# Levenberg- Marquardt Descent Technique -

## Gradient Descent :-

$$x_{n+1} = x_n - \lambda \cdot \nabla f$$

## Newton - Rhapson -

Solution for stationary point :-

Hessian, not Laplacian

$$\nabla f(x_{n+1}) = \nabla f(x_n) + (x_{n+1} - x_n)^T \nabla^2 f(x_n) + \text{Higher Order Terms} = 0$$

- Ignoring Higher Order Terms

$$\Rightarrow x_{n+1} = x_n - (\nabla^2 f(x_n))^{-1} \cdot \nabla f(x_n)$$

- $f(x)$ is a scalar function.

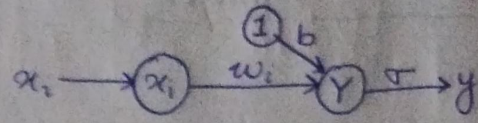
$$H = \nabla^2 f(x_n) = \begin{array}{cc} \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} & \begin{array}{l} \rightarrow \frac{\partial}{\partial x_1} \nabla f \\ \rightarrow \frac{\partial}{\partial x_2} \nabla f \\ \\ \rightarrow \frac{\partial}{\partial x_n} \nabla f \end{array} \end{array}$$

Problems :-
- Matrix may be singular
- Matrix must be positive definite, i.e, eigenvalues are all positive.

Example functions:- $f(x, y) = (a-x)^2 + b(y-x^2)^2$

$$f|_{1,1} = 0$$

~~Ex~~




$$Y = x \cdot w + b$$
$$\hat{y} = \sigma(x \cdot w + b)$$

- $\sigma'(x) = \sigma(x)\,(1 - \sigma(x))$

$$\cancel{\nabla y = \begin{bmatrix} \frac{\partial y}{\partial w} \\ \frac{\partial y}{} \end{bmatrix}}$$

$$f = L(w, b) = (\hat{y} - y)^2$$
$$= \cancel{(y - \sigma(xw + b))^2}$$
$$= (\sigma(xw + b) - y)^2$$

$$\nabla f = \begin{bmatrix} \frac{\partial L}{\partial w} \\ \frac{\partial L}{\partial b} \end{bmatrix} = \begin{bmatrix} 2(\hat{y} - y)\{(\hat{y} - y)\}' \\ 2(\hat{y} - y)(\hat{y} - y)' \end{bmatrix} = \begin{bmatrix} 2(\hat{y} - y)(\sigma(xw + b)) \cdot x \\ 2(\hat{y} - y) \cdot \hat{y} \end{bmatrix} = 2\begin{bmatrix} (\hat{y} - y) \cdot \hat{y} \cdot x \\ (\hat{y} - y) \cdot \hat{y} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial w^2} = x \frac{\partial}{\partial w}\left[(\hat{y} - y) \cdot \hat{y}\right]$$

$$= x \cdot \left\{\hat{y} \cdot \frac{\partial}{\partial w}(\hat{y} - y) + (\hat{y} - y) \cdot \frac{\partial}{\partial w}\hat{y}\right\}$$

$$= x \cdot \left\{\hat{y} \cdot \hat{y}(1 + \hat{y}) + (\hat{y} - y) \cdot \hat{y} \cdot (1 + \hat{y})\right\}$$

$$\frac{\partial^2 f}{\partial b^2} = \frac{\partial}{\partial b}\left[(\hat{y} - y) \cdot \hat{y}\right]$$

$$= (\hat{y} - y)\frac{\partial}{\partial b}(\hat{y}) + \hat{y} \cdot \frac{\partial}{\partial b}(\hat{y} - y)$$

= same as above

$$\Rightarrow \text{Let } c = \hat{y}^2(1-\hat{y}) + (\hat{y}-y)\cdot\hat{y}\cdot(1-\hat{y})$$
$$= (1+\hat{y})\left[\hat{y}^2 + (\hat{y}-y)\cdot\hat{y}\right]$$

Then –

$$\nabla^2 f = 2\cdot\begin{bmatrix} x\cdot c & x\cdot c \\ x\cdot c & c \end{bmatrix}$$

$$= 2c\begin{bmatrix} x & x \\ x & 1 \end{bmatrix}$$

## Levenberg - Marquardt

$$x_{n+1} = \cancel{\nabla} x_n - (\underbrace{\nabla^2 f(x_n) + \lambda I}_{M})^{-1}\nabla f$$

$$M = 2c\begin{bmatrix} x & x \\ x & 1 \end{bmatrix} + \lambda I$$

$$= 2c\underbrace{\begin{bmatrix} x+\lambda & x \\ x & 1+\lambda \end{bmatrix}}_{A}$$

$$M^{-1} = \frac{\text{adj } M}{|M|}$$

$$A^{-1} = \frac{1}{(x+\lambda)(1+\lambda) - x^2}\begin{bmatrix} 1+\lambda & -x \\ -x & x+\lambda \end{bmatrix}$$

$$\because x_{n+1} = x_n - M^{-1} \nabla f$$

$$\Rightarrow \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} w_{old} \\ b_{old} \end{bmatrix} - \frac{2c \cdot 2d}{(x+\lambda)(1+\lambda) - x^2} \begin{bmatrix} 1+\lambda & -x \\ -x & x+\lambda \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} w_{old} \\ b_{old} \end{bmatrix} - \frac{2c \cdot 2d}{(x+\lambda)(1+\lambda) - x^2} \begin{bmatrix} x(1+\lambda) - x \\ -x^2 + x + \lambda \end{bmatrix}$$

where

$c = (1+\hat{y})[\hat{y}^2 + (\hat{y}-y)\hat{y}]$

$d = (\hat{y}-y) \cdot \hat{y}$