1 GLM Part-A

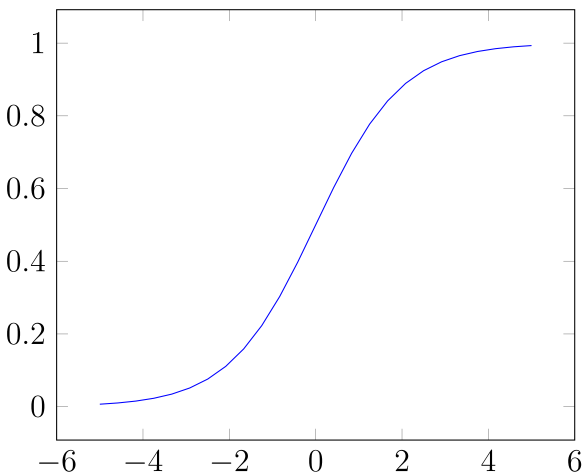
Binary Logistic Regression:
logit(y_i)=$x_i'\beta$, where $y_i \sim$ Bernoulli(w_i), $w_i = \frac{1}{1+exp(-x_i'\beta)}$.

$p(y_i|x_i, \beta) = (w_i(\beta))^{y_i}(1 - w_i(\beta))^{1-y_i}, y_i = 0, 1.$

The likelihood function is:
$L(\beta) = \prod_{i=1}^{N} p(\beta|y_i) = \prod_{i=1}^{N}(w_i(\beta))^{y_i}(1 - w_i(\beta))^{1-y_i}$

Lemma: If $g(z) = \frac{1}{1+exp(-z)}$, then $g'(z) = g(z)(1 - g(z))$.

Proof: $g'(z) = \frac{d}{dz}\frac{1}{1+e^{-z}} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}}(1 - \frac{1}{1+e^{-z}}) = g(z)(1 - g(z))$



Maximizing the log likelihood is equivalent to minimize the negative log likelihood which is known as the Loss function:

$l(\beta) = -log\{\prod_{i=1}^{N} p(\beta|y_i)\}$

$\rightarrow$ -$\sum_{i=1}^{N}\{y_i log(w_i(\beta)) + (1 - y_i)log(1 - w_i(\beta))\}$

Let $l^*(\beta) = y_i log(w_i(\beta)) + (1 - y_i)log(1 - w_i(\beta))$

Using the lemma we get: $w_i'(\beta) = x_{ij}w_i(\beta)(1 - w_i(\beta))$

Thus, $\rightarrow (\nabla l^*(\beta))_j = \frac{y_i x_{ij} w_i(\beta)(1-w_i(\beta))}{w_i(\beta)} - \frac{(1-y_i)x_{ij}w_i(\beta)(1-w_i(\beta))}{(1-w_i(\beta))}$

$\rightarrow (\nabla l^*(\beta))_j = y_i x_{ij}(1 - w_i(\beta)) - (1 - y_i)x_{ij}w_i(\beta)$

Therefore, $(\nabla l(\beta))_j =$ -$\sum_{i=1}^{N}(y_i x_{ij} - w_i(\beta)x_{ij}) = \sum_{i=1}^{N}(w_i(\beta) - y_i)x_{ij}$

2 GLM Part-B

The sample size of my dataset is 404. Outcome variable is small for gestational age (0/1) and the predictors are prenatal phthalate exposures. There are six predictors in the dataset. Let, for ith individual, y_i=small for gestational age, $\boldsymbol{x_i}$=design matrix for phthalate metabolic concentrations: MBP, MIBP, MEP, MBZP, MCPP, DEHP. We specify the model as $y_i \sim$ Bernoulli(p_i), logit(p_i)=$\boldsymbol{x_i'}\beta$.

Application of Gradient descent on real dataset:
Estimates from GLM method:

Intercept	MBP	MIBP	MEP	MBZP	MCPP	DEHP
-1.981	0.206	0.163	-0.086	-0.450	-0.205	0.187

