

SDS385 HW 1

Evan Ott
UT EID: eao466
August 29, 2016

Linear Regression

(A)

WLS objective function:

$$\begin{aligned}\sum_{i=1}^N \frac{w_i}{2} (y_i - x_i^\top \beta)^2 &= \frac{1}{2} \sum_{i=1}^N y_i w_i y_i - \sum_{i=1}^N y_i w_i x_i^\top \beta + \frac{1}{2} \sum_{i=1}^N x_i^\top \beta w_i x_i^\top \beta \\ &= \frac{1}{2} y^\top W y - y^\top W X \beta + \frac{1}{2} (X \beta)^\top W X \beta \\ &= \frac{1}{2} (y - X \beta)^\top W (y - X \beta).\end{aligned}$$

Minimizing this function means setting the gradient (with respect to β) to zero:

$$\nabla_\beta \left[\frac{1}{2} (y - X \beta)^\top W (y - X \beta) \right] = 0$$

That is

$$\begin{aligned}\nabla_\beta \left[\frac{1}{2} (y - X \beta)^\top W (y - X \beta) \right] &= 0 - (y^\top W X)^\top + \frac{2}{2} X^\top W X \hat{\beta} = 0 \\ &\Rightarrow (X^\top W X) \hat{\beta} = X^\top W y \quad \blacksquare\end{aligned}$$

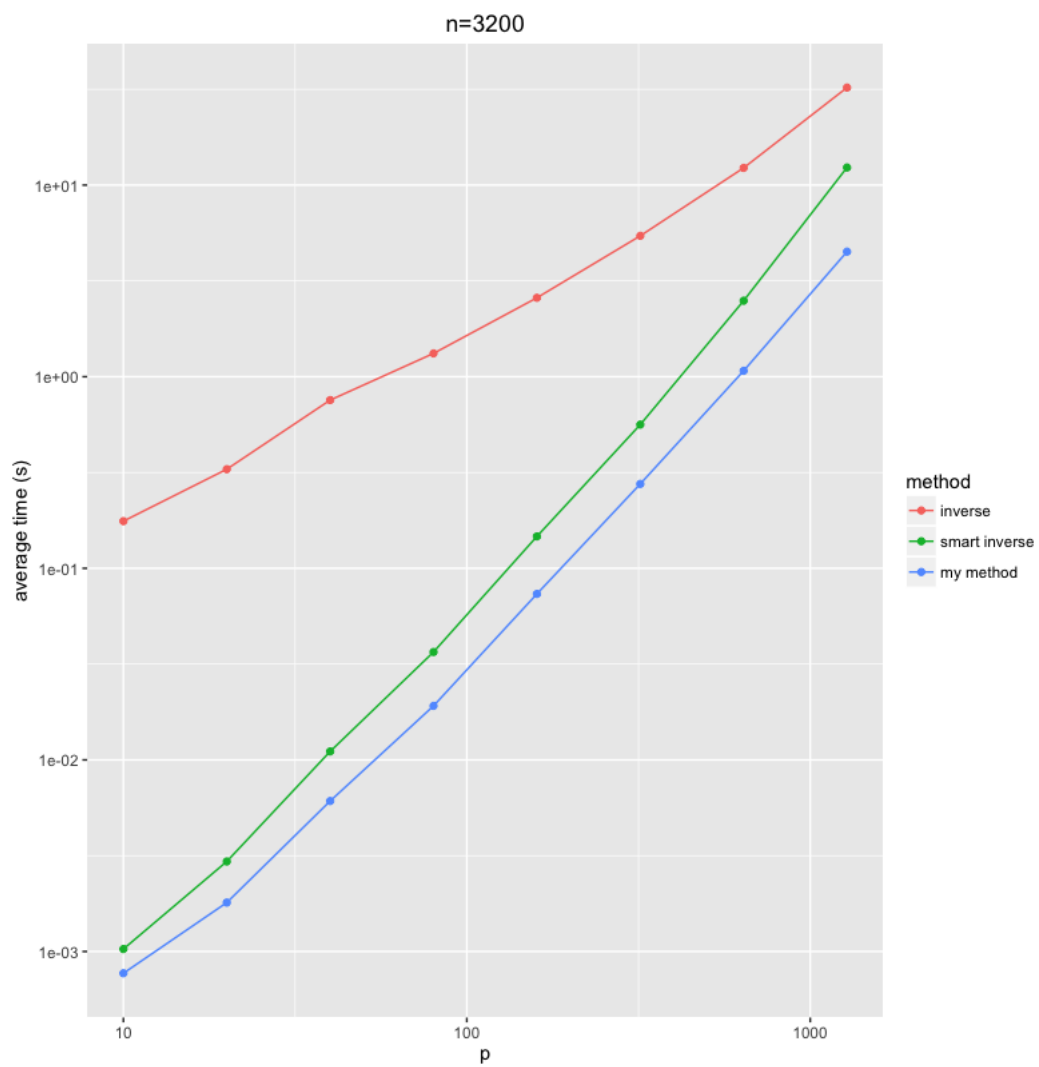
(B)

