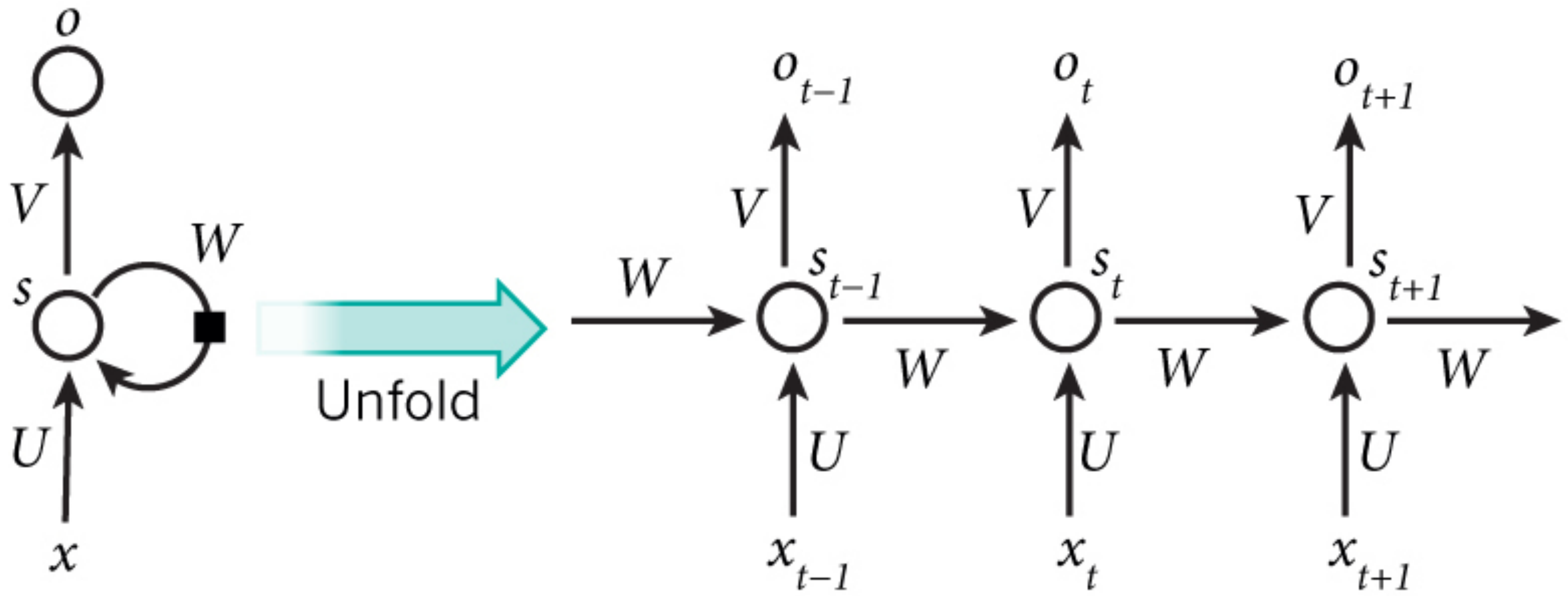
Let's take a simple RNN



$$s_t = f(Ux_t + Ws_{t-1})$$

$$o_t = \mathrm{softmax}(Vs_t)$$

Let's calculate the gradient

- Let's say that U, V, W, x and s are scalars.
- Our formulas will still work for matrices, but it's easier to follow for scalars.
- The hard case is $\frac{\partial s_t}{\partial W}$ because of the recurrent formula for s:

$$s_t = f(Ux_t + Ws_{t-1})$$

Derivative for the 2nd step

Let's denote an argument: a_2

$$s_1 = f(Ux_1 + Ws_0)$$

$$s_2 = f(\boxed{Ux_2 + Ws_1}) = f\big(Ux_2 + Wf(Ux_1 + Ws_0)\big)$$

Im ok until here ->

$$\frac{\partial s_2}{\partial W} = \frac{\partial f}{\partial a_2}\left(W\frac{\partial s_1}{\partial W} + s_1\right) = \boxed{\frac{\partial f}{\partial a_2}W}\frac{\partial s_1}{\partial W} + \boxed{\frac{\partial f}{\partial a_2}s_1}$$

Because s_0 is a constant:

$$\frac{\partial s_1}{\partial W} = \frac{\partial s_1}{\partial W_*}$$

How do we get this? -> $\frac{\partial s_2}{\partial s_1}$

$$\frac{\partial s_2}{\partial W_*}$$

Notation for $s_{j=2}$ in assumption that s_{j-1} is independent of W

The result:

$$\boxed{\frac{\partial s_2}{\partial W} = \frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W_*} + \frac{\partial s_2}{\partial W_*}}$$

Derivative for the 3rd step

$$s_3 = f(Ux_3 + Ws_2)$$

What we know:

$$\boxed{\frac{\partial s_2}{\partial W} = \frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W_*} + \frac{\partial s_2}{\partial W_*}}$$

$$\frac{\partial s_3}{\partial s_2} \qquad \frac{\partial s_3}{\partial W_*}$$

$$\frac{\partial s_3}{\partial W} = \frac{\partial f}{\partial a_3}\left(W\frac{\partial s_2}{\partial W} + s_2\right) = \boxed{\frac{\partial f}{\partial a_3}W}\left(\frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W_*} + \frac{\partial s_2}{\partial W_*}\right) + \boxed{\frac{\partial f}{\partial a_3}s_2}$$

Notation for $s_{j=3}$ in assumption that s_{j-1} is independent of W

The result:

$$\boxed{\frac{\partial s_3}{\partial W} = \frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W_*} + \frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial W_*} + \frac{\partial s_3}{\partial W_*}}$$

Using induction we can get the formula for any step

$$\boxed{\frac{\partial s_2}{\partial W} = \frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W_*} + \frac{\partial s_2}{\partial W_*}}$$

$$\boxed{\frac{\partial s_3}{\partial W} = \frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial s_1}\frac{\partial s_1}{\partial W_*} + \frac{\partial s_3}{\partial s_2}\frac{\partial s_2}{\partial W_*} + \frac{\partial s_3}{\partial W_*}}$$

induction for any positive integer k.

$$\frac{\partial s_k}{\partial W} = \sum_{i=1}^{k}\left(\prod_{j=i+1}^{k}\frac{\partial s_j}{\partial s_{j-1}}\right)\frac{\partial s_i}{\partial W_*}$$