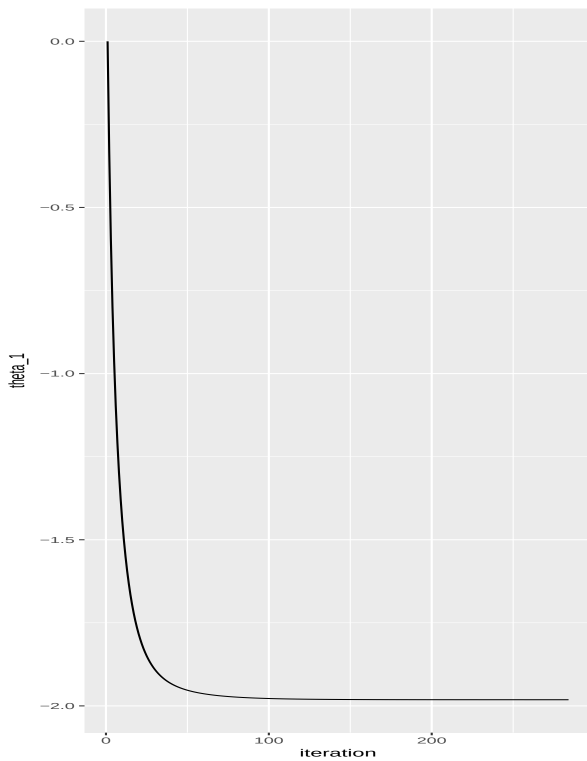
Estimates from Gradient Descent method with step size 0.05:

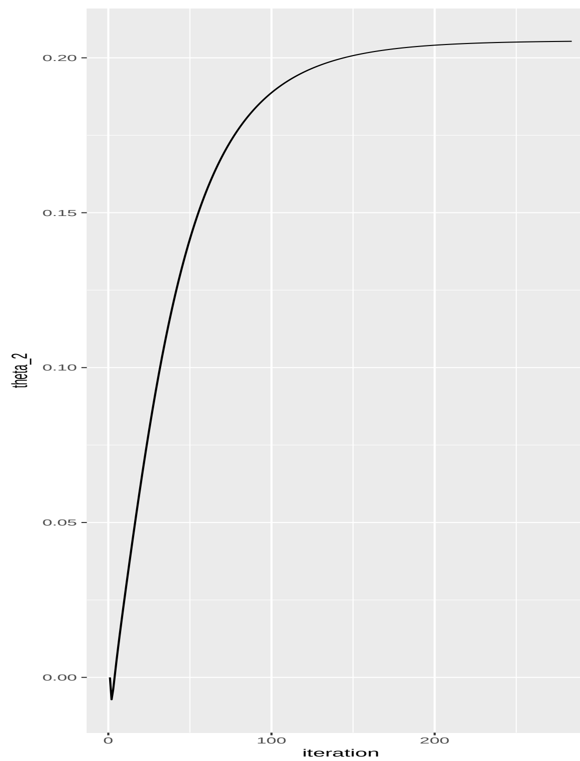
Intercept	MBP	MIBP	MEP	MBZP	MCPP	DEHP
-1.958	0.150	0.150	-0.086	-0.404	-0.165	0.175

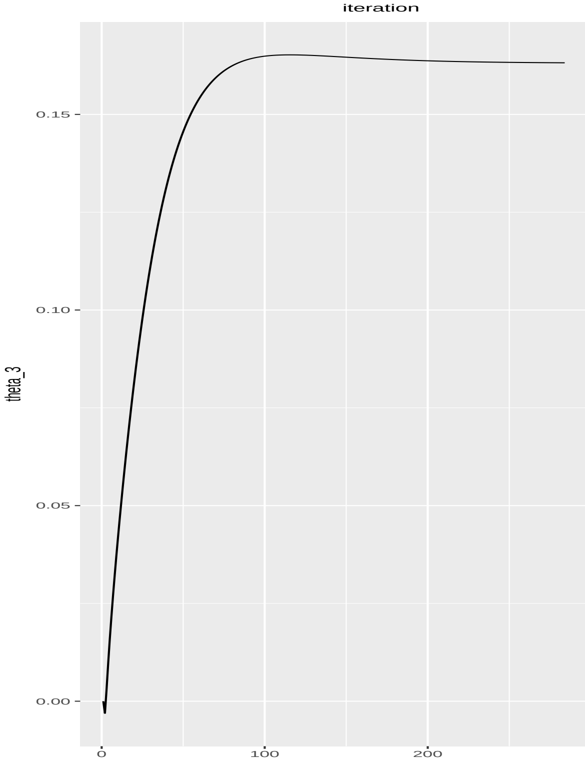
Estimates from Gradient Descent method with step size 0.1:

