

The second neuron in the second layer is dropped out, this means the gradient with respect to second row of weight  $M^{(2)}$  and the second element of Bias  $B^{(2)}$  will be zero.

The second and third neurons in the third layer are dropped out, this means the gradient with respect to the second and third rows of weights  $M^{(3)}$  and the second and third elements of Bias  $B^{(3)}$  will be zero.

Given the loss function with dropout, the architecture of this network is constructed in such a way that a dropout matrix  $D_2$  is multiplied to the output of first layer (or the second weight matrix  $M^{(2)}$ ) before passing to the activation function  $S$ ; and another dropout matrix  $D_3$  is multiplied to the output of second layer (or the third weight matrix  $M^{(3)}$ ) before passing to the activation function  $S$ .

Notice that when applying dropout to a layer's output (or its pre-activation weights), the multiplication between the dropout matrix and the previous layer's output matrix is an element-wise multiplication. A zero in the dropout matrix means the corresponding neuron will be set to 0 after element-wise multiplication. Consider a weight matrix:

$[[W_{1,1} \ W_{1,2} \ W_{1,3}],$

$[W_{2,1} \ W_{2,2} \ W_{2,3}],$

$[W_{3,1} \ W_{3,2} \ W_{3,3}]]$

Each row determines the weighted sum for each corresponding neuron in the next layer, i.e.

$[W_{11}, W_{12}, W_{13}]$  determines the weighted sum for the first neuron in the next layer.

If we element-wise multiply this row by  $[0, 0, 0]$ , then first neuron of the next layer will be 0.

In this specific question, the values of second row of  $D_2$  are 0, so the second neuron in the next layer is dropped out. The values of second row and third row of  $D_3$  are 0, so the second and third neurons in the next layer are dropped out.