Implementing Gradient Descent

Thursday, 27 May 2021 3:30 PM

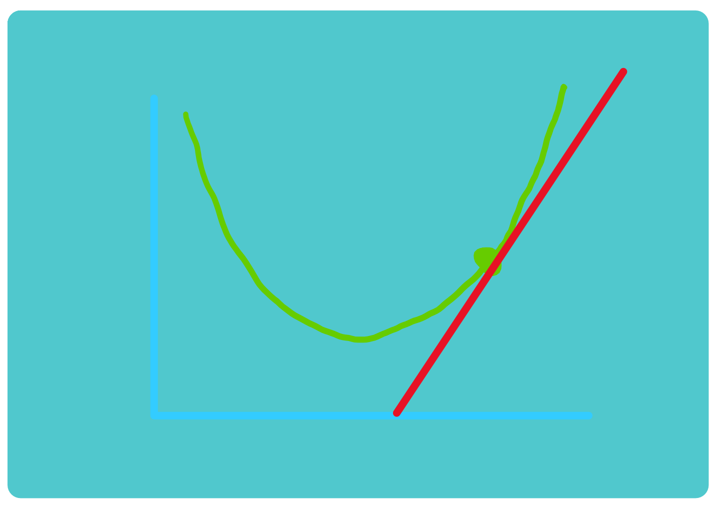
I. Gradient descent

Find set of weight that will minimize the error.

$$MSE = E(x) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2$$

m = # data
n = # classes

* Gradient :- Rate of change in slope at a given point



Gradient descent: Following the negative direction of the slope

Depending on the error function it could have many local minima, one method to avoid it is using momentum

II. Derivative of MSE

$$\nabla E = \left(\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \dots \frac{\partial E}{\partial w_n}, \frac{\partial E}{\partial b}\right)$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial w_i}$$

$$\frac{\partial E}{\partial \hat{y}_i} = \frac{\partial}{\partial \hat{y}_i}\left(\frac{1}{2}(y_i - \hat{y}_i)^2\right) = \frac{2}{2}(y_i - \hat{y}_i)\cdot -1 = -(y_i - \hat{y}_i)$$

$$\frac{\partial E}{\partial w_i} = -(y_i - \hat{y}_i)\cdot\frac{\partial \hat{y}}{\partial w_i} = -(y_i - \hat{y}_i)\cdot f'(h)\cdot x_i \quad , \quad \hat{y} = f(Wx+b) = f(h)$$

Update weights

$$w_i = w_i + \alpha \frac{\partial E}{\partial w_i}$$

$$\delta = -(y_i - \hat{y}_i)\cdot f'(h)$$

$$w_i = w_i + \alpha\,\delta\, x_i$$

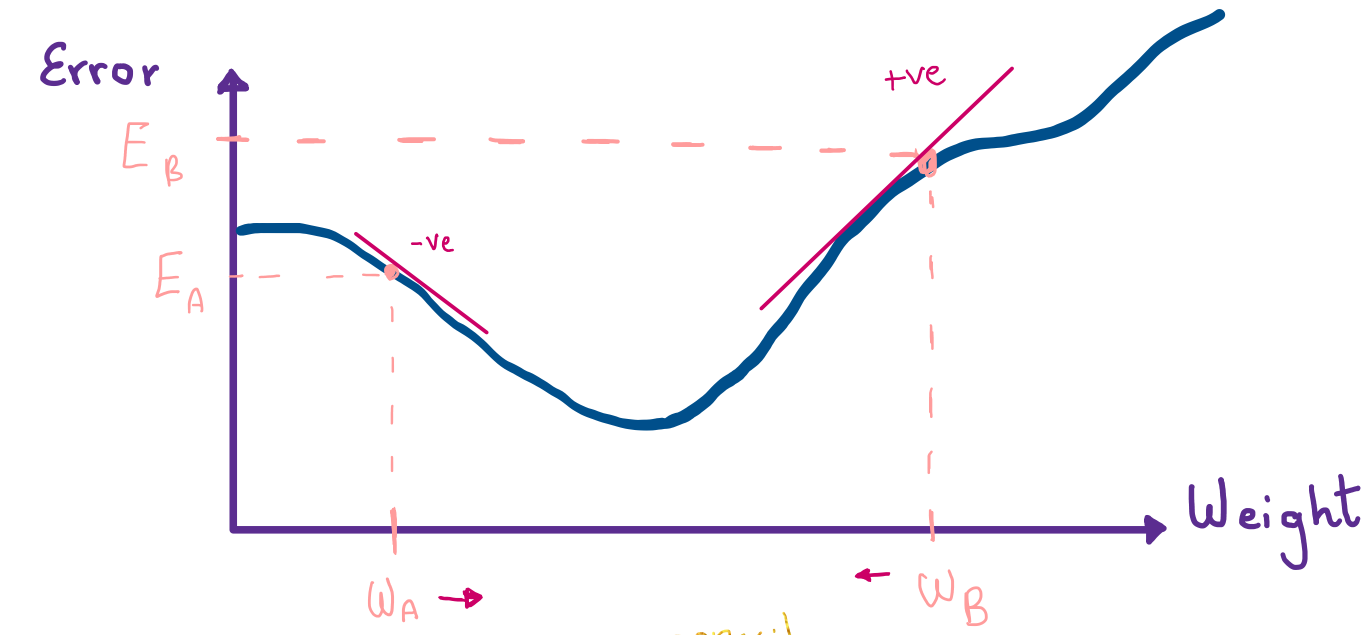
III. Code + Implementation.

in Notebook.

IV. Backpropagation

Propagating error backward to find the set ow weights that reduce the error

⚠ Note :-



opposits

- −ve gradient ▶ increase weight
- +ve gradient ▶ reduce weight

$$W_A' = W_A + \alpha\left(-\frac{\partial E}{\partial W}\right)$$