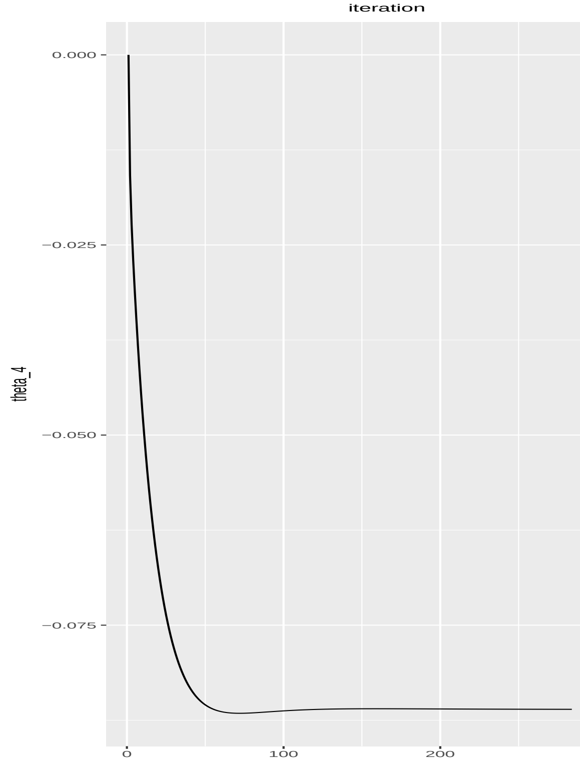
Intercept	MBP	MIBP	MEP	MBZP	MCPP	DEHP
-1.979	0.193	0.165	-0.086	-0.446	-0.195	0.186

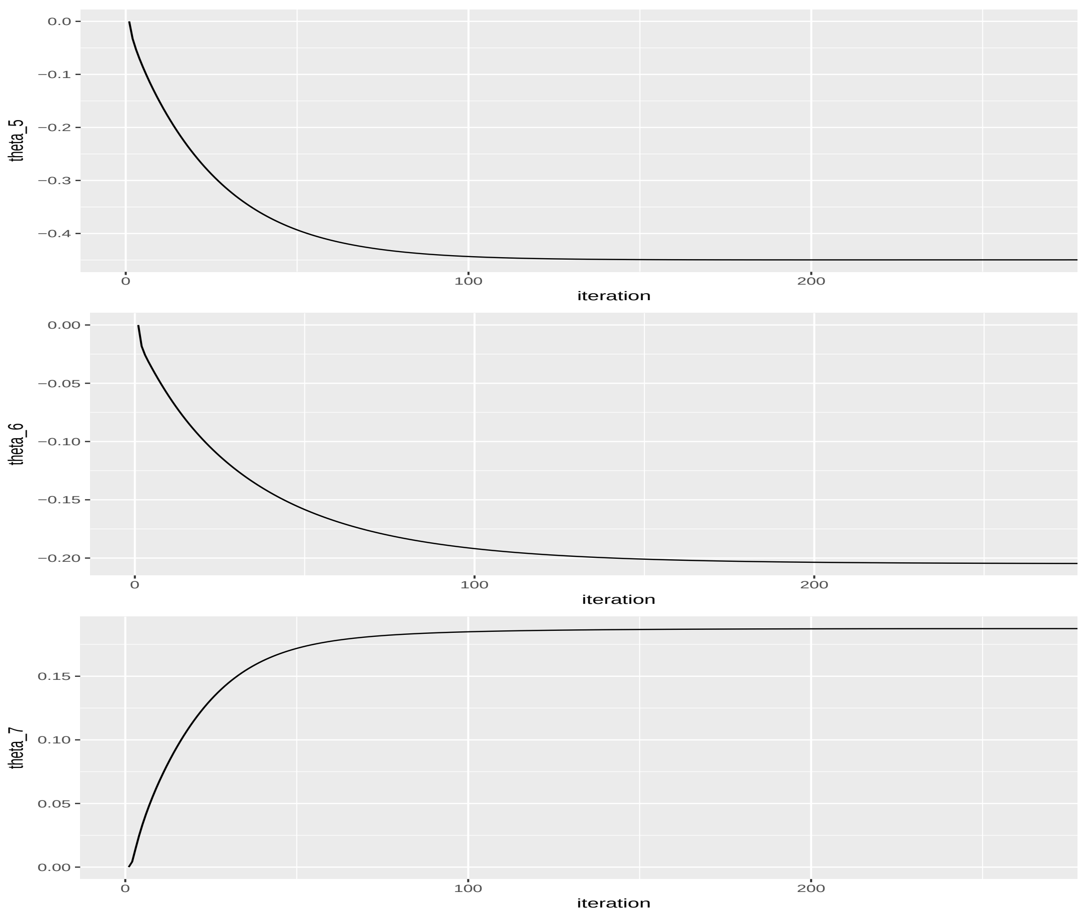
As we can see from the above tables that as the step size increases, the method gives better result and the estimates become very close to the GLM estimates.


