2. The error function can be approximated around ~~the~~ $\hat{w}$ using Taylor's series as

$$E(w) \simeq E(\hat{w}) + b(w-\hat{w})^{\tau} + \frac{1}{2}(w-\hat{w})^{T} H (w-\hat{w}) \quad \text{—①}$$

Here b is the gradient of E at $\hat{w}$.

$$b = \nabla E \big|_{w=\hat{w}}$$

and H is the Hessian Matrix of $W \times W$ dimensions

$$H_{ij} = \frac{\partial E}{\partial w_i \partial w_j}\bigg|_{w=\hat{w}}$$

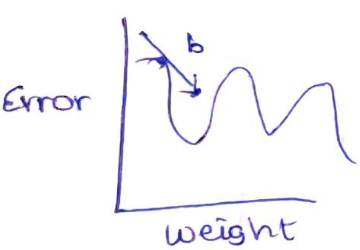
Here weights are from w_1 to w_n (n is the total weight)
In question this 'n' is represented as W

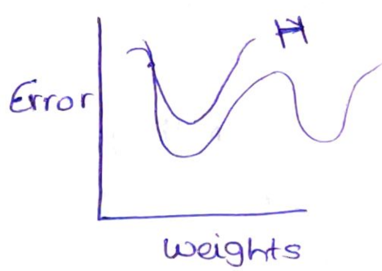
$$H = \begin{bmatrix} \frac{\partial E^n}{\partial w_1^n} & \frac{\partial E^n}{\partial w_1^{n-1} \cdot w_2} & \cdots \\ & \vdots & \\ & & \frac{\partial E^n}{\partial w_n^n} \end{bmatrix}$$

The corresponding local app gradient is (obtained from ①)

$$\nabla E = b + H(w-\hat{w}^{\tau})$$

We go for higher order of derivates using the taylor's series as it helps in better approximate the Error (lower)





∴ The total no. of independent terms in the quadratic error function $E(W)$ is $W(W+3)/2$

In the quadratic ~~approach~~ approximation of error ~~surface~~ function, error function is specified by the quantities 'b' & 'H'.

The total independent term in $b = W$ (W-dimensional vector)

$$H = W + \left(\frac{W^2 - W}{2}\right)$$

[as H is symmetric & we consider upper tria & diagonal elts)

$$\Rightarrow \text{Total} = W + W + \left(\frac{W^2 - W}{2}\right)$$

$$= \frac{4W + W^2 - W}{2}$$

$$= \frac{W(W+3)}{2}$$