1 GLM Part-A

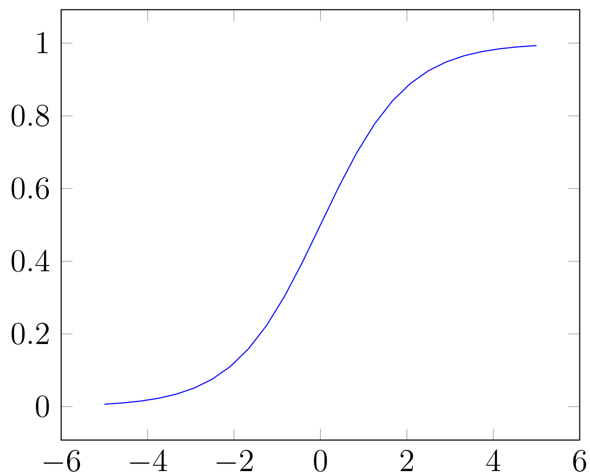
Binary Logistic Regression:
logit(y_i)=$x_i'\beta$, where $y_i \sim$ Bernoulli(w_i), $w_i = \frac{1}{1+exp(-x_i'\beta)}$.

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

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