20MT30017
Millon Madhur Das

First Subtask

2021-07-24
millonmadhurdas@gmail.com

The given dataset comprises of m examples, one example $\left(x^{(i)}, y^{(i)}\right)$ per row. In particular, the i^{th} row contains columns $x_0^{(i)} \in \mathbb{R}, x_1^{(i)} \in \mathbb{R}$, i.e. $x^{(i)} \in \mathbb{R}^2$ and $y^{(i)} \in \{0,1\}$.
The average empirical loss function for Logistic Regression,

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)$$

where $y^{(i)} \in \{0,1\}, h_\theta(x) = g\left(\theta^T x\right)$ and $g(z) = \frac{1}{1+e^{-z}}$ and θ is an n dimensional parameter vector.

We have to find the Hessian matrix H of the empirical loss function with respect to θ, and show that the hessian H is positive semi-definite in nature.

Rewriting $h_\theta\left(x^{(i)}\right)$ and loss function in terms of weight ω,

$$h_\theta\left(x^{(i)}\right) \equiv \sigma(\omega^T x^{(i)}) = \frac{1}{1+e^{-z_i}} = \alpha^{(i)}$$

$$mJ(\omega) = -\sum_{i=1}^{m} y^{(i)}\log\left(\alpha^{(i)}\right) + \left(1-y^{(i)}\right)\log\left(1-\alpha^{(i)}\right)$$

Some helper partial derivatives,

$$\frac{\partial \log \alpha^{(i)}}{\partial \omega^{(i)}} = \frac{1}{\alpha^{(i)}}\frac{\partial \alpha^{(i)}}{\partial \omega^{(i)}} = \frac{1}{\alpha^{(i)}}\frac{\partial \alpha^{(i)}}{\partial z_i}\frac{\partial z_i}{\partial \omega^{(i)}} = (1-\alpha^{(i)})x^{(i)}$$

$$\frac{\partial \log(1-\alpha^{(i)})}{\partial \omega^{(i)}} = \frac{1}{1-\alpha^{(i)}}\frac{\partial(1-\alpha^{(i)})}{\partial \omega^{(i)}} = -\alpha^{(i)}x^{(i)}$$

Computing the partial derivative of $J(\omega)$ wrt. ω

$$\begin{aligned} m\frac{\partial J(\omega)}{\partial \omega^{(j)}} &= -\sum_{i=1}^{m} y^{(i)}x^{(i)}(1-\alpha^{(i)}) - (1-y^{(i)})x_j^{(i)}\alpha^{(i)} \\ &= \sum_{i=1}^{m}(\alpha^{(i)}-y^{(i)})x_j^{(i)} \end{aligned}$$

Representing the above result in matrix form, $m\nabla_\omega J = A^T(\alpha - Y)$, where $A = \begin{pmatrix} x_0^{(1)} & x_1^{(1)} \\ . & . \\ . & . \\ . & . \\ x_0^{(m)} & x_1^{(m)} \end{pmatrix}$

The partial derivative of α wrt ω,

$$\frac{\partial \alpha^{(i)}}{\partial \omega^{(k)}} = x_k^{(i)}\alpha^{(i)}(1-\alpha^{(i)})$$

Computing the Hessian,

$$\begin{aligned} H = m\frac{\partial^2 J(\omega)}{\partial \omega^{(j)}\partial \omega^{(k)}} &= \sum_{i=1}^{m} x_j^{(i)}\frac{\partial \alpha^{(i)}}{\partial \omega^{(k)}} \\ &= \sum_{i=1}^{m} x_j^{(i)}x_k^{(i)}\alpha^{(i)}(1-\alpha^{(i)}) \\ &= (Z^{(j)})^T B Z^{(k)} \end{aligned}$$

where, $(Z^{(j)})^T = (x_j^{(1)}, x_j^{(2)}, ..., x^{(m)})_j$, $x^{(j)} = (x_0^{(j)}, x_1^{(j)})^T$ and

$$B = \begin{pmatrix} \alpha^{(1)}(1-\alpha^{(1)}) & 0 & 0 & ... \\ 0 & \alpha^{(2)}(1-\alpha^{(2)}) & 0 & ... \\ ... & & & \\ 0 & 0 & ... & \alpha^{(m)}(1-\alpha^{(m)}) \end{pmatrix}$$

