Backward Propagation in RNN

Steps in Backward Propagation

- Calculate the loss by comparing **ŷ** (prediction) and **y** (ground truth)

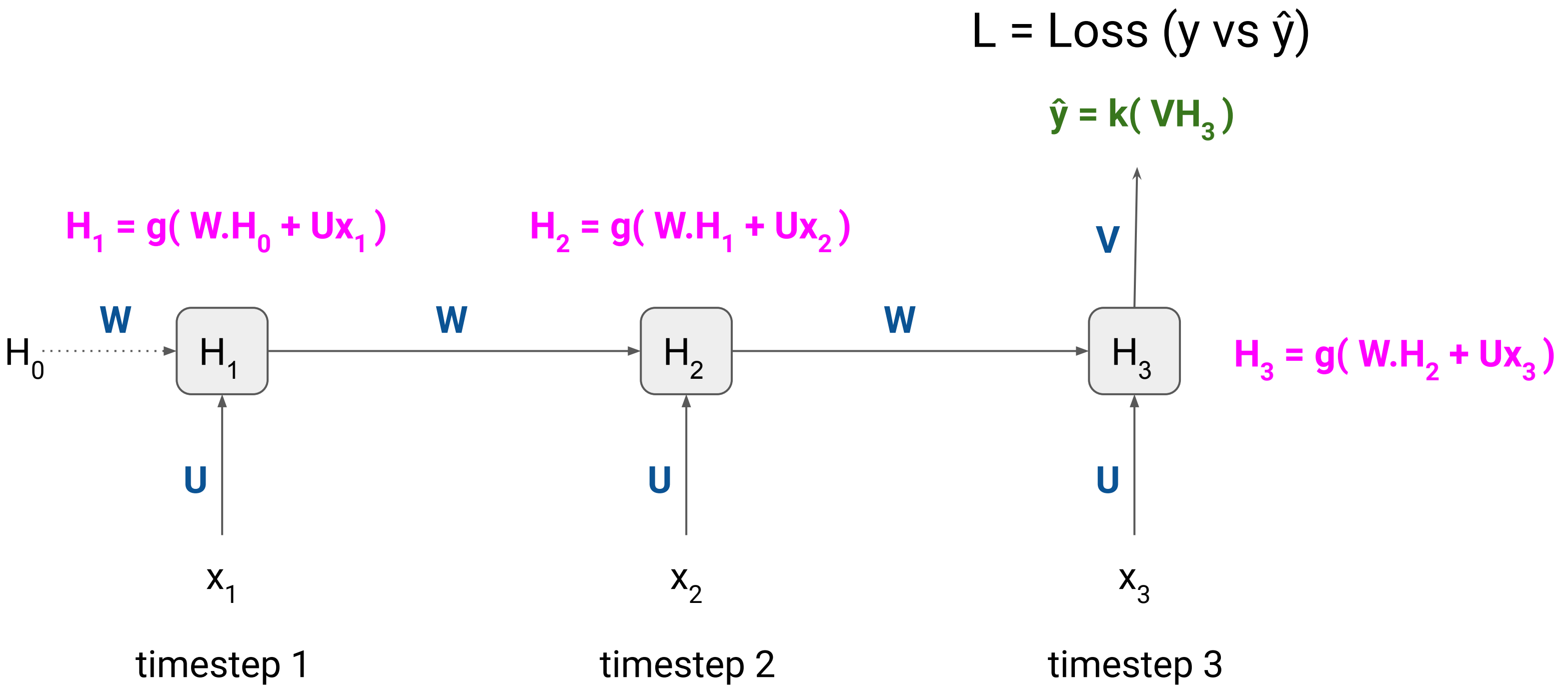
Steps in Backward Propagation

- Calculate the loss by comparing **ŷ** (prediction) and **y** (ground truth)
- Compute gradients with respect to weight matrices **U**, **V**, and **W**

Steps in Backward Propagation

- Calculate the loss by comparing **ŷ** (prediction) and **y** (ground truth)
- Compute gradients with respect to weight matrices **U**, **V**, and **W**
- Update weight matrices **U**, **V**, and **W** by using the gradients

