Logit Model

$$P(1|x) = G(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)$$

Where G is a function considering on value strictly between 0 and 1 for all real number ($\mathbb{R}$).

$$G(z) = \frac{exp^{\theta}}{1 + exp^{\theta}}$$

Or,

$$G(z) = \frac{1}{1 + exp^{-\theta}}$$

Just we have learned from sigmoid (σ) function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

In addition, we have learnt from Binomial and Bernoulli's exponential distribution:

That the relationship between μ and θ

$are\ invertible$:

$$\boxed{\mu = \frac{1}{1 + \exp^{-\theta}}}$$

And,

$$\boldsymbol{\exp^{-\theta} = \frac{1-\mu}{\mu}}$$

Or,

$$\mu = \frac{1}{1 + \frac{1}{\exp^{\theta}}}$$

$$=> \mu = \frac{1}{\frac{\exp^{\theta} + 1}{\exp^{\theta}}}$$

$$=> \mu = \frac{\exp^{\theta}}{\exp^{\theta} + 1}$$

$$\boldsymbol{\mu = \frac{\exp^{\theta}}{\exp^{\theta} + 1}}$$

Similarly:

$$\boldsymbol{\exp^{\theta} = \frac{\mu}{1-\mu}}$$

We can even prove it:

$$y = \frac{1}{1 + e^{-x}}$$

Multiplying $log\ on$ both the sides:

$$log(y) = log\left(\frac{1}{1 + e^{-x}}\right)$$

Dividing $1 - y$ in both the sides, we get:

$$log\left(\frac{y}{1 - y}\right) = log\left(\frac{\frac{1}{1 + e^{-x}}}{1 - \frac{1}{1 + e^{-x}}}\right)$$

Taking RHS we get:

$$log\left(\frac{\frac{1}{1 + e^{-x}}}{\frac{1 + e^{-x} - 1}{1 + e^{-x}}}\right)$$

$$=> log\left(\frac{\frac{1}{1 + e^{-x}}}{\frac{e^{-x}}{1 + e^{-x}}}\right)$$

$$=> log\left(\frac{1}{1 + e^{-x}} \times \frac{1 + e^{-x}}{e^{-x}}\right)$$

$$=> log\left(\frac{1}{e^{-x}}\right)$$

$$=> log\left(\frac{1}{\frac{1}{e^{x}}}\right)$$

$$=> log(e^{x})$$

We can write it as:

$$log_{e}(e^{x})$$

We know:

$$log_{a}a^{b} = b \times log_{a}a \; and \; log_{a}a = 1,$$

$$hence \; log_{a}a^{b} = b$$

i.e.

$$log_{e}e^{x} = x \times 1 = x$$

Therefore:

$$x = \; log\left(\frac{y}{1-y}\right)$$

Same from Bernoulli and Binomial exponential distribution we get:

$$\theta = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right),$$

Or,

$$\eta(\theta) = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right).$$

Hence, by logit model we can define it by:

$$\boldsymbol{Logit(p) = log\left(\frac{p}{1 - p}\right)}$$

Where,

$$\boldsymbol{p = probability,}$$

$$\boldsymbol{\frac{p}{1 - p} = corresponding\ odds.}$$

$\mathbf{Hence, it\ is\ known\ as}\ \boldsymbol{log\ of\ odds\ or\ log(odds)}$

i.e.

$$\boldsymbol{Logit(p) = log(odds)}$$

In addition, it strictly follow Bernoulli and Binomial exponential distribution. And it is applicable in logistic regression.

Logistic Regression

Logistic regression uses this $\boldsymbol{log(odds) or\ log\ of\ odds}$ in its equation.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

We may also represent this in many ways:

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

Or,

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

Etc.

We know from $\boldsymbol{log(odds) that\ if}$:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Then,

$$-x = log\left(\frac{p}{1-p}\right)$$

Where,

$p = no.\, success\; and\; 1 - p\;\; represents$
$no.\, of\; failures.$

And $x\; is\; logit\; or\; log(odds).$

We will take $g(z) and\; proceed$:

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

Then:

$$-z = log\left(\frac{g(z)}{1-g(z)}\right)$$

as log odds.

$$g'(z) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right)$$

$$= \frac{\partial}{\partial z}\left((1+e^{-z})^{-1}\right)$$

Applying chain rule of derivative:

$$\frac{\partial f(u)}{\partial x} = \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial x}$$

Here, $f = u^{-1}$ and $u = (1 + e^{-z})$

$$= \frac{\partial}{\partial u}\left(u^{-1}\right) \times \frac{\partial}{\partial z}(u)$$

We know, $\frac{\partial}{\partial a} a^{n} = n \times a^{n-1}, then$:

1. $$\frac{\partial}{\partial u}\left(u^{-1}\right) = \left(-1(u)^{-1-1}\right)$$

$$= -1(u)^{-2}$$

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

2. $$\frac{\partial}{\partial z}(u) = \frac{\partial}{\partial z}(1 + e^{-z})$$

$$= \left(\frac{\partial}{\partial z}(1) + \frac{\partial}{\partial z}(e^{-z})\right)$$

$$= \left(0 + \frac{\partial}{\partial z}(e^{-z})\right)$$

$$= \left(\frac{\partial}{\partial z} (e^{-z}) \right)$$

Applying chain rule of derivative:

$$\frac{\partial f(u)}{\partial x} = \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial x}$$

