# *Getting Started with Machine Learning*

*Raunak J*

In the last video we learned Log Loss.

In this video we'll try to optimize that loss function.

Using Gradient Descent.

So Gradient Descent is basically a loss optimization algorithm, that derivates the loss function and reaches a global minima to shift the parameters.

So in our example of Logistic Regression Loss Optimization, parameters are constant and regression co-efficient.

So our loss function for Logistic Regression is Log Loss :

$$LL_{(y,\hat{y})} = \frac{1}{N}\sum_{i=1}^{N} -y_i.\log\left(P(\widehat{y_i})\right) - (1 - y_i).\log(1 - P(\widehat{y_i}))$$

This can be elaborated as :

$$LL_{(y,\sigma(Z))} = \frac{1}{N}\sum_{i=1}^{N} -y_i.\log\left(\sigma(Z_i)\right) - (1 - y_i).\log(1 - \sigma(Z_i))$$

$$LL_{(y,\sigma(Z))} = \frac{1}{N}\sum_{i=1}^{N} -y_i.\log\big(\sigma(Z_i)\big) - (1-y_i).\log(1-\sigma(Z_i))$$

$$LL_{(y,\sigma(Z))} = \frac{1}{N}\sum_{i=1}^{N} -y_i.\log\left(\frac{1}{1+e^{-Z_i}}\right) - (1-y_i).\log(1-\frac{1}{1+e^{-Z_i}})$$

$$LL_{(y,\sigma(Z))} = \frac{1}{N}\sum_{i=1}^{N} -y_i.\log\left(\frac{1}{1+e^{-(\beta_0+\beta_1.x_i)}}\right) - (1-y_i).\log(1-\frac{1}{1+e^{-(\beta_0+\beta_1.x_i)}})$$

So for performing Loss Optimization of Logistic Regression :

Step 1 : Derivate the Loss Function.

Step 2 : Converge the Loss Function with parameters iteratively.

$$LL_{(y,\sigma(Z))} = \frac{1}{N}\sum_{i=1}^{N} -y_i.\log\big(\sigma(Z_i)\big) - (1-y_i).\log(1-\sigma(Z_i))$$

Following the step 1 :

$$\nabla_{\beta_i} = \sum_{i=1}^{N} -y_i.\log\left(\frac{1}{1+e^{-(\beta_0+\beta_1.x_i)}}\right) - (1-y_i).\log(1-\frac{1}{1+e^{-(\beta_0+\beta_1.x_i)}})$$

$$\nabla_{\beta_i} = \sum_{i=1}^{N} \left\{-y_i.\log\left(\frac{1}{1+e^{-(\beta_0+\beta_1.x_i)}}\right) - (1-y_i).\log\left(1-\frac{1}{1+e^{-(\beta_0+\beta_1.x_i)}}\right)\right\}.x_i$$

$$\nabla_{\beta_i} = \sum_{i=1}^{N} \left\{ -y_i . \log\left( \frac{1}{1 + e^{-(\beta_0 + \beta_1 . x_i)}} \right) - (1 - y_i) . \log\left( 1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 . x_i)}} \right) \right\} . x_i$$

$$\nabla_{\beta_i} = \sum_{i=1}^{N} LL_{\{y_i, \sigma(Z)\}} . x_i$$

$$\nabla_{\beta_i} = \frac{1}{N} \sum_{i=1}^{N} LL_{\{y_i, \sigma(Z)\}} . x_i$$

Step 2 :

$$\beta_i = \beta_i - \alpha . \nabla_{\beta_i}$$

$$\beta_i = \beta_i - \alpha . \frac{1}{N} \sum_{i=1}^{N} LL_{\{y_i, \sigma(Z)\}} . x_i$$

$$Where\ \beta_i\ are\ parameters\ and\ \alpha\ is\ learning\ rate$$

Iterating {

$$\beta_i = \beta_i - \alpha.\frac{1}{N}\sum_{i=1}^{N} LL_{\{y_i,\sigma(Z)\}}.x_i$$

}