Backward Propagation



Backward Propagation

- **Weights:** V, W, and U

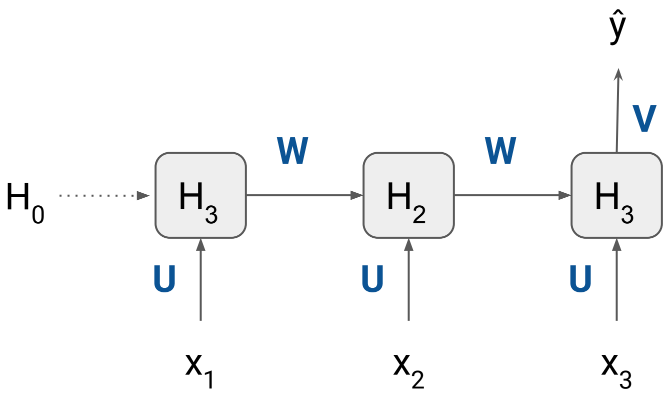
Backward Propagation

- **Weights:** V, W, and U
- **Gradients:** $\partial L/\partial V$, $\partial L/\partial W$, and $\partial L/\partial U$

Backward Propagation

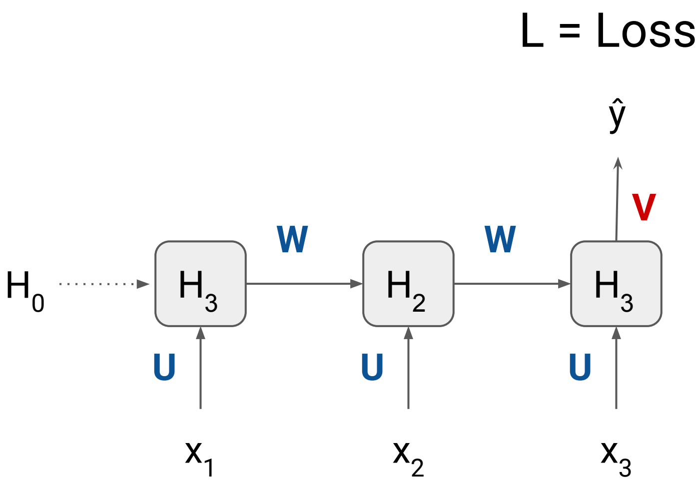
- $\partial L / \partial V = ?$

L = Loss



Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \, . \, (\partial \hat{y}/\partial V)$

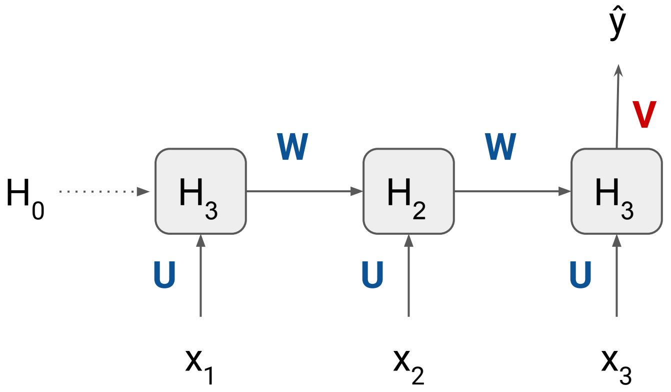


Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \cdot (\partial \hat{y}/\partial V)$

Let $L = \frac{1}{2}(y - \hat{y})^2$

L = Loss



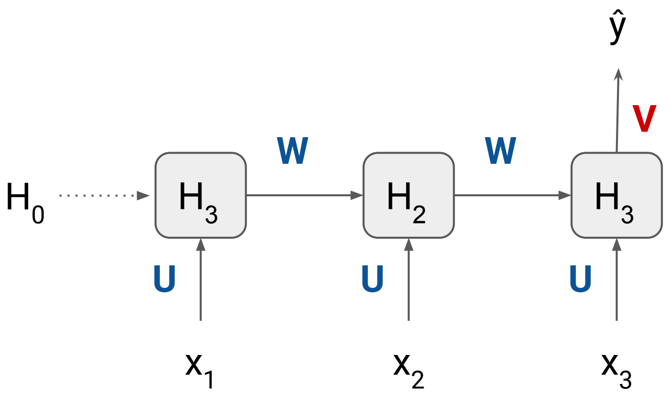
Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \cdot (\partial \hat{y}/\partial V)$

