I) Fprop →

for layer L,

① $z^L = \omega^L \times a^{L-1} + b^L$
   (n^L, m) (n^L, n^{L-1}) (n^{L-1}, m) (n^L, m)

② $a^L = g^L(z^L)$, $g^L()$ is some nonlinearity.
   (n^L, m) (n^L, m)

③ a) Save $(a^L, z^L, \omega^L, b^L)$ to dictionary
   b) return (a^L)

④ if layer L is the last layer,
   i) compute cost (scalar) J
   ii) Save $(a^L, z^L, \omega^L, b^L), (J)$ to dictionary.
   iii) return (a^L, J)

II) Back Prop →

for Layer L,

① If L is the last layer,
   a) $\Delta a^L = dJ$
      (scalar) (scalar)

   b) $\Delta z^L = \Delta a^L \cdot a^{L'}$, $a^{L'} = g^L(z) \cdot \frac{\partial}{\partial z}$ (nonlinearity)
      (n^L, m) (scalar) (n^L, m)

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

   d) $\Delta b^L = \Delta z^L \rightsquigarrow (n^L, m)$
      $\Delta \omega^L = \Delta z^L \times a^{L-1^T}$
      (n^L, n^{L-1}) (n^L, m) (m, n^{L-1})

② If L is not last layer,
   a) $\Delta a^L = \Delta a^L$ from previous iteration.

      ⋮
      Same as before.

III) Grad. dec. →

for any layer L,
   $\omega^L = \omega^L - \eta \Delta \omega^L$
   $b^L = b^L - \eta \Delta b^L$
   where η is the learning rate.

IV Loop. →

Loop over I→II→III till specified epoches.