Here, $f = e^u \; and \; u = -z$

Therefore,

$$1. \frac{\partial}{\partial u}(f) = \frac{\partial}{\partial u}(e^u) = e^u = e^{-z}$$

$$2. \; \frac{\partial}{\partial z}(u) = \frac{\partial}{\partial z}(-z) = -1$$

Therefore,

$$\frac{\partial}{\partial z}(e^{-z}) = (e^{-z}) \times -1 = -e^{-z}$$

Therefore,

$$\frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = -\frac{1}{(1+e^{-z})^2} \times -e^{-z}$$

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

Hence,

$$\boldsymbol{g'(x) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2}}$$

[Alternative Calculation:

$$\boldsymbol{\frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right)}$$

$$\textit{Applying here } \frac{\partial}{\partial x} \times \frac{f(x)}{g(x)}$$
$$= \frac{g(x) \times \frac{\partial}{\partial x} \times f(x) - f(x) \times \frac{\partial}{\partial x} \times g(x)}{\left(g(x)\right)^2}$$

$$\boldsymbol{= \frac{(1+e^{-z}) \times \frac{\partial}{\partial z}(1) - 1 \times \frac{\partial}{\partial z}(1+e^{-z})}{(1+e^{-z})^2}}$$

$$\boldsymbol{= \frac{(1+e^{-z}) \times 0 - 1 \times \frac{\partial}{\partial z}(1+e^{-z})}{(1+e^{-z})^2}}$$

$$= \frac{0 - 1 \times (\frac{\partial}{\partial z}(1) + \frac{\partial}{\partial z}(e^{-z}))}{(1 + e^{-z})^2}$$

$$= \frac{-1 \times (0 + \frac{\partial}{\partial z}(e^{-z}))}{(1 + e^{-z})^2}$$

By chain rule, we already proved that:

$$\frac{\partial}{\partial z}(e^{-z}) = -e^{-z}$$

Therefore,

$$= \frac{-1 \times (0 + (-e^{-z}))}{(1 + e^{-z})^2}$$

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

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

$$g'(x) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2}$$

]

We know,

$$-z = log\left(\frac{g(z)}{1-g(z)}\right)$$

Then, applying it in above equation we get:

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

We can rewrite it as:

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

$$= \frac{\frac{g(z)}{1-g(z)}}{\left(1+\frac{g(z)}{1-g(z)}\right)^2}$$

$$= \frac{\frac{\boldsymbol{g(z)}}{\mathbf{1} - \boldsymbol{g(z)}}}{\left(\frac{\mathbf{1} - \boldsymbol{g(z)} + \boldsymbol{g(z)}}{\mathbf{1} - \boldsymbol{g(z)}}\right)^{\mathbf{2}}}$$

$$= \frac{\frac{\boldsymbol{g(z)}}{\mathbf{1} - \boldsymbol{g(z)}}}{\left(\frac{\mathbf{1}}{\mathbf{1} - \boldsymbol{g(z)}}\right)^{\mathbf{2}}}$$

$$= \frac{\boldsymbol{g(z)}}{\mathbf{1} - \boldsymbol{g(z)}} \times \left(\mathbf{1} - \boldsymbol{g(z)}\right)^{\mathbf{2}}$$

$$= g(z) \times \left(\mathbf{1} - \boldsymbol{g(z)}\right)$$

[Alternative Calculation:

$$-\boldsymbol{z} = \boldsymbol{log}\left(\frac{\boldsymbol{g(z)}}{\mathbf{1} - \boldsymbol{g(z)}}\right)$$

Or,

$$\boldsymbol{z} = -\boldsymbol{log}\left(\frac{\boldsymbol{g(z)}}{\mathbf{1} - \boldsymbol{g(z)}}\right)$$

Then, applying it in above equation we get:

$$\frac{\boldsymbol{e^{-z}}}{(\boldsymbol{1}+\boldsymbol{e^{-z}})^{\boldsymbol{2}}}=\frac{\boldsymbol{e}^{-\left(-\boldsymbol{log}\left(\frac{\boldsymbol{g(z)}}{\boldsymbol{1-g(z)}}\right)\right)}}{\left(\boldsymbol{1}+\boldsymbol{e}^{-\left(-\boldsymbol{log}\left(\frac{\boldsymbol{g(z)}}{\boldsymbol{1-g(z)}}\right)\right)}\right)^{\boldsymbol{2}}}$$

$$=\frac{\frac{1}{e^{\left(-\log\left(\frac{\boldsymbol{g(z)}}{\boldsymbol{1-g(z)}}\right)\right)}}}{\left(1+\frac{1}{e^{\left(-\log\left(\frac{\boldsymbol{g(z)}}{\boldsymbol{1-g(z)}}\right)\right)}}\right)^{2}}$$

$$=\frac{\frac{1}{\frac{1}{\boldsymbol{e}^{\boldsymbol{log}\left(\frac{\boldsymbol{g(z)}}{\boldsymbol{1-g(z)}}\right)}}}}{\left(1+\frac{1}{\frac{1}{\boldsymbol{e}^{\boldsymbol{log}\left(\frac{\boldsymbol{g(z)}}{\boldsymbol{1-g(z)}}\right)}}}\right)^{2}}$$

We can rewrite it as:

