Logistic Regression

Partial Derivative of the Cost Function

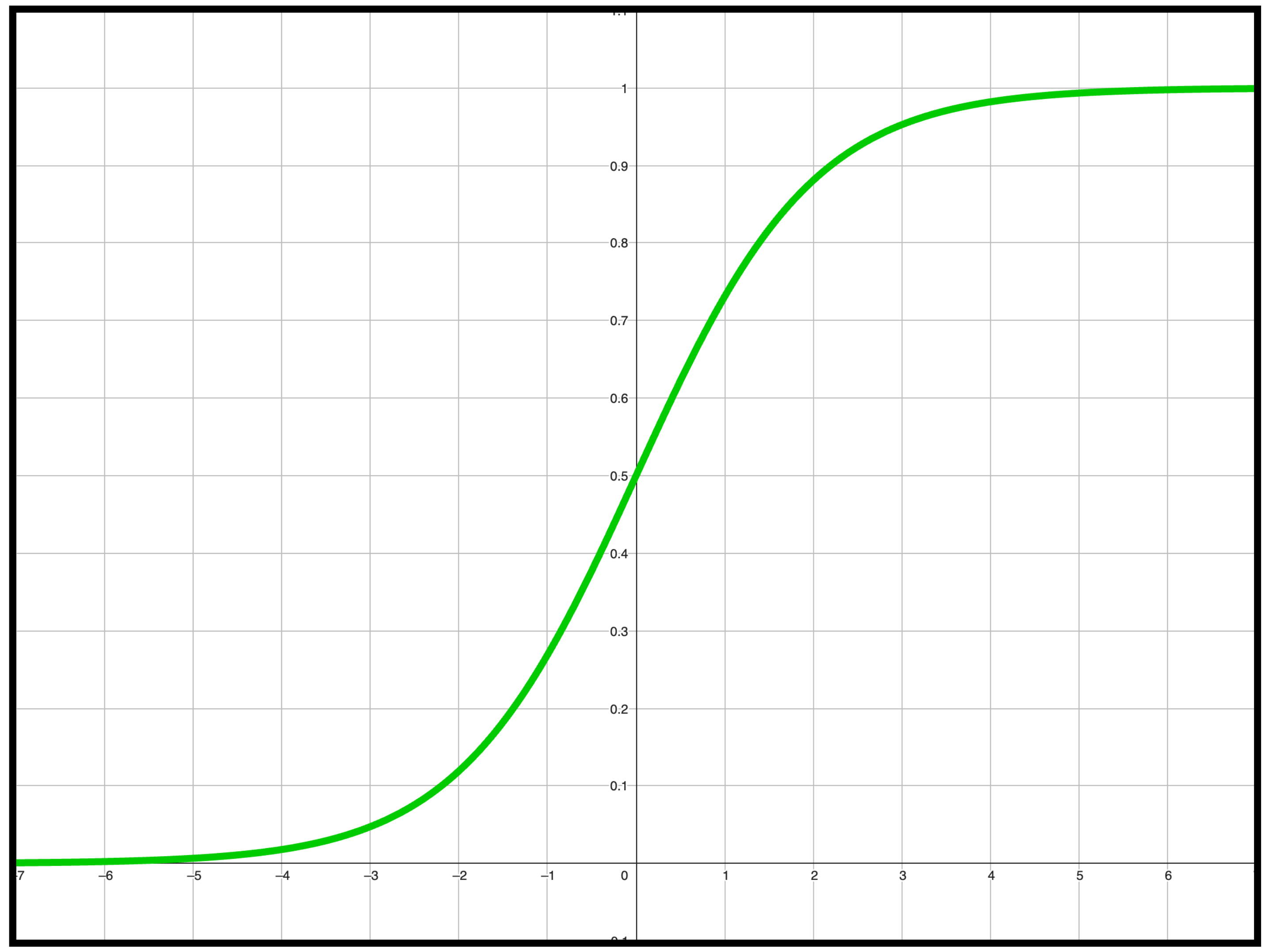
Rahul Singh
rsingh@arrsingh.com

Logistic Regression

$$\hat{Y} = \sigma(W^T X + \beta)$$

W is a $k \times 1$ vector of weights $\omega_1, \omega_2, \omega_3 \ldots \omega_k$
X is a $k \times n$ matrix of observations $x_1, x_2, x_3 \ldots x_k$
β is a scalar