0.0
−0.5
−1.0
−1.5
−2.0
theta_1
0
100
200
iteration


0.20
0.15
0.10
0.05
0.00
theta_2
0
100
200
iteration


0.15
0.10
0.05
0.00
theta_3
0
100
200
iteration


0.000
−0.025
−0.050
−0.075
theta_4
0
100
200
iteration



After implementing gradient descent on logistic regression, we notice that the procedure takes a number of steps to converge. The above plots show the patterns of convergence with numbers of iterations for each parameter.

Convergence check:

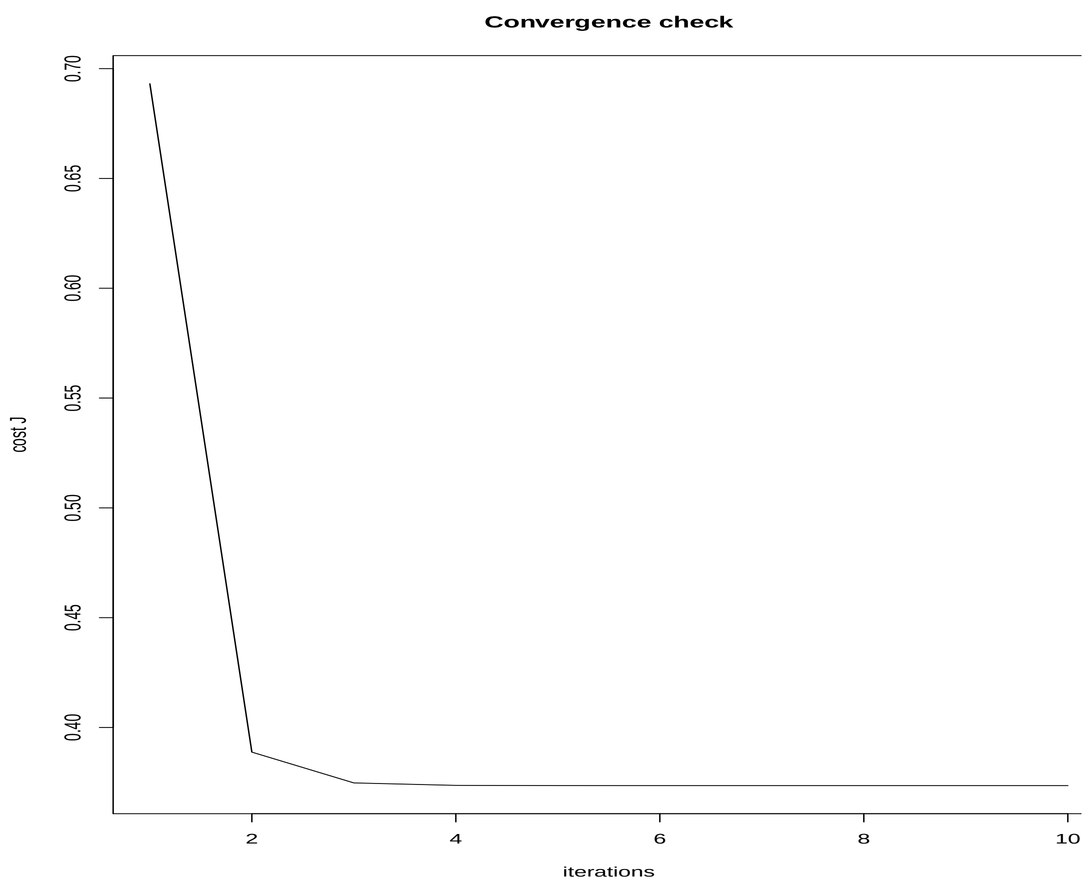
To check the convergence, we change the number of iterations and obtain the values of cost function. Then we store those values as cost1000 and cost1001. The difference between two values is 2.574919e-08 which is infinitely small. So we can claim that the difference meets our convergence criteria.

3 GLM Part-D

Application of Newton's method of real dataset:
We use the same dataset as we did in Part B. After implementing the procedure on logistic regression, we get the following estimates:

Intercept	MBP	MIBP	MEP	MBZP	MCPP	DEHP
-1.981	0.206	0.163	-0.086	-0.450	-0.205	0.187

It seems that the estimates from Newton method are very close to the estimates from GLM.



From this plot, we can infer that Newton's Method has converged by around 10 iterations. In the previous part, gradient descent took hundreds or even thousands of iterations to converge. Newton's Method is much faster in comparison.