Notice, B is a diagonal matrix with all diagonal entries positive and less than 1, as $\alpha^{(i)} \equiv \frac{1}{1+e^{-z_i}} < 1$. Similarly, $(1-\alpha^{(i)}) < 1$

Finally representing the Hessian in matrix form,

$$\begin{aligned} H = m\nabla_\omega^2 J &= A^T B A \\ &= A^T B^{\frac{1}{2}} B^{\frac{1}{2}} A \\ &= (B^{\frac{1}{2}} A)^T (B^{\frac{1}{2}} A) \\ &= K^T K \end{aligned}$$

Notice $(K^T K)^T = K^T K$, i.e. H is symmetric. As $|K^T K| = |K|^2 \geq 0$, the Hessian H is positive semi-definite.

Alternate proof for positive semi-definite,

$$\text{H} = \begin{pmatrix} x_0^{(1)} & x_0^{(2)} & ... & x_0^{(m)} \\ x_1^{(1)} & x_1^{(2)} & ... & x_1^{(m)} \end{pmatrix} \begin{pmatrix} \alpha^{(1)}(1-\alpha^{(1)}) & 0 & 0 & ... \\ 0 & \alpha^{(2)}(1-\alpha^{(2)}) & 0 & ... \\ ... & & & \\ 0 & 0 & ... & \alpha^{(m)}(1-\alpha^{(m)}) \end{pmatrix} \begin{pmatrix} x_0^{(1)} & x_1^{(1)} \\ x_0^{(2)} & x_1^{(2)} \\ . & . \\ . & . \\ . & . \\ x_0^{(m)} & x_1^{(m)} \end{pmatrix}$$

$$= \begin{pmatrix} x_0^{(1)} & x_0^{(2)} & ... & x_0^{(m)} \\ x_1^{(1)} & x_1^{(2)} & ... & x_1^{(m)} \end{pmatrix} \begin{pmatrix} x_0^{(1)}\alpha^{(1)}(1-\alpha^{(1)}) & x_1^{(1)}\alpha^{(1)}(1-\alpha^{(1)}) \\ x_0^{(2)}\alpha^{(2)}(1-\alpha^{(2)}) & x_1^{(2)}\alpha^{(2)}(1-\alpha^{(2)}) \\ . & . \\ . & . \\ . & . \\ x_0^{(m)}\alpha^{(m)}(1-\alpha^{(m)}) & x_1^{(m)}\alpha^{(m)}(1-\alpha^{(m)}) \end{pmatrix}$$

$$= \begin{pmatrix} (x_0^{(1)})^2\alpha^{(1)}(1-\alpha^{(1)}) + ... + (x_0^{(m)})^2\alpha^{(m)}(1-\alpha^{(m)}) & x_0^{(1)}x_1^{(1)}\alpha^{(1)}(1-\alpha^{(1)}) + ... + x_0^{(m)}x_0^{(m)}\alpha^{(m)}(1-\alpha^{(m)}) \\ x_0^{(1)}x_1^{(1)}\alpha^{(1)}(1-\alpha^{(1)}) + ... + x_0^{(m)}x_0^{(m)}\alpha^{(m)}(1-\alpha^{(m)}) & (x_1^{(1)})^2\alpha^{(1)}(1-\alpha^{(1)}) + ... + (x_1^{(m)})^2\alpha^{(m)}(1-\alpha^{(m)}) \end{pmatrix}$$

As discussed earlier, it is clear that the Hessian is symmetric.
$det(H) = 0$
$tr(H) = 2[(x_0^{(1)})^2\alpha^{(1)}(1-\alpha^{(1)}) + (x_0^{(2)})^2\alpha^{(1)}(1-\alpha^{(1)}) + ... + (x_0^{(m)})^2\alpha^{(m)}(1-\alpha^{(m)})] > 0$

So, the eigenvalues are positive, precisely, $\lambda = \sqrt{tr(H)}$

This establishes that the Hessian is positive semi-definite.