$\hat{Y}$ is the predicted values for the model with parameters W and β



Logistic Regression uses Maximum Likelihood Estimation to find the parameters W and β

Logistic Regression Cost Function

$$\hat{Y} = \sigma(W^T X + \beta)$$

$$L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} y\, log_e \hat{y} + (1 - y)\, log_e(1 - \hat{y})$$

Partial Derivatives of the Cost Function w.r.t W and β

$$\frac{\partial}{\partial W} L(W, \beta) = (\hat{Y} - Y)\, X$$

$$\frac{\partial}{\partial \beta} L(W, \beta) = (\hat{Y} - Y)$$

Lets walk through the steps to compute the partial derivatives of the cost function w.r.t W and β

Logistic Regression Cost Function

$$\hat{Y} = \sigma(W^T X + \beta)$$

$$Z = W^T X + \beta$$

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

$$L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} y\, log_e \hat{y} + (1 - y)\, log_e(1 - \hat{y})$$

Partial Derivatives of the Cost Function w.r.t W and β

$$\frac{\partial}{\partial W}L(W, \beta) = \frac{\partial}{\partial \hat{Y}}L(W, \beta)\frac{\partial}{\partial Z}\hat{Y}\,\frac{\partial}{\partial W}Z$$

$$\frac{\partial}{\partial \beta}L(W, \beta) = \frac{\partial}{\partial \hat{Y}}L(W, \beta)\frac{\partial}{\partial Z}\hat{Y}\,\frac{\partial}{\partial \beta}Z$$

Chain Rule

See Tutorial on Derivatives

Lets calculate these four partial derivatives

$$\frac{\partial}{\partial \hat{Y}}L(W, \beta)$$

$$\frac{\partial}{\partial Z}\hat{Y}$$

$$\frac{\partial}{\partial W}Z$$

$$\frac{\partial}{\partial \beta}Z$$

Logistic Regression Model

$$Z = W^T X + \beta$$

Partial Derivative of Z w.r.t β

$$\frac{\partial}{\partial \beta} Z = \frac{\partial}{\partial \beta} W^T X + \beta = 1$$

$W^T X$ is a constant for the partial derivative w.r.t β

$$\frac{\partial}{\partial \beta} \beta = 1$$

$$\frac{\partial}{\partial \hat{Y}} L(W, \beta)$$

$$\frac{\partial}{\partial Z} \hat{Y}$$

$$\frac{\partial}{\partial W} Z$$

$$\frac{\partial}{\partial \beta} Z$$

Logistic Regression Model

$$Z = W^T X + \beta$$

Partial Derivative of Z w.r.t W

$$\frac{\partial}{\partial W} Z = \frac{\partial}{\partial W} W^T X + \beta = X$$

β is a constant for the partial derivative w.r.t W

$$\frac{\partial}{\partial W} W^T X = X$$

$$\frac{\partial}{\partial \hat{Y}} L(W, \beta)$$

$$\frac{\partial}{\partial Z} \hat{Y}$$

$$\frac{\partial}{\partial W} Z$$

✓ $$\frac{\partial}{\partial \beta} Z$$

Partial Derivative of $\hat{Y}$ w.r.t Z

$$\frac{\partial}{\partial Z}\hat{Y} = \frac{\partial}{\partial Z}(1+e^{-z})^{-1}$$

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = \frac{\partial}{\partial Z}(1+e^{-z})^{-1}\frac{\partial}{\partial Z}(1+e^{-Z})$$

Chain Rule

See Tutorial on Derivatives

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = -1(1+e^{-z})^{-2}\frac{\partial}{\partial Z}(1+e^{-Z})$$

