Since a new regularization term was introduced, the derivative has been changed.

$$\frac{\partial J(\vec{w},b)}{\partial w_j} \longrightarrow \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}w_j$$

$$\frac{\partial J(\vec{w},b)}{\partial b} \longrightarrow \frac{1}{m}\sum_{i=1}^{m} f_{\vec{w},b}(x^{(i)}) - y^{(i)}$$

it remains the same because we don't regularize b

Gradient Descent :-

repeat {

$$w_j = w_j - \alpha \frac{\partial J(\vec{w},b)}{\partial w_j} \rightarrow \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}w_j$$

$$b = b - \alpha \frac{\partial J(\vec{w},b)}{\partial b} \rightarrow \frac{1}{m}\sum_{i=1}^{m} f_{\vec{w},b}(x^{(i)}) - y^{(i)}$$

} simultaneous update

How is regularization shrinking all the parameters?

Gradient descent is as follows:-

$$w_j = w_j - \alpha \underbrace{\left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x_j^{(i)} + \frac{\lambda}{m} \sum_{j=1}^{n} w_j \right]}_{\frac{\partial J(\vec{w},b)}{\partial w_j}}$$

$$\Rightarrow w_j = w_j - \frac{\alpha}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x_j^{(i)} - \frac{\alpha\lambda}{m} w_j$$

On rearrangement

$$\Rightarrow w_j = w_j - \underline{\frac{\alpha\lambda}{m} w_j} - \underline{\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x_j^{(i)}}$$

$$\Rightarrow w_j = \underbrace{w_j \left[1 - \alpha \frac{\lambda}{m} \right]}_{\text{taking } w_j \text{ common}} - \underbrace{\frac{\alpha}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x_j^{(i)}}_{\text{usual term}}$$

Let's take $\alpha = 0.01$, $\lambda = 1$ and $m = 50$, for example.

$$w_j = w_j [1 - 0.0002] - \frac{\alpha}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x^{(i)}$$

$$w_j = w_j \underbrace{(0.9998)}_{\text{very small}} - \frac{\alpha}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x^{(i)}$$

Earlier we used to subtract just the usual term $\left(\frac{\alpha}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(x^{(i)}) - y^{(i)}) x^{(i)}\right)$ from w_j, but now we have to multiply a small number like 0.9998 (< 1) to w_j and then subtract the usual term.

This makes the w_j shrinking process more effective and faster by decreasing w_j in each iteration by a little bit.

How we get the derivative term (optional)

$$\frac{\partial}{\partial w_j} J(\vec{w}, b) = \frac{d}{d w_j}\left[\frac{1}{2m}\sum_{i=1}^{m}\Big(\underbrace{f(\vec{x}^{(i)})}_{\vec{w}\cdot\vec{x}^{(i)}+b} - y^{(i)}\Big)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2\right]$$

$$= \frac{1}{\cancel{2}m}\sum_{i=1}^{m}\left[\left(\vec{w}\cdot\vec{x}^{(i)} + b - y^{(i)}\right)\cancel{2}x_j^{(i)}\right] + \frac{\lambda}{\cancel{2}m}\cancel{2}w_j \quad \text{No } \sum_{j=1}^{n}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[\Big(\underbrace{\vec{w}\cdot\vec{x}^{(i)} + b}_{f(\vec{x})} - y^{(i)}\Big) x_j^{(i)}\right] + \frac{\lambda}{m} w_j$$

no need for summation because in gradient descent w will automatically be updated.