FOML Assignment1

SADINENI ABHINAY - CS21BTECH11055
NIMMALA AVINASH - CS21BTECH11039
Group 64

October 2023

Question 4

Let us consider the following notations

(i) $h_w(x)$ be model function

(ii) w be parameter

$$h_w(x) = g(w^Tx) = \frac{1}{1+e^{-w^Tx}} \tag{1}$$

$$g(z) = \frac{1}{1+e^{-z}} \tag{2}$$

Let's assume that

$$P(y=1|x;w) = h_w(x) \tag{3}$$

$$P(y=0|x;w) = 1 - h_w(x) \tag{4}$$

$$P(y|x;w) = (h_w(x)))^y(1-h_w(x))^{(1-y)} \tag{5}$$

(a) (i) **Log likelihood:**

$$L(w) = p(y|x;w) \tag{6}$$

$$= \prod_{i=1}^{n} P(y_i|x_i;w) \tag{7}$$

$$= \prod_{i=1}^{i} (h_w(x_i)))^{y_i}(1-h_w(x_i))^{(1-y_i)} \tag{8}$$

$$l(w) = log(L(w)) \tag{9}$$

$$= \sum_{i=1}^{n} [y_i log(h_w(x_i)) + (1-y_i)log(1-h_w(x_i))] \tag{10}$$

(ii) **Gradient**:

$$\frac{\partial l(w)}{\partial w_j} = \sum_{i=1}^{n} \left[y_i \frac{1}{h_w(x_i)} \frac{\partial}{\partial w_j} h_w(x_i) \right] \tag{11}$$

$$- \left[(1-y_i) \frac{1}{1-h_w(x_i)} \frac{\partial}{\partial w_j} h_w(x_i) \right] \tag{12}$$

$$\frac{\partial}{\partial w_j} h_w(x_i) = h_w(x_i)\left(1-h_w(x_i)\right) \frac{\partial}{\partial w_j}\left(w^T x_i\right) \tag{13}$$

$$\frac{\partial l(w)}{\partial w_j} = \sum_{i=1}^{n} \left[y_i x_{ij}\left(1-h_w(x_i)\right) - (1-y_i)\, x_{ij} h_w(x_i) \right] \tag{14}$$

$$= \sum_{i=1}^{n} \left[x_{ij}\left(y_i - h_w(x_i)\right) \right] \tag{15}$$

$$\nabla_w l(w) = [\frac{\partial l(w)}{\partial w_1} \frac{\partial l(w)}{\partial w_2} \cdots \frac{\partial l(w)}{\partial w_n}]^T \tag{16}$$

$$= X^T.(YV - HV) \tag{17}$$

Here X is design matrix of [x...] , YV is vector containing target values from data point 1 to n , HV is vector containing predicted values from data point 1 to n

(iii) **Hessian Matrix**: $(H_{l(w)})$

$$H_{ij} = \frac{\partial^2 l(w)}{\partial w_i \partial w_j} \tag{18}$$

$$= \frac{\partial}{\partial w_i}(\sum_{k=1}^{n} [x_{kj}\left(y_k - h_w(x_k)\right)]) \tag{19}$$

$$= \sum_{k=1}^{n} [-x_{kj} x_{ki} h_w(x_k)\left(1-h_w(x_k)\right)] \tag{20}$$

$$= -X^T D X \tag{21}$$

where D is diagonal matrix with $D_{ii} = h_w(x_i)(1-h_w(x_i))$

(iv) **Update equation**:

$$w := w + H_{l(w)}^{-1} \nabla_w l(w) \tag{22}$$

(v) **Algorithm:**

Algorithm 1: Gradient function

Data: w, x, y
Result: The gradient vector
foreach j *in* 1 *to* n **do**
  $partial\_derivative_j \leftarrow \sum_{i=1}^{n} x_{ij}(y_i - h_w(x_i))$
end
return $[partial\_derivative_1, \ldots, partial\_derivative_n]^T$

Algorithm 2: Hessian matrix

Data: w, x, y
Result: The Hessian matrix of the function at w

Initialize an empty square matrix H with dimensions $n \times n$;
foreach i *in* 1 *to* n **do**
  foreach j *in* 1 *to* n **do**
    `// Compute each element of the Hessian matrix`
    `//` x_{ij} `is the` j`-th component of` x_i
    $H_{ij} \leftarrow \sum_{k=1}^{n}[-x_{kj}x_{ki}h_w(x_k)(1 - h_w(x_k))]$;
  end
end
return H;

Algorithm 3: Newton-Raphson iterative method

Data: initial_guess, x , y , $constant_1$, $constant_2$
iterations $\leftarrow constant_1$
convergence $\leftarrow constant_2$
$w \leftarrow$ initial_guess
i $\leftarrow$ 1
$gradient_{new} \leftarrow$ gradient(w, x ,y)
$H \leftarrow$ hessian(w, x , y)
while $gradient_{new} > convergence$ *and* $i <= iterations$ **do**
  $w^{i+1} \leftarrow w^i + \text{inverse}(H) \cdot gradient_{new}$
  $i \leftarrow i + 1$
  $gradient_{new} \leftarrow$ gradient(w^i, x , y)
  $H \leftarrow$ hessian(w^i, x , y)
end
return w^i

(b) Let us start by evaluating the updated parameter for each iteration.

$$
\begin{aligned}
w^{i+1} &= w^i - \text{inverse}(H) \cdot gradient_{new} && (23)\\
&= w^i - (X^T DX)^{-1} X^T (YV - HV) && (24)\\
&= (X^T DX)^{-1}(X^T DX)w^i - (X^T DX)^{-1} X^T (YV - HV) && (25)\\
&= (X^T DX)^{-1}((X^T DX)w^i - X^T (YV - HV)) && (26)\\
&= (X^T DX)^{-1}(X^T DP) && (27)
\end{aligned}
$$

where $P = Xw^i - D^{-1}(YV - HV)$
D is invertiable matrix since diagonals are non-zero since $w^T x_i$

(i) The solution is similar to solution for **weighted least squares ($(\Phi^T R\Phi)^{-1}(\Phi^T Rt)$) in problem 3c**

(ii) This solution for logistic regression has weights for it error function which are dependent on parameter and data.

(iii) these weights are changed for every iteration D matrix changes as w changes

(iv) these are the reason it is called re weighted iterative method.

(c) Error function of a logistic regression is given by

$$Err(h_w, y) = -l(w) \tag{28}$$

Error function is convex if its hessian matrix is semi-definite positive

$$\begin{aligned} H_{Err(w)} &= -H_{l(w)} && (29) \\ &= X^T D X && (30) \\ w H_{Err(w)} w^T &= w X^T D X w^T && (31) \\ &= (w X^T D^{1/2})(w X^T D^{1/2})^T && (32) \\ &= ||w X^T D^{1/2}||^2 && (33) \end{aligned}$$

Norm of vector / matrix is always greater than or equal to zero So Hessian matrix is positive semi-definite .therefore Error function is convex function on w.