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

$$= \frac{\frac{1}{\frac{1}{\frac{g(z)}{1-g(z)}}}}{\left(1 + \frac{1}{\frac{1}{\frac{g(z)}{1-g(z)}}}\right)^2}$$

$$= \frac{\frac{1}{\frac{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}{\boldsymbol{g}(\boldsymbol{z})}}}{\left(1 + \frac{1}{\frac{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}{\boldsymbol{g}(\boldsymbol{z})}}\right)^2}$$

$$= \frac{\frac{\boldsymbol{g}(\boldsymbol{z})}{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}}{\left(1 + \frac{\boldsymbol{g}(\boldsymbol{z})}{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}\right)^{\mathbf{2}}}$$

$$= \frac{\frac{\boldsymbol{g}(\boldsymbol{z})}{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}}{\left(\frac{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z}) + \boldsymbol{g}(\boldsymbol{z})}{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}\right)^{\mathbf{2}}}$$

$$= \frac{\frac{\boldsymbol{g}(\boldsymbol{z})}{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}}{\left(\frac{\mathbf{1}}{\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})}\right)^{\mathbf{2}}}$$

$$= \frac{g(z)}{1 - g(z)} \times \left(1 - g(z)\right)^2$$

$$= g(z) \times \left(1 - g(z)\right)$$

Hence,

$$g'(z) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2} = g(z) \times \left(1 - g(z)\right)$$

]

Some Notes on Logistic Regression

1.From Linear Regression we have learnt that:

$$y = \beta_0 + \beta_1 x + \epsilon$$
$$E(y) = y\hat{} = \beta_0 + \beta_1 x$$

Then for logistic regression:

$$\sigma(\beta_0 + \beta_1 x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$Similarly\ for\ multiple\ linear\ regression$

$: y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + .. + \beta_p x_{pi} + \epsilon_i$

Therefore estimation of $y\ or\ y\ hat,$

$$E(y_i) = \hat{y}_i$$
$$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + .. + \beta_p x_{pi}$$

Then,

$$\beta^T = [\beta_0, \beta_1, \beta_2, .., \beta_p]$$
$$x = [1, x_1, x_2, .., x_p]$$

Note we can write $\beta\ in\ matrix\ format$:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ . \\ \beta_p \end{pmatrix}$$

$$\beta^T = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ . \\ \beta_p \end{pmatrix}^T$$

$$= [\boldsymbol{\beta_0}\ \boldsymbol{\beta_1}\ \boldsymbol{\beta_2} \dots \boldsymbol{\beta_p}]$$

i.e. β^T *or beta transpose is row vector of* $\boldsymbol{\beta}$

and , x is inputs attached with β *(beta)matrix while*

i.e.$x = [1\ , x_{1,} x_2, \dots , x_p]$, while

β *to be considered as weights.*

i.e.

$$\boldsymbol{\beta^T x} = \begin{pmatrix} \boldsymbol{\beta_0} \\ \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ . \\ \boldsymbol{\beta_p} \end{pmatrix}^T \begin{pmatrix} \mathbf{1} \\ \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ . \\ \boldsymbol{x_p} \end{pmatrix}$$

$$= (\boldsymbol{\beta_0} + \boldsymbol{\beta_1 x_1} + \boldsymbol{\beta_2 x_2} + .. + \boldsymbol{\beta_p x_p})$$

$$= \boldsymbol{y\hat{}} = \boldsymbol{E(y) i.e. Estimation\ of\ y}.$$

We have already studied during analysing of $\boldsymbol{\sigma\ sigmoid\ function} =$

$\boldsymbol{y\hat{} = \sigma(x) = \sigma(w_0 x_0 + w_1 x_1 + \cdots + w_n x_n), while\ x_0 = 1}$

$\boldsymbol{and\ w\ are\ weights\ , here\ w = \ \beta.}$

Or,

$$\boldsymbol{\sigma}(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2 + .. + \boldsymbol{\beta}_p \boldsymbol{x}_p) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2 + .. + \boldsymbol{\beta}_p \boldsymbol{x}_p)}}$$

$$= \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\boldsymbol{\beta}^T \boldsymbol{x})}}$$

Here, as we are considering $\boldsymbol{g}(\boldsymbol{z})\boldsymbol{and}\ \boldsymbol{z} = \boldsymbol{\beta}^T \boldsymbol{x}$,

$$\boldsymbol{then}, \boldsymbol{g}(\boldsymbol{\beta}^T \boldsymbol{x}) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\boldsymbol{\beta}^T \boldsymbol{x})}}$$

2. $\boldsymbol{\sigma}\ (\boldsymbol{sigmoid\ function})\boldsymbol{or}\ \boldsymbol{g}\ \boldsymbol{here\ is\ called}$ $\boldsymbol{Activation\ Function}$

Why Activation Function? Because it takes the input $\boldsymbol{\beta}^T \boldsymbol{x}$ and put into the equation:

$$\boldsymbol{\sigma}(\boldsymbol{\beta}^T \boldsymbol{x}) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\boldsymbol{\beta}^T \boldsymbol{x})}}$$

More precisely the equation suggests:

$$\boldsymbol{p}^{\wedge} = \boldsymbol{h}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{\sigma}(\boldsymbol{\beta}^T \boldsymbol{x})$$

Now,

$\boldsymbol{h_\beta(x)\ represents\ y^{\wedge}\ , estimated\ of\ y}$.

$\boldsymbol{Moreover\ , h_\beta(x)\ can\ also\ be\ simply\ represented\ as\ h(x)}$.

