1. Preliminaries

1. 0
2. 1/x
3. σ(x)(1−σ(x))
4. $sech^2(x)$
5. 8x

2. Forward (code)

3. Backward

1. $\frac{\partial L(U,V)}{\partial U} = \frac{\partial A}{\partial U} + \frac{\partial B}{\partial U} = (\frac{\partial A}{\partial h} + \frac{\partial B}{\partial h})\ \frac{\partial h}{\partial U}$

$$\frac{\partial A}{\partial U} = \frac{\partial y\ log\ (\sigma(V\,h))}{\partial \sigma\ (V\,h)} \times \frac{\partial \sigma\ (V\,h)}{\partial\ V\,h} \times \frac{\partial V\,h}{\partial h} \times \frac{\partial h}{\partial U}$$
$$\frac{\partial A}{\partial U} = \frac{y}{\sigma(V\,h)} \times \sigma(V\,h)\,(1-\sigma(V\,h)) \times V \times \frac{\partial h}{\partial U}$$
$$\frac{\partial A}{\partial U} = y\ (1-\sigma(V\,h))\ V \times \frac{\partial h}{\partial U}$$

$$\frac{\partial B}{\partial U} = \frac{\partial (1-y)\ log(1-\sigma(V\,h))}{\partial (1-\sigma\ (V\,h))} \times \frac{\partial (1-\sigma\ (V\,h))}{\partial\ V\,h} \times \frac{\partial V\,h}{\partial h} \times \frac{\partial h}{\partial U}$$
$$\frac{\partial B}{\partial U} = \frac{1-y}{1-\sigma\ (V\,h)} \times -\sigma(V\,h)\,(1-\sigma(V\,h)) \times V \times \frac{\partial h}{\partial U}$$
$$\frac{\partial B}{\partial U} = -(1-y)\ \sigma(V\,h)\ V \times \frac{\partial h}{\partial U}$$

$\frac{\partial h}{\partial U} = \frac{\partial\ (max\ (tanh\ (\Sigma\ u))}{\partial U}$ substituting sum of some cells of U for $\Sigma\ u$

$\frac{\partial h}{\partial U} = \frac{\partial\ (max\ (tanh\ (U))}{\partial U}$ if I only use the max value, it becomes

$$\frac{\partial h}{\partial U} = \frac{\partial\ (tanh\ (U))}{\partial U}$$
$$\frac{\partial h}{\partial U} = sech^{\,2}(U)$$

$$\frac{\partial L(U,V)}{\partial U} = (\frac{\partial A}{\partial h} + \frac{\partial B}{\partial h})\ \frac{\partial h}{\partial U}$$
$$\frac{\partial L(U,V)}{\partial U} = [\ [y\ (1-\sigma(V\,h))\ V] - [(1-y)\ \sigma(V\,h)\ V)]\ ] \times \frac{\partial h}{\partial U}$$

$$\frac{\partial L(U,V)}{\partial U} = [\ (y - \sigma(V\,h))\ V\ \ ] \times \frac{\partial h}{\partial U}$$

$$\frac{\partial L(U,V)}{\partial U} = [\ (y - \sigma(V\,h))\ V\ \ ] \times sech^{\,2}(U)$$

2. Implementation (code)

4. Gradient Checking

1. V diff: 0.000000000000000000000036, U diff: 0.00000022