

ALGORITHM BLUEPRINT

I) Fprop →

for layer L ,

$$\textcircled{1} \quad z^L = \underset{(n^L, m)}{w^L} \times \underset{(n^{L-1}, m)}{a^{L-1}} + \underset{(n^L, m)}{b^L}$$

$$\textcircled{2} \quad a^L = \underset{(n^L, m)}{g^L(z^L)}, \quad g^L() \text{ is some nonlinearity.}$$

③ a) Save (a^L, z^L, w^L, b^L) to dictionary.
 #return (a^L)

b) if layer L is the last layer,

i) compute cost (scalar) J

ii) Save $(a^L, z^L, w^L, b^L), (J)$ to dictionary.

iii) return (a^L, J)

II) Back Prop →

for Layer L ,

① If L is the last layer,

$$a) \quad \Delta a^L = dJ$$

(scalar) (scalar)

$$b) \quad \Delta z^L = \Delta a^L \cdot a^{L'} \quad , \quad a^{L'} = g^{L'}(z) = \frac{\partial}{\partial z} (\text{non linearity})$$

(n^L, m) (scalar) (n^L, m)

$$c) \quad \Delta a^{L-1} = \underset{(n^{L-1}, m)}{w^{L,T}} \times \underset{(n^L, m)}{\Delta z^L}$$

$$d) \quad \Delta b^L = \Delta z^L \rightarrow (n^L, m)$$

$$\Delta w^L = \Delta z^L \times a^{L-1,T}$$

(n^L, n^{L-1}) (n^L, m) (m, n^{L-1})

Save $(\Delta w^L, \Delta b^L, \Delta a^{L'})$
 return (Δa^{L-1})

② If L is not last layer,

$$a) \quad \Delta a^L = \Delta a^L \text{ from previous iteration.}$$

⋮

same as before.

III) Grad. dec. →

for any layer L ,

$$w^L = w^L - \eta \Delta w^L$$

$$b^L = b^L - \eta \Delta b^L$$

} we can use other update algorithms such as ADAM, etc.

where η is the learning rate.

IV) Loop →

Loop over I → II → III till specified epochs.