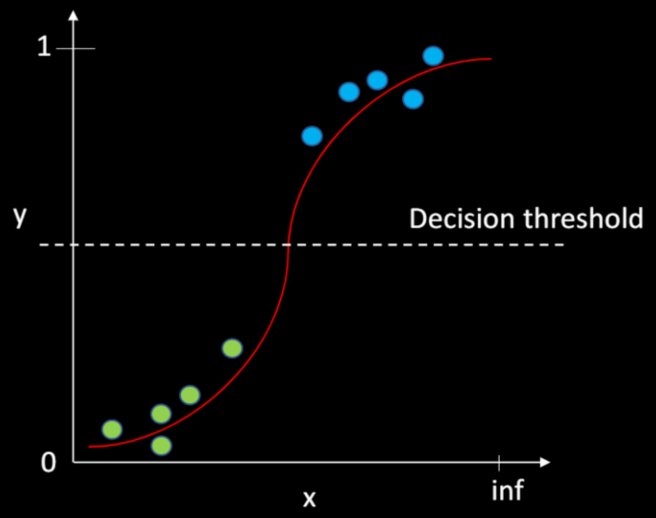
Training a Logistic Regression model



Logistic Regression

(used to solve Classification problems)

Prabhjot Kaur

Background

- Supervised Machine Learning
- Classification
- Sigmoid function
- Cost function
- Gradient descent

What is logistic regression?

$$y = w_0 + w_1 x_1 + w_2 x_2$$

Input samples	Feature 1 (x1)	Feature 2 (x2)	Labels
1	50	32	1
2	10	22	1
3	55	39	1
4	34	67	0

$$y = w_0 + [w_1 \quad w_2][x_1 \quad x_2]^T$$

$$y = w_0 + wx^T$$

(scalar) (scalar) (1x2)*(2x1) (scalar)

$$Y = W_0 + WX^T$$

(1x4) (1x4) (1x2)*(2x4)

Process entire dataset at once

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

$$y\_predicted = \begin{cases} 0 \text{ if } \bar{y} < 0.5 \\ 1 \text{ if } \bar{y} >= 0.5 \end{cases}$$

Training the model

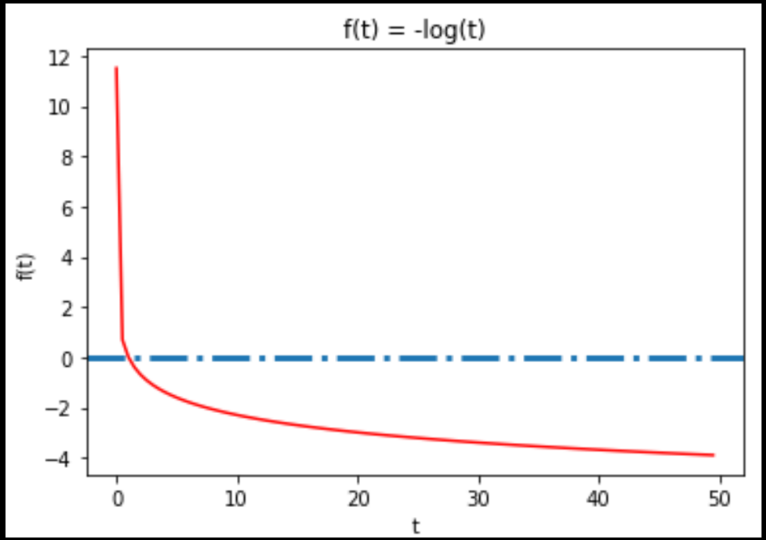
1. Start with a random set of parameters (W)
2. For epoch 1: N
   i. For all samples in the training set (say, m number of samples)
      i. Predict the output using a LR model based on the random set of weights
      ii. Find the cost function (error between true and predicted output)
   ii. Find average cost for all training samples
   iii. Minimize the cost function (using gradient descent)
   iv. Update model parameters
   v. Repeat step 2 until convergence or for N iterations

$$C(\theta) = \begin{cases} -\log(\bar{y}), if\ ytrue = 1 \\ -\log(1-\bar{y}), if\ ytrue = 0 \end{cases}$$

Cost function for the LR

Cost function for LR

$$C(w) = \begin{cases} -\log(\bar{y}), if\ ytrue = 1 \\ -\log(1-\bar{y}), if\ ytrue = 0 \end{cases}$$



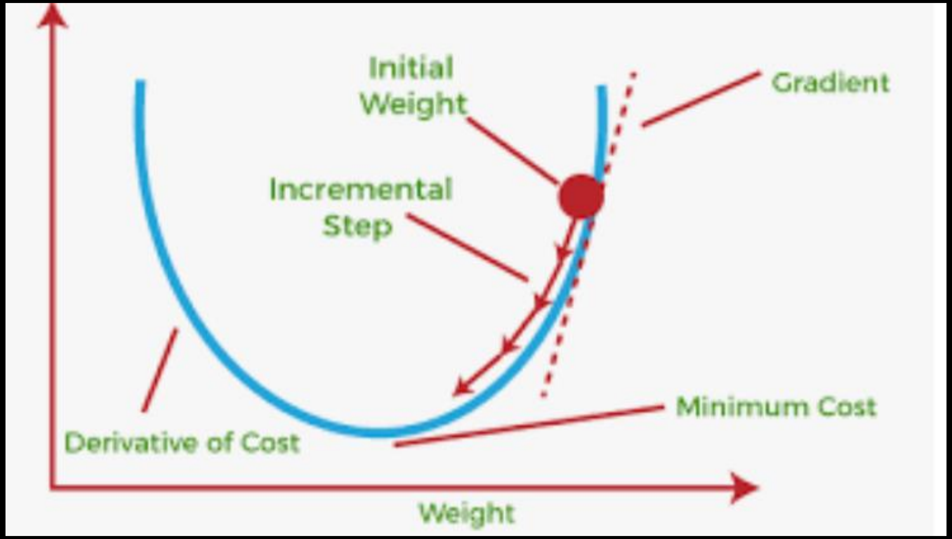
$$J(w) = \frac{-1}{m}\sum_{i=0}^{m}\left[ytrue^{i}\log(\bar{y}i) + \left(1 - ytrue^{i}\right)\log(1-\bar{y}i)\right]$$

$$\frac{\partial}{\partial w_j}J(w) = \frac{1}{m}\sum_{i=1}^{m}\left(\bar{y}^{i} - y^{i}\right)x^{i}$$

Derivative of the cost function – used to update the parameters using gradient descent

Update parameters using gradient descent

$$\frac{\partial}{\partial w_j} J(w) = \frac{1}{m} \sum_{i=1}^{m} \left(\bar{y}^i - y^i \right) x^i$$



$$w_{new} = w - \alpha \frac{\partial}{\partial w} J(w)$$

After the model is trained, we will have the model parameters (W) and bias(Wo).

Let’s code!