Let $L = ½(y - \hat{y})^2$

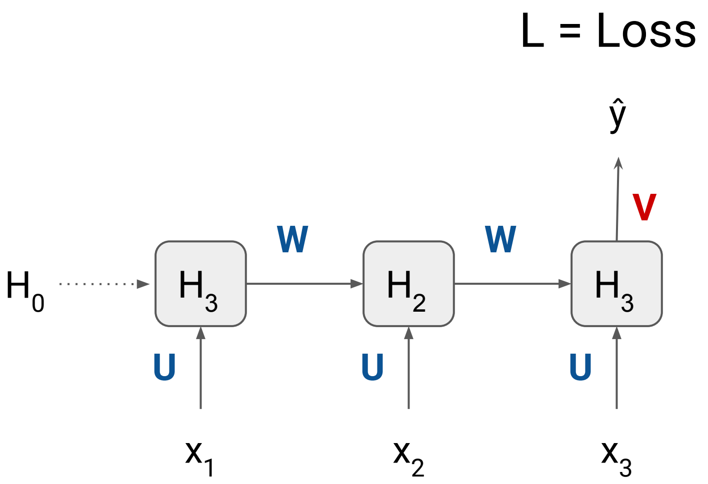
Then $\partial L/\partial \hat{y} = (\hat{y} - y)$

L = Loss



Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \, . \, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \, . \, (\partial \hat{y}/\partial V)$

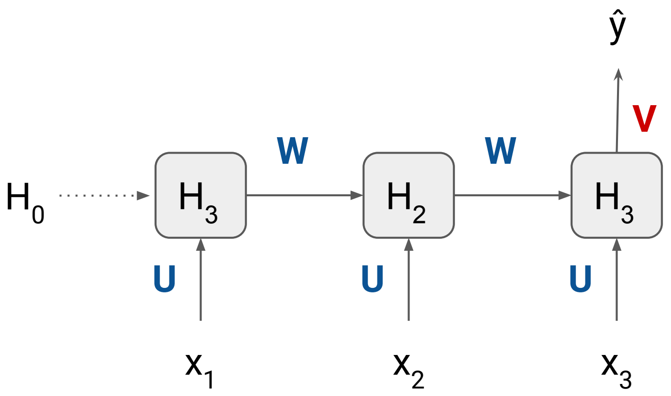


Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (\partial \hat{y}/\partial V)$

$\hat{y} = k(VH_3)$

L = Loss



Backward Propagation

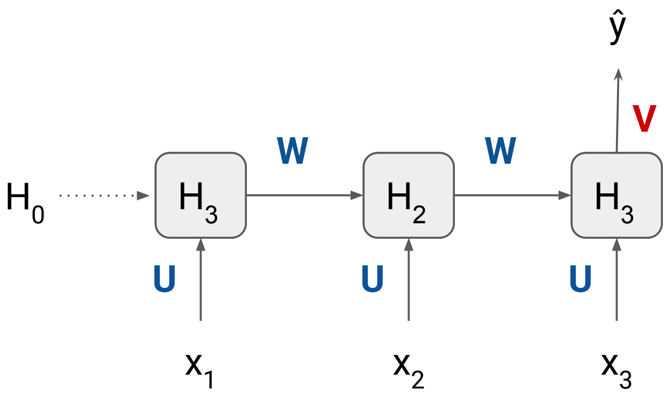
- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (\partial \hat{y}/\partial V)$

$\hat{y} = k(VH_3)$

Assuming Linear activation function

$\hat{y} = VH_3$

L = Loss



Analytics Vidhya
Learn everything about analytics

Backward Propagation

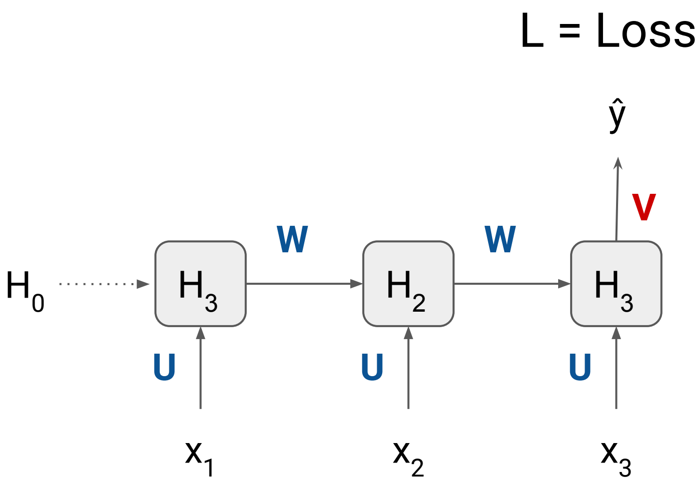
- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (\partial \hat{y}/\partial V)$