Power Rule

See Tutorial on Derivatives

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = -1(1+e^{-z})^{-2}(-1)(e^{-Z})$$

$$\frac{\partial}{\partial x}e^{-x} = -e^{-x}$$

See Tutorial on Derivatives

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = (1+e^{-z})^{-2}e^{-Z}$$

Logistic Regression Model

$$Z = W^TX + \beta$$

$$\hat{Y} = \sigma(Z) = \frac{1}{1+e^{-Z}} = (1+e^{-z})^{-1}$$

$$\frac{\partial}{\partial \hat{Y}}L(W,\beta)$$

$$\frac{\partial}{\partial Z}\hat{Y}$$

✓ $$\frac{\partial}{\partial W}Z$$

✓ $$\frac{\partial}{\partial \beta}Z$$

Partial Derivative of $\hat{Y}$ w.r.t Z

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = (1 + e^{-z})^{-2}\, e^{-Z}$$

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = \frac{e^{-Z}}{(1 + e^{-Z})^2}$$

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = (\hat{Y})^2 e^{-Z}$$

$$\hat{Y} = \frac{1}{1 + e^{-Z}}$$

$$\Rightarrow (\hat{Y})^2 = \frac{1}{(1 + e^{-Z})^2}$$

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = (\hat{Y})^2 \frac{(1 - \hat{Y})}{\hat{Y}}$$

$$e^{-Z} = \frac{(1 - \hat{Y})}{\hat{Y}}$$

$$\Rightarrow \frac{\partial}{\partial Z}\hat{Y} = \hat{Y}(1 - \hat{Y})$$

Logistic Regression Model

$$Z = W^T X + \beta$$

$$\hat{Y} = \sigma(Z) = \frac{1}{1 + e^{-Z}} = (1 + e^{-z})^{-1}$$

$$\frac{\partial}{\partial \hat{Y}} L(W, \beta)$$

$$\frac{\partial}{\partial Z}\hat{Y}$$

✓ $$\frac{\partial}{\partial W} Z$$

✓ $$\frac{\partial}{\partial \beta} Z$$

Logistic Regression Model

$$Z = W^T X + \beta$$

$$L(W,\beta) = -\frac{1}{n}\sum_{i=0}^{n} y\, log_e \hat{y} + (1-y)\, log_e(1-\hat{y})$$

Partial Derivative of $L(W,\beta)$ w.r.t $\hat{Y}$

$$\frac{\partial}{\partial \hat{Y}} L(W,\beta) = \frac{\partial}{\partial \hat{Y}}\left(-\frac{1}{n}\sum_{i=0}^{n} y\, log_e \hat{y} + (1-y)\, log_e(1-\hat{y})\right)$$

$$\Rightarrow \frac{\partial}{\partial \hat{Y}} L(W,\beta) = -\frac{1}{n}\sum_{i=0}^{n} \frac{\partial}{\partial \hat{Y}}\left(y\, log_e \hat{y} + (1-y)\, log_e(1-\hat{y})\right)$$

$$\Rightarrow \frac{\partial}{\partial \hat{Y}} L(W,\beta) = -\frac{1}{n}\sum_{i=0}^{n} \frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\frac{\partial}{\partial \hat{Y}}(1-\hat{y})$$

$\frac{\partial}{\partial \hat{Y}} L(W,\beta)$

✓ $\frac{\partial}{\partial Z}\hat{Y}$

✓ $\frac{\partial}{\partial W} Z$

✓ $\frac{\partial}{\partial \beta} Z$

Logistic Regression Model

$$Z = W^T X + \beta$$

$$L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} y\, log_e \hat{y} + (1-y)\, log_e (1-\hat{y})$$

Partial Derivative of $L(W, \beta)$ w.r.t $\hat{Y}$

$$\Rightarrow \frac{\partial}{\partial \hat{Y}} L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} \frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}} \frac{\partial}{\partial \hat{Y}}(1-\hat{y})$$

$$\Rightarrow \frac{\partial}{\partial \hat{Y}} L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} \frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}(-1)$$

$$\Rightarrow \frac{\partial}{\partial \hat{Y}} L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} \frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}} = \frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

