Minimizing Loss functions with Gradient descent (Adaline)

$$L(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \sigma(z^{(i)})\right)^2$$

Why? $\sigma(z)$ defines σ: classifier z: desciscion boundary $= w^T x + b$

we are suming up the square of all missclassifications (MSAE)

$$\therefore \nabla L(w,b) = \left[\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b}\right]$$

1. calc $\frac{\partial L}{\partial w}$ (chain rule)

$$\frac{\partial L}{\partial w} = \frac{2}{n}\sum_{i=1}^{n}\left(y^{(i)} - \sigma(z^{(i)})\right)\frac{\partial L}{\partial w_j}\left(y^{(i)} - \sum_j w_j x_j^{(i)} + b\right)$$

$$\Rightarrow \frac{\partial L}{\partial w} = \frac{2}{n}\sum_{i=1}^{n}\left(y^{(i)} - \sigma(z^{(i)})\right)\left(-x_j^{(i)}\right)$$

$$\frac{\partial L}{\partial b} = \frac{2}{n}\sum_{i=1}^{n}\left(y^{(i)} - \sigma(z^{(i)})\right)\frac{\partial L}{\partial b}\left(y^{(i)} - \sum_j (w_j x_j^{(i)} + b)\right)$$

$$\Rightarrow \frac{\partial L}{\partial b} = \frac{2}{n}\sum_{i=1}^{n}\left(y^{(i)} - \sigma(z^{(i)})\right)\cdot -1$$

$$\boxed{\nabla L(w,b) = \left[\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b}\right]}$$