$\hat{y} = k(VH_3)$

Assuming Linear activation function

$\hat{y} = VH_3$

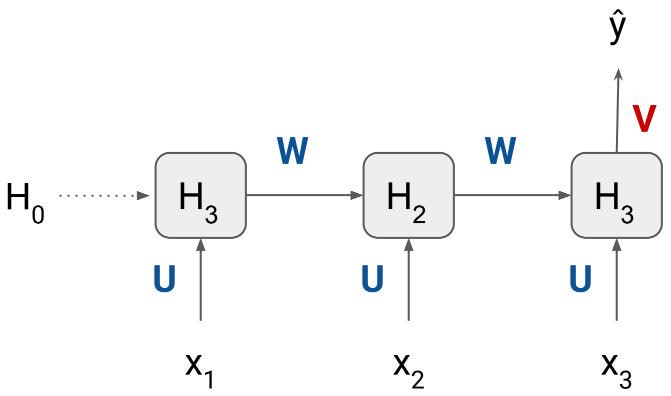
$\partial \hat{y}/\partial V = H_3$



Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

L = Loss

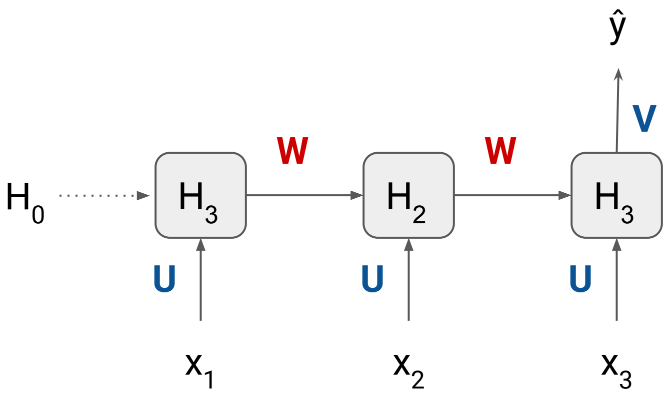


Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

- $\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$

L = Loss



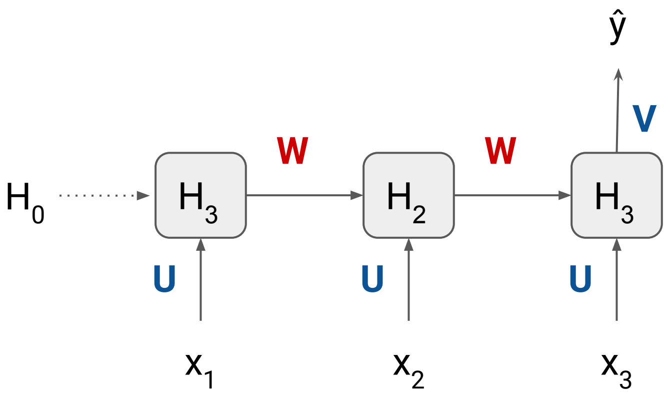
Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

- $\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$

$\hat{y} = VH_3$

L = Loss



Backward Propagation

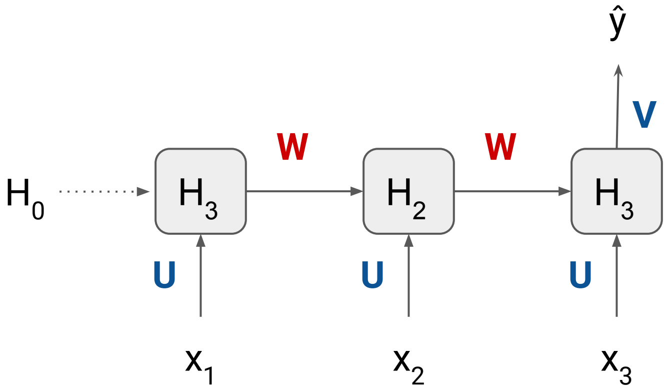
- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