- ✓ $\frac{\partial}{\partial \hat{Y}} L(W, \beta)$
- ✓ $\frac{\partial}{\partial Z} \hat{Y}$
- ✓ $\frac{\partial}{\partial W} Z$
- ✓ $\frac{\partial}{\partial \beta} Z$

Logistic Regression Cost Function

$$\hat{Y} = \sigma(W^T X + \beta)$$

$$Z = W^T X + \beta$$

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

$$L(W, \beta) = -\frac{1}{n}\sum_{i=0}^{n} y\, log_e \hat{y} + (1 - y)\, log_e (1 - \hat{y})$$

Partial Derivatives of the Cost Function w.r.t W and β

$$\frac{\partial}{\partial W} L(W, \beta) = \frac{\partial}{\partial \hat{Y}} L(W, \beta) \frac{\partial}{\partial Z} \hat{Y} \frac{\partial}{\partial W} Z$$

$$\frac{\partial}{\partial \beta} L(W, \beta) = \frac{\partial}{\partial \hat{Y}} L(W, \beta) \frac{\partial}{\partial Z} \hat{Y} \frac{\partial}{\partial \beta} Z$$

We've computed the 4 partial derivatives

$$\frac{\partial}{\partial \hat{Y}} L(W, \beta) = \frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

$$\frac{\partial}{\partial Z} \hat{Y} = \hat{Y}(1 - \hat{Y})$$

$$\frac{\partial}{\partial W} Z = X$$

$$\frac{\partial}{\partial \beta} Z = 1$$

Partial Derivatives of the Cost Function w.r.t W

We've computed the 4 partial derivatives

$$\frac{\partial}{\partial W}L(W,\beta) = \frac{\partial}{\partial \hat{Y}}L(W,\beta)\frac{\partial}{\partial Z}\hat{Y}\,\frac{\partial}{\partial W}Z$$

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\hat{Y}(1-\hat{Y})\,X$$

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\left[\frac{-y+\hat{y}y+\hat{y}-\hat{y}y}{\hat{y}(1-\hat{y})}\right]\right)\hat{Y}(1-\hat{Y})\,X$$

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\left[\frac{\hat{y}-y}{\hat{y}(1-\hat{y})}\right]\right)\hat{Y}(1-\hat{Y})\,X$$

$$\frac{\partial}{\partial \hat{Y}}L(W,\beta) = \frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

$$\frac{\partial}{\partial Z}\hat{Y} = \hat{Y}(1-\hat{Y})$$

$$\frac{\partial}{\partial W}Z = X$$

$$\frac{\partial}{\partial \beta}Z = 1$$

Logistic Regression

Partial Derivatives of the Cost Function w.r.t W

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\left[\frac{\hat{y}-y}{\hat{y}(1-\hat{y})}\right]\right)\hat{Y}(1-\hat{Y})\,X$$

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = \left(\frac{\frac{1}{n}\sum_{i=0}^{n}(\hat{y}-y)}{\hat{Y}(1-\hat{Y})}\right)\hat{Y}(1-\hat{Y})\,X$$

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\hat{y}(1-\hat{y})\right)X$$

$$\Rightarrow \frac{\partial}{\partial W}L(W,\beta) = (\hat{Y}-Y)X$$

Q.E.D

We've computed the 4 partial derivatives

$$\frac{\partial}{\partial \hat{Y}}L(W,\beta) = \frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

$$\frac{\partial}{\partial Z}\hat{Y} = \hat{Y}(1-\hat{Y})$$

$$\frac{\partial}{\partial W}Z = X$$

$$\frac{\partial}{\partial \beta}Z = 1$$

Partial Derivatives of the Cost Function w.r.t β

We've computed the 4 partial derivatives

$$\frac{\partial}{\partial \beta}L(W,\beta) = \frac{\partial}{\partial \hat{Y}}L(W,\beta)\frac{\partial}{\partial Z}\hat{Y}\frac{\partial}{\partial \beta}Z$$