The matrix factorization idea basically amounts to trying to prevent the full inverse operation. Overall, you will still probably need something $O(n^3)$, but the constants matter when actually doing computation as opposed to asymptotics. We don't actually want the inverse anyway, we just want to solve $(X^\top W X) \hat{\beta} = X^\top W y$ for β . Using a matrix decomposition can help with that a lot.

I based my solution on the Cholesky decomposition (see <http://www.seas.ucla.edu/~vandenbe/103/lectures/chol.pdf>). This creates matrices $X = LL^\top$, where L is lower-triangular. So, then solve $Lz = X^\top W y$ for z and $R\beta = z$ for β . The decomposition is still $O(n^3)$, but faster than inverse. The two final steps are each $O(n^2)$ which are faster.

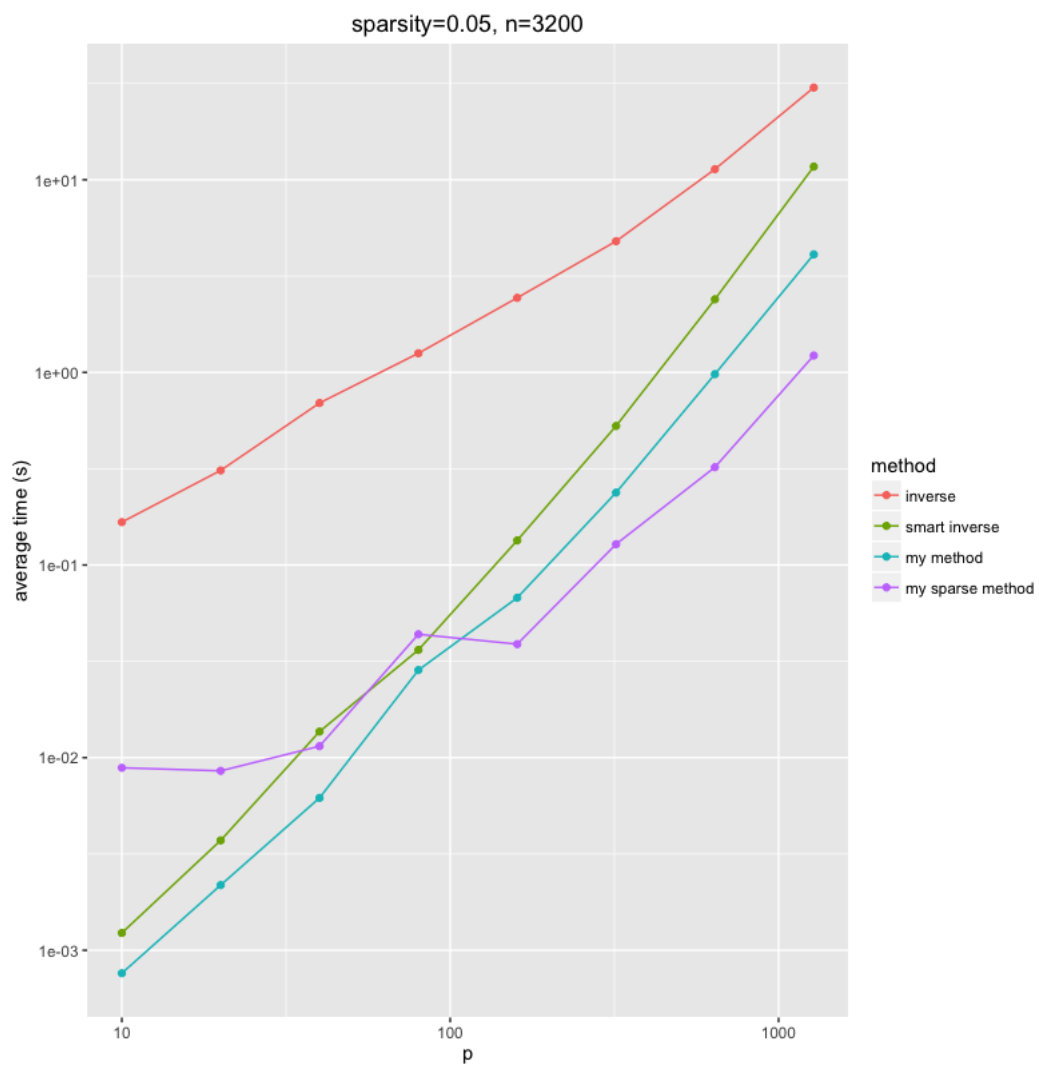
(C)