- $\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$

$\hat{y} = VH_3$

$\partial \hat{y}/\partial H_3 = V$

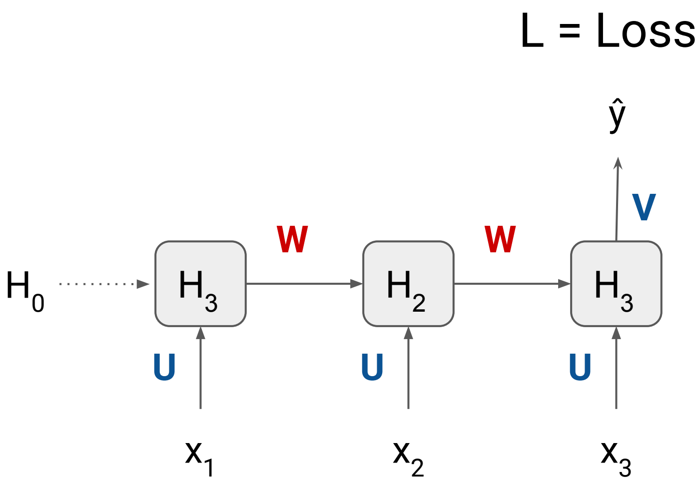
L = Loss



Backward Propagation

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

- $\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$
  $= (\hat{y} - y) \,.\, V \,.\, ?$



Backward Propagation

- $H_3 = g(WH_2 + Ux_3)$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial W = (\partial g(z_3)/\partial z_3) \, . \, (\partial z_3/\partial W)$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial W = (\partial g(z_3)/\partial z_3) \cdot \boxed{(\partial z_3/\partial W)}$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial W = (\partial g(z_3)/\partial z_3) \,.\, (\partial z_3/\partial W)$

$$= (\partial g(z_3)/\partial z_3)[\, H_2 + W(\partial H_2/\partial W)]$$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial W = (\partial g(z_3)/\partial z_3) \,.\, (\partial z_3/\partial W)$

  $= (\partial g(z_3)/\partial z_3)[\, H_2 + W(\partial H_2/\partial W)]$

- $\partial H_2/\partial W = (\partial g(z_2)/\partial z_2)[\, H_1 + W(\partial H_1/\partial W)]$ ...where $z_2 = WH_1 + Ux_2$

Analytics Vidhya
Learn everything about analytics

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial W = (\partial g(z_3)/\partial z_3) \cdot (\partial z_3/\partial W)$

  $= (\partial g(z_3)/\partial z_3)[\, H_2 + W(\partial H_2/\partial W)]$

- $\partial H_2/\partial W = (\partial g(z_2)/\partial z_2)[\, H_1 + W(\partial H_1/\partial W)]$ ...where $z_2 = WH_1 + Ux_2$
- $\partial H_1/\partial W = (\partial g(z_1)/\partial z_1)[\, H_0 + W(\partial H_0/\partial W)]$ ...where $z_1 = WH_0 + Ux_1$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial W = (\partial g(z_3)/\partial z_3) \;.\; (\partial z_3/\partial W)$

$$= (\partial g(z_3)/\partial z_3)[\, H_2 + W((\partial g(z_2)/\partial z_2)[\, H_1 + W((\partial g(z_1)/\partial z_1)[\, H_0 + W(\partial H_0/\partial W)])]\,)]$$

