Why Adaline is called as a Full-Batch Gradient Descent?

- First of all, Adaline uses MSE as a object function.

$$L(W,b) = MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(z_i)\right)^2$$

MSE is differentiable because $\sigma(z)$, the activate function is also differentiable.

With respect to $W := W + \Delta W$, we are going to follow $-\nabla$'s direction.

So $\Delta W_j = -\eta \nabla_{w_j} L(w,b)$

$$= -\eta \frac{\partial}{\partial w_j} \cdot \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(z_i)\right)^2 = -\eta \cdot \frac{2}{n}\sum_{i=1}^{n}\left(y_i - \sigma(z_i)\right) \cdot \frac{\partial}{\partial w_j} \cdot \left(-W_j^T X_i - b\right)$$

$$= \eta \cdot \frac{2}{n}\sum_{i=1}^{n}\left(y_i - \sigma(z_i)\right) \cdot X_i,$$

So j^{th} weight value $\Delta W_j = \eta \cdot \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(z_i)\right) \cdot X_i$

With respect to $\sum_{i=1}^{n}\underbrace{\left(y_i - \sigma(z_i)\right)}_{\text{Array}} \cdot \underbrace{X_i}_{\text{Array}}$, $\Rightarrow$ Matrix Dot Operation.

So this operation is look like batch-operation, we call Adaline algorithm as a batch algorithm.