And $p^{\wedge}\ (or\ p\ hat)\ represents\ estimated\ probability$.

i.e.

$$\boldsymbol{p^{\wedge} = h_\beta(x) or\ h(x) = \sigma(\beta^T x)}$$

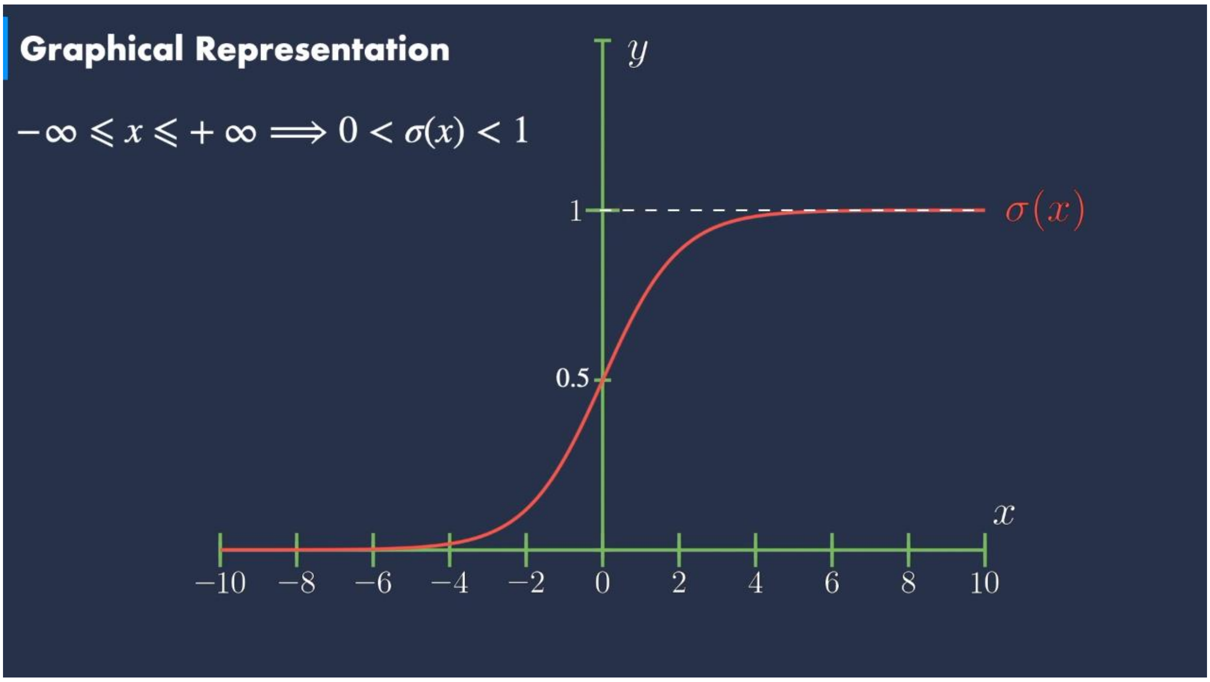
Or if we represent $\boldsymbol{\sigma(x)\ as\ g(z), then}$:

$$\boldsymbol{p^{\wedge} = h_\beta(x) or\ h(x) = g(\beta^T x)}$$

3.Relationship between estimated y ($y^{\wedge}$) and estimated probability ($p^{\wedge}$):

$$\boldsymbol{y^{\wedge} = \begin{cases} 0, & p^{\wedge} < 0.5 \\ 1, & p^{\wedge} \geq 0.5 \end{cases}}$$

Forming a graph:



Where,

$$-\infty \leq x \leq +\infty \text{ then } 0 < \sigma(x) < 1$$

Then,

If:

$$\boldsymbol{\sigma(x) = \frac{1}{1+e^{-x}}}$$

Then:

$$\boldsymbol{\sigma(x) = \begin{cases} 0, & x \to -\infty \\ 1, & x \to \infty \end{cases}}$$

Or if $\boldsymbol{\sigma(x) = g(z), then}$:

$$\boldsymbol{g(z) = \begin{cases} 0, & x \to -\infty \\ 1, & x \to \infty \end{cases}}$$

4. Cost function of one training instance.

We know,

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Then,

$$-x = log\left(\frac{p^{\wedge}}{1 - p^{\wedge}}\right)$$

Or,

$$x = -log\left(\frac{p^{\wedge}}{1 - p^{\wedge}}\right)$$

Then, $x = \left(-log(p^{\wedge}) - \left(-log(1 - p^{\wedge})\right)\right)$

$$if\ \beta = \theta\ then\ , c(\theta) = \begin{cases} -\log(p^{\wedge}), & if\ y = 1 \\ -\log(1 - p^{\wedge}), & if\ y = 0 \end{cases}$$

Where, $p\ is\ probability\ and\ p^{\wedge}\ is\ estimated\ probability\ and\ c(\theta)\ represents\ cost\ function\ of\ one\ training\ instance.$

5. Cost function of entire training set:

$$J(\boldsymbol{\theta}) = -\frac{1}{m}\prod_{i=1}^{m} p^{\wedge(i)^{y^{(i)}}} \times \left(1 - p^{\wedge(i)}\right)^{1-y^{(i)}}$$

Remember **Bernoulli's exponential family distribution**:

Putting log we get:

$$= -\frac{1}{m}\prod_{i=1}^{m} log\left(p^{\wedge(i)^{y^{(i)}}} \times \left(1 - p^{\wedge(i)}\right)^{1-y^{(i)}}\right)$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} log\left(p^{\wedge(i)}\right) + 1 - y^{(i)} log\left(1 - p^{\wedge(i)}\right)\right)$$

Note here is a portion that belongs to calculate likelihood and maximum likelihood of logistic regression:

Likelihood

$$P(Y|X, \boldsymbol{\beta}) = \prod_{i=1}^{n}\left(p^{\wedge(i)^{y^{(i)}}} \times \left(1 - p^{\wedge(i)}\right)^{1-y^{(i)}}\right)$$

We know: $p^{\wedge} = h_{\beta}(x)$

Therefore,

$$P(Y|X,\beta) = \prod_{i=1}^{n}\left(h_\beta(x)^{y^{(i)}} \times \left(1 - h_\beta(x)\right)^{1-y^{(i)}}\right)$$

And,

$$h_\beta(x) = \sigma\left(\beta^T x\right) = \frac{1}{1 + e^{-(\beta^T x)}}$$

Or,

$$g\left(\beta^T x\right) = \frac{1}{1 + e^{-(\beta^T x)}}$$

And,

$$\beta_j = \beta_j + \alpha \nabla_{\beta_j} l(\beta_j)$$

[∇ known as nabla and update occurs according to the above equation.]

