{M+1}} \sqrt{N \bar{b}}$

Also, $\Sigma_N^{-1} - \tilde{\Sigma}_N^{-1} = \tau (X^T X - U U^T X U U^T)$,

hence $\|\Sigma_N^{-1} - \tilde{\Sigma}_N^{-1}\|_2 = \tau \lambda_{M+1}^2$.

Corollary (consistency):

$\tilde{\mu}_N \approx \mu_N$, the maximum a priori vector satisfying $U^T \tilde{\mu} = U^T \beta$

Corollary (conservativeness):

$\tilde{p}(\beta|X, Y)$ is no less uncertain than $p(\beta|X, Y)$; Formally, $\tilde{\Sigma}_N \succeq \Sigma_N$ and $H(\tilde{p}(\beta|X, Y)) \geq H(p(\beta|X, Y))$.

Results for Logistic Regression

Approximate posterior mean and standard deviation across a subset of parameters as $M$ varies. X-axis represents ground truth from running Hamiltonian Monte Carlo.

Acknowledgements

We thank Raj Agrawal for helpful comments and discussion.