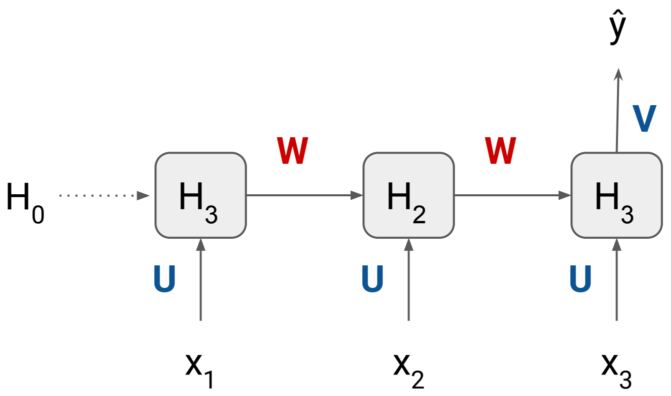
Backward Propagation

L = Loss

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

- $\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$

  $= (\hat{y} - y) \,.\, V \,.\, ?$

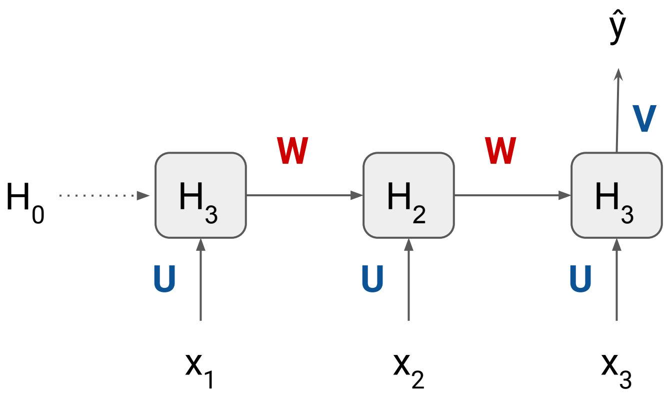


Backward Propagation

L = Loss

- $\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$
  $= (\hat{y} - y) \,.\, (H_3)$

- $\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$
  $= (\hat{y} - y) \,.\, V \,.\, (\partial g(z_3)/\partial z_3)$
  $[\, H_2 + W(\partial H_2/\partial W)]$

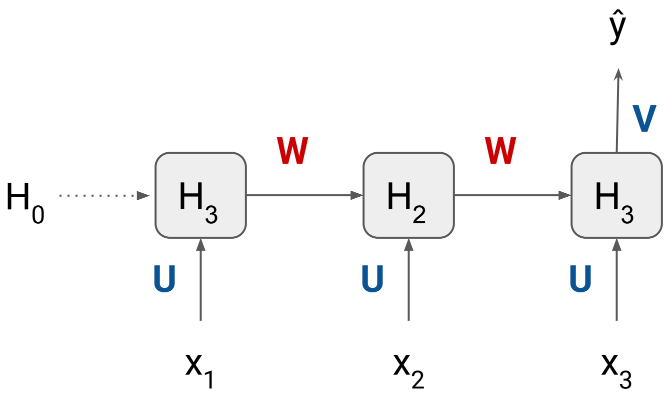


Backward Propagation

- $$\partial L/\partial V = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial V)$$
  $$= (\hat{y} - y) \,.\, (H_3)$$

- $$\partial L/\partial W = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial W)$$
  $$= (\hat{y} - y) \,.\, V \,.\, (\partial g(z_3)/\partial z_3)\,[\, H_2 + W\underbrace{(\partial H_2/\partial W)}_{\text{Recursive}}]$$

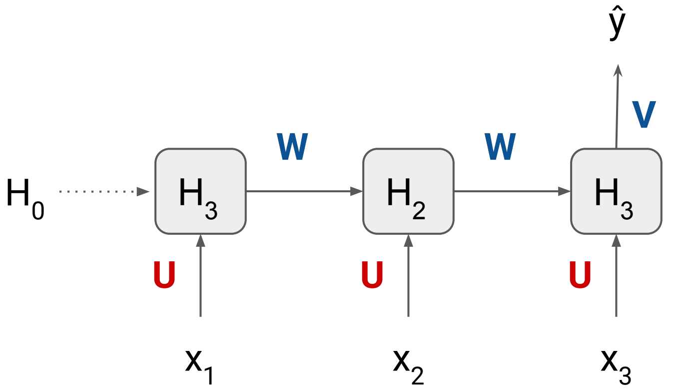
L = Loss



Backward Propagation

- $\partial L/\partial U = (\partial L/\partial \hat{y}) \cdot (\partial \hat{y}/\partial H_3) \cdot (\partial H_3/\partial U)$
  $= (\hat{y} - y) \cdot V \cdot ?$