Where,

$$\beta_j = initial\ beta(\beta).$$

$$\alpha = learning\ rate.$$

$$\nabla_{\beta_j} = partial\ derivative\ of\ \beta.$$

$$l(\beta_j) = log\ likelihood\ of\ \beta$$

Therefore,

$$l(\beta) = P(Y|X, \beta)$$

$$= \prod_{i=1}^{n} log\left(h_\beta(x)^{y^{(i)}} \times \left(1 - h_\beta(x)\right)^{1-y^{(i)}}\right)$$

$$= \sum_{i=1}^{n} log\left(h_\beta(x)^{y^{(i)}}\right) + log\left(\left(1 - h_\beta(x)\right)^{1-y^{(i)}}\right)$$

$$= \sum_{i=1}^{n} \left(y^{(i)} log\left(h_\beta(x)\right) + 1 - y^{(i)} log\left(1 - h_\beta(x)\right)\right)$$

Maximization of $log\ likelihood\ of\ \beta(beta)$

Now partial derivative of $l(\beta), log\ likelihood\ of\ \beta$.

$$\frac{\partial}{\partial(\beta)}\left(l(\beta)\right)$$

$$= \frac{\partial}{\partial(\beta)}\left(\sum_{i=1}^{n}\left(y^{(i)}log\left(h_\beta(x)\right) + 1 - y^{(i)}log\left(1 - h_\beta(x)\right)\right)\right)$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial(\beta)}\left(\left(y^{(i)}log\left(h_\beta(x)\right) + 1 - y^{(i)}log\left(1 - h_\beta(x)\right)\right)\right)$$

Now,

$$let\ A = y^{(i)}log\left(h_\beta(x)\right)\ and\ B = 1 - y^{(i)}log\left(1 - h_\beta(x)\right)$$

Therefore ,

$$= \sum_{i=1}^{n}\frac{\partial}{\partial(\beta)}(A + B)$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial(\beta)}(A) + \frac{\partial}{\partial(\beta)}(B)$$

Now taking:

$$\frac{\partial}{\partial(\beta)}(A) = \frac{\partial}{\partial(\beta)}\left(y^{(i)}log\left(h_\beta(x)\right)\right)$$

$$= log\left(h_\beta(x)\right) \times \frac{\partial}{\partial(\beta)}\left(y^{(i)}\right) + \left(y^{(i)}\right) \times \frac{\partial}{\partial(\beta)} \times log\left(h_\beta(x)\right)$$

Putting $h_\beta(x) = g\left(\beta^T x\right) we\ get$:

$$= log\left(g\left(\beta^T x\right)\right) \times \frac{\partial}{\partial(\beta)}\left(y^{(i)}\right) + \left(y^{(i)}\right) \times \frac{\partial}{\partial(\beta)} \times log\left(g\left(\beta^T x\right)\right)$$

$$= log\left(g\left(\beta^T x\right)\right) \times 0 + \left(y^{(i)}\right) \times \frac{\partial}{\partial(\beta)} \times log\left(g\left(\beta^T x\right)\right)$$

$$= 0 + \left(y^{(i)}\right) \times \frac{\partial}{\partial(\beta)} \times log\left(g\left(\beta^T x\right)\right)$$

$$= \left(y^{(i)}\right) \times \frac{\partial}{\partial(\beta)} \times log\left(g\left(\beta^T x\right)\right)$$

Now,

$$\frac{\partial}{\partial(\beta)} \times log\left(g\left(\beta^T x\right)\right)$$

Applying chain rule of derivative:

$$\frac{\partial f(u)}{\partial x} = \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial x}$$

Here, $f = log(u)\ and\ u = g(\beta^T x)$

$$= \frac{\partial}{\partial u}(log(u)) \times \frac{\partial}{\partial \beta} g\left(\beta^T x\right)$$

$$= \frac{\partial}{\partial g(\beta^T x)}\left(log\left(g(\beta^T x)\right)\right) \times \frac{\partial}{\partial \beta} g\left(\beta^T x\right)$$

$$= \frac{1}{g\left(\beta^T x\right)} \times \frac{\partial}{\partial \beta} g\left(\beta^T x\right)$$

Applying chain rule of derivative:

$$\frac{\partial f(u)}{\partial x} = \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial x}$$

Here, $f = g(u)\ and\ u = \beta^T x$

$$\frac{\partial f(u)}{\partial \beta} = \frac{\partial}{\partial u}\big(g(u)\big) \times \frac{\partial}{\partial \beta}(u)$$

$$= \frac{\partial}{\partial \beta^T x}(g(\beta^T x)) \times \frac{\partial}{\partial \beta}(\beta^T x)$$

$$= \frac{\partial}{\partial \beta^T x}(g(\beta^T x)) \times (x \times \frac{\partial}{\partial \beta}\big(\beta^T\big) + \big(\beta^T\big) \times \frac{\partial}{\partial \beta}(x)$$

$$= \frac{\partial}{\partial \beta^T x}\Big(g\big(\beta^T x\big)\Big) \times \big(x \times 1 + \big(\beta^T\big) \times 0\big)$$

$$= \frac{\partial}{\partial \beta^T x}\Big(g\big(\beta^T x\big)\Big) \times (x)$$

