Part 1

1.

$$Loss = \sum_t L^{(t)} = \sum_t L(y^{(t)}, \widehat{y^{(t)}})$$

$$\delta V = \eta \cdot \frac{\partial L}{\partial V} = \eta \cdot \sum_t \frac{\partial L^{(t)}}{\partial y^{(t)}} \cdot \frac{\partial y^{(t)}}{\partial O^{(t)}} \cdot h^{(t)}$$

$$\frac{\partial L^{(1)}}{\partial U} = \frac{\partial L^{(1)}}{\partial y^{(1)}} \cdot \frac{\partial y^{(1)}}{\partial O^{(1)}} \cdot \frac{\partial O^{(1)}}{\partial h^{(1)}} \cdot \frac{\partial h^{(1)}}{\partial U}$$

$$\frac{\partial L^{(2)}}{\partial U} = \frac{\partial L^{(2)}}{\partial y^{(2)}} \cdot \frac{\partial y^{(2)}}{\partial O^{(2)}} \cdot \frac{\partial O^{(2)}}{\partial h^{(2)}} \cdot (\frac{\partial h^{(2)}}{\partial U} + \frac{\partial h^{(2)}}{\partial h^{(1)}} \cdot \frac{\partial h^{(1)}}{\partial U})$$

$$\frac{\partial L^{(3)}}{\partial U} = \frac{\partial L^{(3)}}{\partial y^{(3)}} \cdot \frac{\partial y^{(3)}}{\partial O^{(3)}} \cdot \frac{\partial O^{(3)}}{\partial h^{(3)}} \cdot (\frac{\partial h^{(3)}}{\partial U} + \frac{\partial h^{(3)}}{\partial h^{(2)}} \cdot \frac{\partial h^{(2)}}{\partial U} + \frac{\partial h^{(3)}}{\partial h^{(2)}} \cdot \frac{\partial h^{(2)}}{\partial h^{(1)}} \cdot \frac{\partial h^{(1)}}{\partial U})$$

$$\delta U = \eta \cdot \sum_t \frac{\partial L^{(t)}}{\partial U}$$

The upgrade formula for W is similar.

2.

For sigmoid:

$$\sigma'(x) = \sigma(x)\big(1 - \sigma(x)\big) < 1$$

Hence there will be gradient vanishing.

For tanh:

$$\tanh'(x) = \frac{1}{1 + x^2} < 1$$

Hence there will also be gradient vanishing.

For relu:

There will not be gradient vanishing problem if $x > 0$.