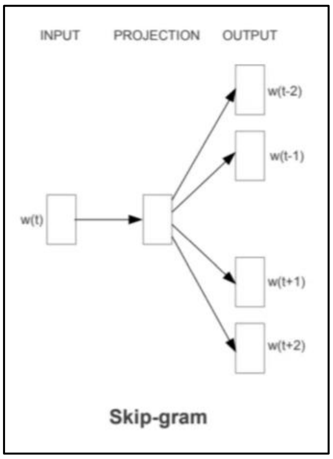
Question 1:

a- Skip-gram model can learn the target word by selecting a word window k and then taking its neighboring words as context words. It will aim to maximize the similarities between the embedding/representation of "shimmered" and its selected neighboring words. Afterwards, it will aim to minimize the similarities (for example, cosine similarity) between other non-neighboring words in this sentence and "shimmered". Also, similarities between "shimmered" and unrelated randomly chosen words may also need to be minimized. The main goal of skip-gram model is learning how to predict neighboring words by looking at a target word using this strategy.

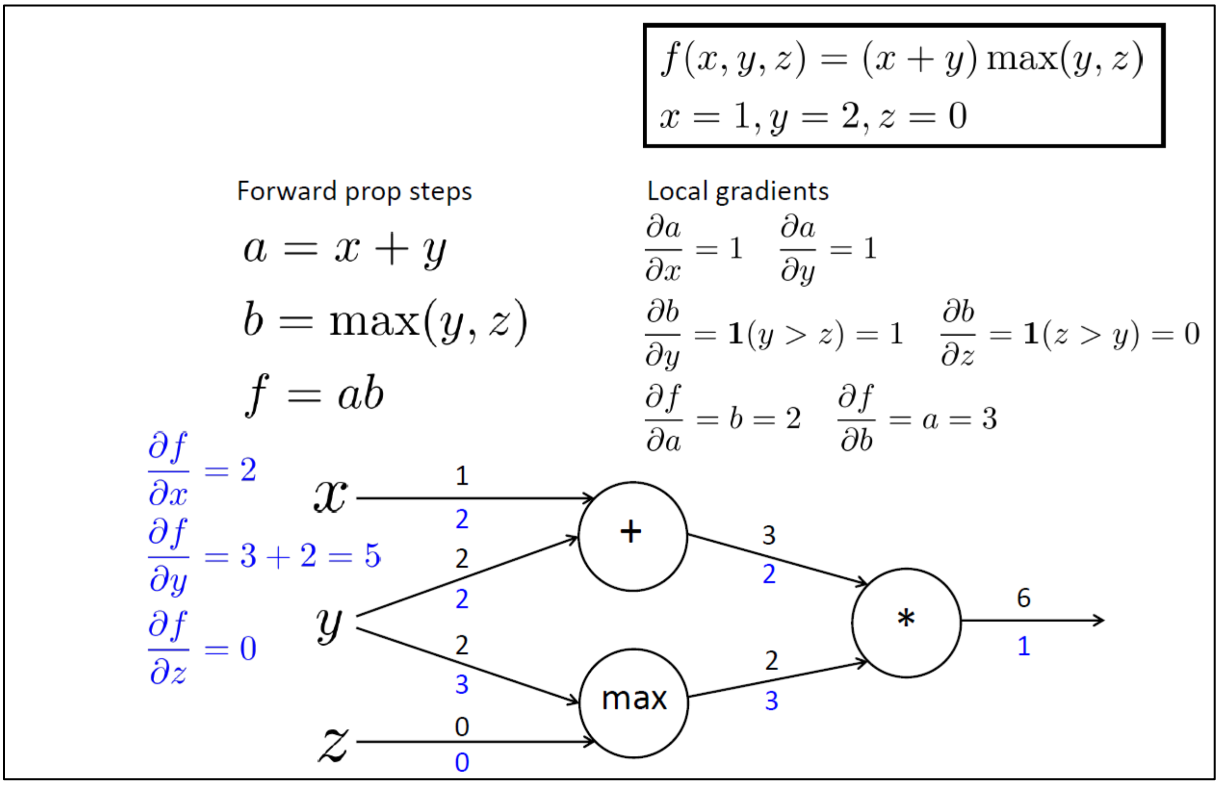


b- k=5 means the total size of our word window should be 5. Our target word ("shimmered") should be in the middle of this window. So, we will be taking its +2 and -2 neighboring words as context words (i.e. positive examples). These are: "wide", "road", "in", "the". These words similarity to "shimmered" should be maximized so that skip-gram model can learn that these positive examples are more likely to be close to the target word.

c- For negative examples, we can choose non-neighboring words like "hot" or "sun" from the sentence. Otherwise, we can randomly select unrelated words as negative examples as well. Then, we can minimize similarities between these negative examples and "shimmered" so that skip-gram model can learn that these negative examples are less likely to be close to the target word.

Question 2:

(You were allowed to omit showing the steps involving chain rule due to their simplicity. So, answer shown below is enough to get full points. However, showing each step with chain rule is a much more appreciated solution.)



$$\boxed{\begin{array}{l} f(x,y,z) = (x+y)\max(y,z) \\ x = 1, y = 2, z = 0 \end{array}}$$

Forward prop steps

$$a = x + y$$

$$b = \max(y, z)$$

$$f = ab$$

Local gradients

$$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$$

$$\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0$$

$$\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3$$

$$\frac{\partial f}{\partial x} = 2$$

$$\frac{\partial f}{\partial y} = 3 + 2 = 5$$

$$\frac{\partial f}{\partial z} = 0$$