$$= g'\big(\beta^T x\big) \times (x)$$

we know

$$g(z)\ where\ z = \beta^T x\ ,\ then\ ,$$

$$g'(z) = \ g(z) \times \left(1 - g(z)\right)$$

$$Now, \frac{\partial}{\partial \beta} g\left(\beta^T x\right) = \frac{\partial}{\partial \beta^T x}\left(g\left(\beta^T x\right)\right) \times (x)$$

$$=> \frac{\partial}{\partial \beta} g(z) = \frac{\partial}{\partial z}\left(g(z)\right) \times (x)$$

$$= \ g^{'}(z) \times (x)$$

$$= g(z) \times \left(1 - g(z)\right) \times x$$

Now if $x\ is\ continuous\ of\ jth\ term\ then$

$$= g(z) \times \left(1 - g(z)\right) \times x_j$$

Now,

$$\frac{\partial}{\partial(\beta)}(A) = \frac{\partial}{\partial(\beta)}\left(y^{(i)} log\left(h_\beta(x)\right)\right)$$

We get,

$$\frac{\partial}{\partial(\beta)}\left(y^{(i)} log\left(h_\beta(x)\right)\right) = \frac{\partial}{\partial(\beta)}\left(y^{(i)} log\left(g(\beta^T x)\right)\right)$$

$$= \frac{(\left(y^{(i)}\right) \times (g(z) \times \left(1 - g(z)\right)) \times x_j}{g\left(\beta^T x\right)}$$

Now we can put either $\beta^T x = z \; or \; z = \beta^T x$, we proceeding with $z = \beta^T x \; we \; get$:

$$= \frac{(\left(y^{(i)}\right) \times g\left(\beta^T x\right) \times \left(1 - g\left(\beta^T x\right)\right) \times x_j}{g\left(\beta^T x\right)}$$

$$= \left(\left(y^{(i)}\right) \times \left(1 - g\left(\beta^T x\right)\right)\right) \times x_j$$

$$\frac{\partial}{\partial(\beta)}\left(y^{(i)} log\left(h_\beta(x)\right)\right) = \frac{\partial}{\partial(\beta)}\left(y^{(i)} log\left(g(\beta^T x)\right)\right)$$

$$= \frac{(\left(y^{(i)}\right) \times (g(z) \times \left(1 - g(z)\right)) \times x_j}{g\left(\beta^T x\right)}$$

$$= \left(\left(y^{(i)}\right) \times \left(1 - g\left(\beta^T x\right)\right)\right) \times x_j$$

$$\frac{\partial}{\partial(\beta)}(B) = \frac{\partial}{\partial(\beta)}\left(1 - y^{(i)} \times log\left(1 - h_\beta(x)\right)\right)$$

$$= \left(1 - y^{(i)} \times \frac{\partial}{\partial(\beta)} log\left(1 - h_\beta(x)\right) + log\left(1 - h_\beta(x)\right) \times \frac{\partial}{\partial(\beta)}\left(1 - y^{(i)}\right)\right)$$

$$= \left(\left(1 - y^{(i)} \times \frac{\partial}{\partial(\beta)} log\left(1 - h_\beta(x)\right) + log\left(1 - h_\beta(x)\right) \times \left(\frac{\partial}{\partial(\beta)}(1) - \frac{\partial}{\partial(\beta)} y^{(i)}\right)\right)\right)$$

$$= \left(\left(1 - y^{(i)} \times \frac{\partial}{\partial(\beta)} log\left(1 - h_\beta(x)\right) + log\left(1 - h_\beta(x)\right) \times (0 - 0)\right)\right)$$

$$= \left(\left(1 - y^{(i)} \times \frac{\partial}{\partial(\beta)} log\left(1 - h_\beta(x)\right) + log\left(1 - h_\beta(x)\right) \times (0)\right)\right)$$

$$=\left(\left(1-y^{(i)}\times\frac{\partial}{\partial(\beta)}log\left(1-h_\beta(x)\right)+0\right)\right)$$

$$=\left(\left(1-y^{(i)}\times\frac{\partial}{\partial(\beta)}log\left(1-h_\beta(x)\right)\right)\right)$$

Now, $h_\beta(x)=g(\beta^T x)$

$$=\left(1-y^{(i)}\times\frac{\partial}{\partial(\beta)}\left(log\left(1-g(\beta^T x)\right)\right)\right)$$

Now,

$$\frac{\partial}{\partial(\beta)}\left(log\left(1-g(\beta^T x)\right)\right)$$

Applying chain rule of derivative:

$$\frac{\partial f(u)}{\partial x}=\frac{\partial f}{\partial u}\times\frac{\partial u}{\partial x}$$

Here, $f=log(u) and\ u=1-g(\beta^T x)$

$$\frac{\partial f(u)}{\partial\beta}=\frac{\partial}{\partial u}\left(log(u)\right)\times\frac{\partial}{\partial\beta}(u)$$

$$= \frac{\partial}{\partial\left(1 - g\left(\beta^T x\right)\right)}\left(log\left(1 - g\left(\beta^T x\right)\right)\right)$$
$$\times \frac{\partial}{\partial\beta}\left(1 - g\left(\beta^T x\right)\right)$$

$$= \frac{1}{\left(1 - g\left(\beta^T x\right)\right)} \times \frac{\partial}{\partial\beta}\left(1 - g\left(\beta^T x\right)\right)$$

$$= \frac{1}{\left(1 - g\left(\beta^T x\right)\right)} \times \left(\frac{\partial}{\partial\beta}(1) - \frac{\partial}{\partial\beta}\left(g\left(\beta^T x\right)\right)\right)$$

$$= \frac{1}{\left(1 - g\left(\beta^T x\right)\right)} \times \left(0 - \frac{\partial}{\partial\beta}\left(g\left(\beta^T x\right)\right)\right)$$

$$= \frac{1}{\left(1 - g\left(\beta^T x\right)\right)} \times \left(-\frac{\partial}{\partial\beta}\left(g\left(\beta^T x\right)\right)\right)$$