$$\Rightarrow \frac{\partial}{\partial \beta}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\frac{\partial}{\partial Z}\hat{Y}\frac{\partial}{\partial \beta}Z$$

$$\Rightarrow \frac{\partial}{\partial \beta}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\hat{Y}(1-\hat{Y})\,1$$

$$\Rightarrow \frac{\partial}{\partial \beta}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\left[\frac{-y+\hat{y}y+\hat{y}-\hat{y}y}{\hat{y}(1-\hat{y})}\right]\right)\hat{Y}(1-\hat{Y})\,1$$

$$\frac{\partial}{\partial \hat{Y}}L(W,\beta) = \frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

$$\frac{\partial}{\partial Z}\hat{Y} = \hat{Y}(1-\hat{Y})$$

$$\frac{\partial}{\partial W}Z = X$$

$$\frac{\partial}{\partial \beta}Z = 1$$

Partial Derivatives of the Cost Function w.r.t β

We've computed the 4 partial derivatives

$$\frac{\partial}{\partial \beta}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\left[\frac{-y+\hat{y}y+\hat{y}-\hat{y}y}{\hat{y}(1-\hat{y})}\right]\right)\hat{Y}(1-\hat{Y})\,1$$

$$\frac{\partial}{\partial \beta}L(W,\beta) = \left(\frac{1}{n}\sum_{i=0}^{n}\left[\frac{\hat{y}-y}{\hat{y}(1-\hat{y})}\right]\right)\hat{Y}(1-\hat{Y})$$

$$\frac{\partial}{\partial \beta}L(W,\beta) = \left(\frac{\frac{1}{n}\sum_{i=0}^{n}(\hat{y}-y)}{\hat{Y}(1-\hat{Y})}\right)\hat{Y}(1-\hat{Y})$$

$$\frac{\partial}{\partial \beta}L(W,\beta) = \frac{1}{n}\sum_{i=0}^{n}(\hat{y}-y) = \hat{Y}-Y$$

Q.E.D

$$\frac{\partial}{\partial \hat{Y}}L(W,\beta) = \frac{1}{n}\sum_{i=0}^{n} -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

$$\frac{\partial}{\partial Z}\hat{Y} = \hat{Y}(1-\hat{Y})$$

$$\frac{\partial}{\partial W}Z = X$$

$$\frac{\partial}{\partial \beta}Z = 1$$

Related Tutorials & Textbooks

[Logistic Regression](#)

An introduction to Logistic Regression. A Logistic Regression model use used to predict a binary value (the dependent variable) for one or more independent variables using a threshold to classify a probability.

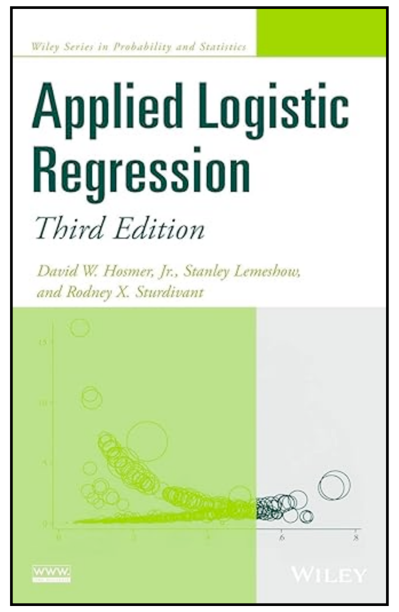
[Multiple Regression](#)

Multiple regression extends the two dimensional linear model introduced in Simple Linear Regression to k + 1 dimensions with one dependent variable, k independent variables and k+1 parameters.

[Cost Function & Gradient Descent for Logistic Regression](#)

An introduction to the Cost function for Logistic Regression long with its partial derivative (the gradient vector). The model parameters (B & W) are then optimized using Maximum Likelihood Estimation and Gradient Descent.

Recommended Textbooks



Applied Logistic Regression

by David W. Hosmer Jr., Stanley Lemeshow, Rodney X. Sturdivant

For a complete list of tutorials see:
https://arrsingh.com/ai-tutorials