See code on GitHub.



(D)

See code on GitHub.



Generalized linear models

(A)

$$\begin{aligned}
l(\beta) &= -\log \left\{ \prod_{i=1}^N p(y_i|\beta) \right\} \\
&= -\log \left\{ \prod_{i=1}^N \binom{m_i}{y_i} w_i(\beta)^{y_i} (1 - w_i(\beta))^{m_i - y_i} \right\} \\
&= -\sum_{i=1}^N \log \left\{ \binom{m_i}{y_i} w_i(\beta)^{y_i} (1 - w_i(\beta))^{m_i - y_i} \right\} \\
&= -\sum_{i=1}^N \log \binom{m_i}{y_i} + y_i \log(w_i(\beta)) + (m_i - y_i) \log(1 - w_i(\beta)) \\
\nabla l(\beta) &= -\sum_{i=1}^N 0 + \frac{y_i}{w_i(\beta)} \nabla w_i(\beta) - \frac{m_i - y_i}{1 - w_i(\beta)} \nabla w_i(\beta) \\
\nabla w_i(\beta) &= w_i^2(\beta) e^{-x_i^\top \beta} x_i \\
\nabla l(\beta) &= -\sum_{i=1}^N w_i^2(\beta) e^{-x_i^\top \beta} x_i \left(\frac{y_i}{w_i(\beta)} - \frac{m_i - y_i}{1 - w_i(\beta)} \right) \\
&= -\sum_{i=1}^N w_i^2(\beta) e^{-x_i^\top \beta} x_i \left(\frac{y_i - y_i w_i(\beta) - m_i w_i(\beta) + y_i w_i(\beta)}{w_i(\beta)(1 - w_i(\beta))} \right) \\
&= -\sum_{i=1}^N w_i(\beta) e^{-x_i^\top \beta} x_i \left(\frac{y_i - m_i w_i(\beta)}{1 - w_i(\beta)} \right) \\
&= -\sum_{i=1}^N w_i(\beta) \left(\frac{1}{w_i(\beta)} - 1 \right) x_i \left(\frac{y_i - m_i w_i(\beta)}{1 - w_i(\beta)} \right) \\
&= -\sum_{i=1}^N w_i(\beta) \frac{1 - w_i(\beta)}{w_i(\beta)} x_i \left(\frac{y_i - m_i w_i(\beta)}{1 - w_i(\beta)} \right) \\
&= -\sum_{i=1}^N [y_i - m_i w_i(\beta)] x_i
\end{aligned}$$