Now let’s solve,

$$\frac{\partial}{\partial\beta}\left(g\left(\beta^T x\right)\right)$$

Applying chain rule of derivative:

$$\frac{\partial f(u)}{\partial x} = \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial x}$$

Here, $f = g(u)\ and\ u = \beta^T x$

$$\frac{\partial f(u)}{\partial \beta} = \frac{\partial}{\partial u}\big(g(u)\big) \times \frac{\partial}{\partial \beta}(u)$$

$$= \frac{\partial}{\partial \beta^T x}(g(\beta^T x)) \times \frac{\partial}{\partial \beta}(\beta^T x)$$

$$= \frac{\partial}{\partial \beta^T x}(g(\beta^T x)) \times (x \times \frac{\partial}{\partial \beta}\left(\beta^T\right) + \left(\beta^T\right) \times \frac{\partial}{\partial \beta}(x)$$

$$= \frac{\partial}{\partial \beta^T x}\left(g\left(\beta^T x\right)\right) \times \left(x \times 1 + \left(\beta^T\right) \times 0\right)$$

$$= \frac{\partial}{\partial \boldsymbol{\beta}^T \boldsymbol{x}} \left(\boldsymbol{g}(\boldsymbol{\beta}^T \boldsymbol{x}) \right) \times (\boldsymbol{x})$$

$$= \boldsymbol{g}'(\boldsymbol{\beta}^T \boldsymbol{x}) \times (\boldsymbol{x})$$

we know

$\boldsymbol{g}(\boldsymbol{z})$ ***where*** $\boldsymbol{z} = \boldsymbol{\beta}^T \boldsymbol{x}$, ***then*** ,

$\boldsymbol{g}'(\boldsymbol{z}) = \ g(z) \times \left(\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})\right)$

$\boldsymbol{Now,} \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{g}\left(\boldsymbol{\beta}^T \boldsymbol{x}\right) = \ \frac{\partial}{\partial \boldsymbol{\beta}^T \boldsymbol{x}} \left(\boldsymbol{g}(\boldsymbol{\beta}^T \boldsymbol{x}) \right) \times (\boldsymbol{x})$

$=> \ \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{g}(\boldsymbol{z}) = \ \frac{\partial}{\partial \boldsymbol{z}} \left(\boldsymbol{g}(\boldsymbol{z}) \right) \times (\boldsymbol{x})$

$= \ \boldsymbol{g}'(\boldsymbol{z}) \times (\boldsymbol{x})$

$= g(z) \times \left(\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})\right) \times \boldsymbol{x}$

Now if $\boldsymbol{x}$ ***is continuous of jth term then***

$= g(z) \times \left(\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})\right) \times \boldsymbol{x_j}$

So,

$$\frac{\partial}{\partial(\boldsymbol{\beta})}\left(\log\left(1-\boldsymbol{g}(\boldsymbol{\beta}^{T}\boldsymbol{x})\right)\right)$$

$$= \frac{1}{\left(\mathbf{1}-\boldsymbol{g}\left(\boldsymbol{\beta}^{T}\boldsymbol{x}\right)\right)} \times \left(-\left(g(z) \times \left(\mathbf{1}-\boldsymbol{g}(\boldsymbol{z})\right) \times \boldsymbol{x}_{\boldsymbol{j}}\right)\right)$$

and ,

$$\frac{\partial}{\partial(\boldsymbol{\beta})}(\boldsymbol{B})$$

$$= \left(1-y^{(i)} \times \frac{1}{\left(\mathbf{1}-\boldsymbol{g}\left(\boldsymbol{\beta}^{T}\boldsymbol{x}\right)\right)} \times \left(-\left(g(z) \times \left(\mathbf{1}-\boldsymbol{g}(\boldsymbol{z})\right) \times \boldsymbol{x}_{\boldsymbol{j}}\right)\right)\right)$$

$$= -\mathbf{1}\left(\frac{\mathbf{1} - y^{(i)} \times (g(z) \times \left(\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})\right) \times \boldsymbol{x}_j}{\left(\mathbf{1} - \boldsymbol{g}\left(\boldsymbol{\beta}^T \boldsymbol{x}\right)\right)} \right)$$

$$\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}(\boldsymbol{\beta})}(\boldsymbol{B}) = -\frac{(\mathbf{1} - y^{(i)} \times (g(z) \times \left(\mathbf{1} - \boldsymbol{g}(\boldsymbol{z})\right) \times \boldsymbol{x}_j\,)}{\left(\mathbf{1} - \boldsymbol{g}\left(\boldsymbol{\beta}^T \boldsymbol{x}\right)\right)}$$

Now we can put either $\beta^T x = z\ or\ z = \ \beta^T x$, we proceeding with $z = \ \beta^T x\ we\ get$:

$$= -\frac{(\mathbf{1} - y^{(i)} \times (g\left(\boldsymbol{\beta}^T \boldsymbol{x}\right) \times \left(\mathbf{1} - \boldsymbol{g}\left(\boldsymbol{\beta}^T \boldsymbol{x}\right)\right) \times \boldsymbol{x}_j\,)}{\left(\mathbf{1} - \boldsymbol{g}\left(\boldsymbol{\beta}^T \boldsymbol{x}\right)\right)}$$

$$= -\left(\mathbf{1} - y^{(i)} \times \left(g\left(\boldsymbol{\beta}^T \boldsymbol{x}\right)\right)\right) \times \boldsymbol{x}_j$$

$$\frac{\partial}{\partial(\beta)}(B)$$
$$= -\frac{\left(1 - y^{(i)} \times \left(g(z) \times \left(1 - g(z)\right) \times x_j\right)\right.}{\left(1 - g\left(\beta^T x\right)\right)}$$
$$= -\left(1 - y^{(i)} \times \left(g\left(\beta^T x\right)\right)\right) \times x_j$$

