KUBRA IQBAL.
Homework 4.

1) Backward step - propagate errors from output to hidden layers.
FORward step → propogate activation from input to output layer.
The forward pass creates the activations list - where activations [i] contains a vector of the activations of neurons in layer [i].
If there are more layers that are hidden layers - propogation will not work. Mostly we have to keep track of activations for each layer.
In the second line of the code, it is helping to find delta. For the dot product we can see the vector output.
For the last step, the value we have is scalor. In this scenario, we are looking for weights - that is why we are doing the dot product.
Transposing - the values are saved in layers and we want to calculate each

node. Backpropagation - nodes are saved in layers. We are transposing, so it becomes a node and a different value. X vector results give us the nabla values.

2) $\boxed{C = y \log(a) + (1-y)\log(1-a)}$

When y is $\longrightarrow$ low value

1) $a = 0.01$ / $y = 0.1$

$c = (0.1)\log(0.01) + (1 - 0.1)\log(1 - 0.01)$
$ce = 0.469$

2) $a = 0.1$ / $y = 0.1$
$ce = 0.325$ —— X

3) $a = 0.2$ / $y = 0.1$
$ce = 0.3618$

4) $a = 0.3$ / $y = 0.1$
$ce = 0.441$

5) $a = 0.4$ / $y = 0.1$

ce = 0.551

6) $a = 0.5 / y = 0.1$
ce = 0.693

7) $a = 0.55 / y = 0.1$
ce =

8) $a = 0.6 / y = 0.1$
ce = 0.875

9) $a = 0.7 / y = 0.1$
ce = 1.119

10) $a = 0.8 / y = 0.1$
ce = 1.470

11) $a = 0.9 / y = 0.1$
ce = 2.08

When Y is $\longrightarrow$ high value

1) $a = 0.01 / y = 0.9$
$c = 0.9 \log(0.01) + (1 - 0.9) \log(1 - 0.01)$

$ce = 4.145$

(2) $a = 0.1$ / $y = 0.9$

$ce = 2.082$

(3) $a = 0.2$ / $y = 0.9$

$ce = 1.470$

(4) $a = 0.3$ / $y = 0.9$

$ce = 1.119$

(5) $a = 0.4$ / $y = 0.9$

$ce = 0.875$

(6) $a = 0.5$ / $y = 0.9$

$Ce = 0.693$

(7) $a = 0.6$ / $y = 0.9$

$ce = 0.5519$

(8) $a = 0.7$ / $y = 0.9$

$ce = 0.4414$

(9) $a = 0.8$ / $y = 0.9$

$Ce = 0.361$

(16) $a = 0.9$ / $y = 0.9$

$ce = 0.325$

(3) Cross entropy for a single neuron can be found from the following equation

$$C = -\frac{1}{n} \sum_x y \log a + (1-y) \log (1-a)$$

After partial derivative

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n} \sum_x \left(\frac{y}{\sigma(z)} - \frac{(1-y)}{1\ \sigma(z)} \frac{\partial \sigma}{\partial w_j} \right)$$

$$= -\frac{1}{n} \sum_x \left[\frac{y}{\sigma(z)} - \frac{(1-y)}{1-\sigma(z)} \right] \sigma'(z) x_j$$

For this situation, as long the derivative of the cost function is taken to respect to w, the x term can not be removed. because it is the result of the partial derivative.

Any cost function will not be able to be removed an input, that will not be possible. This relates with the cost function relying on the weight - which is attached to the x input.