title: "Titre"
date: Fecha
output:
pdf_document:
includes:
in_header: m yst yles.st y

Logistic regression is given by :

$p(y \mid x) = \frac{1}{1+e^{(y<\theta x>)}}$

$p(y=+1 \mid x) = \frac{1}{1+e^{<\theta x>}}$

$Large(y=-1 \mid \ x) = 1 - p(y=+1 \mid x) = \frac{e^{<\theta x>}}{1+e^{<\theta x>}}$

where $<\theta, x>$ is the dot product of the transpose of the vector of weight parameter and the feature vectore

$\theta_{MLE} = \prod_{i=1}^{N} p(y=y_i \mid x=x_i, \theta) = \sum_{i=1}^{N} \ln p(y=y_i \mid p=x_i, \theta)$

Now I want to introduce an indicator variable $\mathbb{I}(y=+1)$, this indicated that output label(y) is +1 and $\mathbb{I}(y=-1)$ indicate that the output label(y) is -1, using this knowledge

$\mathcal{LL}(\theta) = \sum_{i=1}^{N} \mathbb{I}(y=+1) \ln p(y=+1 \mid x=x_i, \theta) + \mathbb{I}(y=-1) \ln p(y=-1 \mid x=x_i, \theta)$

we can also write $\mathbb{I}(y=-1) = 1 - \mathbb{I}(y=+1)$

$\mathcal{LL}(\theta) = \sum_{i=1}^{N} \mathbb{I}(y=+1) \ln p(y=+1 \mid x_i, \theta) + (1-\mathbb{I}(y=+1)) \ln p(y=-1 \mid x_i, \theta)$

substituting the value for logistic regression and simpl ying for a single datapoint we have,

$\mathcal{LL}(\theta) = \mathbb{I}(y_i=+1) \ln \frac{1}{1+e^{<\theta x_i>}} + (1-\mathbb{I}(y_i=+1)) \ln \frac{e^{<\theta x_i>}}{1+e^{<\theta x_i>}}$

$= -\mathbb{I}(y_i=+1) \ln(1+e^{<\theta x_i>}) + (1-\mathbb{I}(y_i=+1))(<\theta x_i> - \ln(1+e^{<\theta x_i>}))$

$\mathcal{LL}(\theta) = -\mathbb{I}(y_i=+1) \ln(1+e^{<\theta x_i>}) + <\theta x_i> - \ln(1+e^{<\theta x_i>}) - \mathbb{I}(y_i=+1) <\theta x_i> + \mathbb{I}(y_i=+1) \ln(1+ e^{<\theta x_i>})$

$= <\theta x_i> - \mathbb{I}(y_i=+1) <\theta x_i> - \ln(1+e^{<\theta x_i>})$

$= (1-\mathbb{I}(y_i=+1)) <\theta x_i> - \ln(1+e^{<\theta x_i>})$

$\frac{\partial \mathcal{LL}}{\partial \theta_j} = (1-\mathbb{I}(y_i=+1)) \frac{\partial(<\theta x_i>)}{\partial \theta_j} - \frac{\partial(\ln(1+e^{<\theta x_i>}))}{\partial \theta_j}$

$= (1-\mathbb{I}(y_i=+1))_j <x_i> - \frac{{}_j<x_i>e^{<\theta x_i>}}{(1+e^{<\theta x_i>})}$

$= (1-\mathbb{I}(y_i=+1))_j <x_i> -_j <x_i> p(y=-1 \mid x_i, \theta)$

$=_j <x_i> (p(y=+1 \mid x_i, \theta) - \mathbb{I}(y_i=+1))$

The derivative/gradient for one data point is given as:

${}_j <x_i> (p(y=+1 \mid x_i, \theta) - \mathbb{I}(y_i=+1))$

adding all the data point we have :

$\frac{\partial \mathcal{LL}}{\partial \theta_j} = \sum_{i=1}^{N} {}_j <x_i> (p(y=+1 \mid x_i, \theta) - \mathbb{I}(y_i=+1))$

where ${}_j <x_i>$ is the feature vector associated with j^{th} parameter .