Now,

$$\sum_{i=1}^{n} \frac{\partial}{\partial(\beta)}(A) + \frac{\partial}{\partial(\beta)}(B)$$
$$= \left(\left(\left(y^{(i)}\right) \times \left(1 - g\left(\beta^T x\right)\right)\right) \times x_j\right)$$
$$+ \left(-\left(1 - y^{(i)} \times \left(g\left(\beta^T x\right)\right)\right) \times x_j\right)$$

$$= \left(\left(y^{(i)}\right) \times \left(1 - g\left(\beta^T x\right)\right)\right) - \left(1 - y^{(i)} \times \left(g\left(\beta^T x\right)\right)\right)$$
$$\times x_j$$

$$= \left(y^{(i)} - g\left(\beta^T x\right) y^{(i)} - g\left(\beta^T x\right) + g\left(\beta^T x\right) y^{(i)}\right) \times x_j$$

$$= \left(\boldsymbol{y}^{(i)} - \boldsymbol{g}\left(\boldsymbol{\beta}^{T}\boldsymbol{x}\right)\right) \times \boldsymbol{x}_j$$

Therefore,

$$\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}(\boldsymbol{\beta})}\left(\boldsymbol{l}(\boldsymbol{\beta})\right) = \left(\boldsymbol{y}^{(i)} - \boldsymbol{g}\left(\boldsymbol{\beta}^{T}\boldsymbol{x}\right)\right) \times \boldsymbol{x}_j$$

We know,

$$p\hat{} = h_\beta(x) or\ h(x) = g(\beta^T x)$$

Therefore ,

$$\frac{\partial}{\partial(\beta)}\left(l(\beta)\right) = \left(y^{(i)} - h_\beta(x)\right) \times x_j$$

Therefore,

$$\beta_j = \beta_j + \alpha \nabla_{\beta_j} l(\beta_j)$$

or,

$$\beta_j = \beta_j + \alpha\left(\left(y^{(i)} - h_\beta(x)\right) \times x_j\right)$$

Derivation of cost function of logistic regression:

Cost Function:

$$\boldsymbol{J(\beta) = -\frac{1}{m}\prod_{i=1}^{m} p^{\wedge(i)^{y^{(i)}}} \times \left(1 - p^{\wedge(i)}\right)^{1-y^{(i)}}}$$

Remember **Bernoulli's exponential family distribution**:

Putting log we get:

$$\boldsymbol{= -\frac{1}{m}\prod_{i=1}^{m} log\left(p^{\wedge(i)^{y^{(i)}}} \times \left(1 - p^{\wedge(i)}\right)^{1-y^{(i)}}\right)}$$

$$\boldsymbol{= -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} log\left(p^{\wedge(i)}\right) + 1 - y^{(i)} log\left(1 - p^{\wedge(i)}\right)\right)}$$

We know $p^{\wedge} = h_{\beta}(x) = \sigma(\beta^{T}x)$

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

$$\frac{\partial}{\partial \beta_j} J(\beta) =$$

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

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

We already came to know from Maximum Likelihood calculation:

$$\frac{\partial}{\partial(\beta)}\left(\left(y^{(i)} log\left(h_\beta(x)\right) + 1 - y^{(i)} log\left(1 - h_\beta(x)\right)\right)\right) = \left(y^{(i)} - g\left(\beta^T x\right)\right) \times x_j$$

Putting the value in above we get:

$$= \left(-\frac{1}{m} \sum_{i=1}^{n} \left(y^{(i)} - g\left(\beta^T x\right)\right) \times x_j \right)$$

Or,

$$= \left(-\frac{1}{m} \sum_{i=1}^{n} \left(-\left(g\left(\beta^T x\right) - y^{(i)}\right)\right) \times x_j \right)$$

Or,

$$= \left(\frac{1}{m} \sum_{i=1}^{n} \left(g\left(\beta^T x\right) - y^{(i)}\right) \times x_j \right)$$

Now $g = \sigma\ function\ and\ \beta = \theta\ then,$

$$=> \frac{\partial}{\partial \theta_j} J(\theta) = \left(\frac{1}{m} \sum_{i=1}^{n} \left(\sigma\left(\theta^T x\right) - y^{(i)}\right) \times x_j \right)$$

$$\frac{\partial}{\partial \beta_j} J(\beta) = \left(\frac{1}{m} \sum_{i=1}^{n} \left(g\left(\beta^T x\right) - y^{(i)}\right) \times x_j \right)$$

Or,

$$\frac{\partial}{\partial \beta_j} J(\beta) = \left(\frac{1}{m} \sum_{i=1}^{n} \left(h_\beta x - y^{(i)} \right) \times x_j \right)$$

Alternatively,

$$\frac{\partial}{\partial \theta_j} J(\theta) = \left(\frac{1}{m} \sum_{i=1}^{n} \left(\sigma\left(\theta^T x\right) - y^{(i)} \right) \times x_j \right)$$

Or,

$$\frac{\partial}{\partial \theta_j} J(\theta) = \left(\frac{1}{m} \sum_{i=1}^{n} \left(h_\beta x - y^{(i)} \right) \times x_j \right)$$