L = Loss



Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial U = (\partial g(z_3)/\partial z_3) \, . \, (\partial z_3/\partial U)$

$$= (\partial g(z_3)/\partial z_3)[\, x_3 + U(\partial x_3/\partial U) + (\partial WH_2/\partial U) \,]$$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial U = (\partial g(z_3)/\partial z_3) \,.\, (\partial z_3/\partial U)$

$$= (\partial g(z_3)/\partial z_3)[\, x_3 + (\partial WH_2/\partial U) \,]$$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial U = (\partial g(z_3)/\partial z_3) . (\partial z_3/\partial U)$

$$= (\partial g(z_3)/\partial z_3)[\, x_3 + (\partial WH_2/\partial U)\,]$$

- $\partial WH_2/\partial U = W(\partial H_2/\partial U)$

$$= W\,(\partial g(z_2)/\partial z_2) . (\partial z_2/\partial U) \text{ ...where } z_2 = WH_1 + Ux_2$$

$$= W\,(\partial g(z_2)/\partial z_2) . [\, x_2 + (\partial WH_1/\partial U)\,]$$

Backward Propagation

- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial U = (\partial g(z_3)/\partial z_3) \,.\, (\partial z_3/\partial U)$

  $= (\partial g(z_3)/\partial z_3)[\, x_3 + (\partial WH_2/\partial U)\,]$
- $\partial WH_2/\partial U = W\,(\partial g(z_2)/\partial z_2) \,.\, [\, x_2 + (\partial WH_1/\partial U)\,]$
- $\partial WH_1/\partial U = W\,(\partial g(z_1)/\partial z_1) \,.\, [\, x_1 + (\partial WH_0/\partial U)\,]$

Backward Propagation

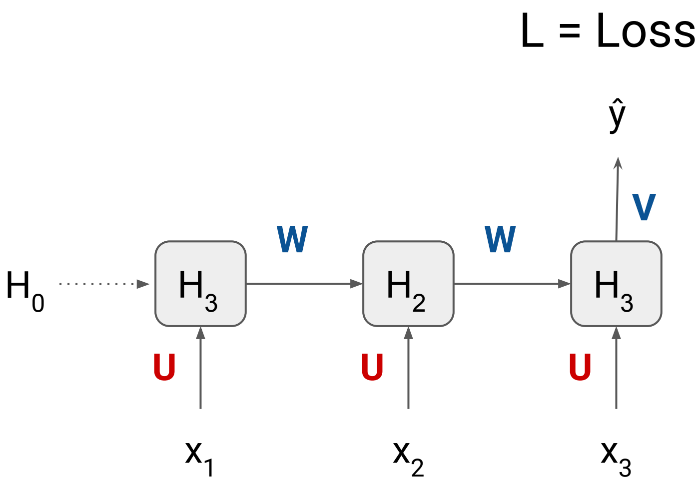
- $H_3 = g(WH_2 + Ux_3) = g(z_3)$
- $\partial H_3/\partial U = (\partial g(z_3)/\partial z_3) . (\partial z_3/\partial U)$

$$
\begin{aligned}
&= (\partial g(z_3)/\partial z_3)[\, x_3 + \\
&\quad (W\, (\partial g(z_2)/\partial z_2) . [\, x_2 + \\
&\quad\quad (W\, (\partial g(z_1)/\partial z_1) . [\, x_1 + \\
&\quad\quad\quad (\partial WH_0/\partial U) \,]) \,]) \,]
\end{aligned}
$$

Backward Propagation

- $\partial L/\partial U = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial U)$

$$= (\hat{y} - y) \,.\, V \,.\, (\partial g(z_3)/\partial z_3) \left[x_3 + \underbrace{(\partial W H_2/\partial U)}_{\text{Recursive}} \right]$$



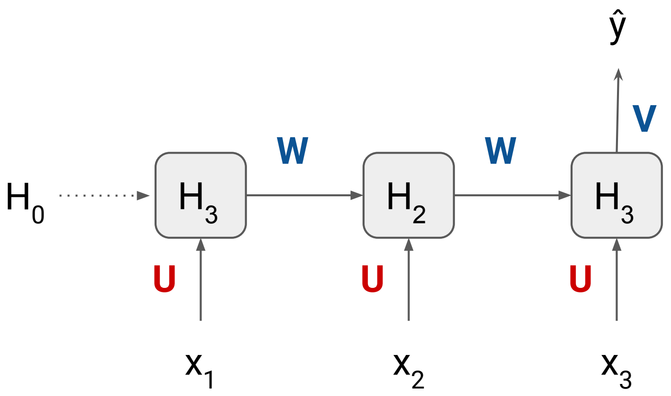
Thank You

Backward Propagation

- $\partial L/\partial U = (\partial L/\partial \hat{y}) \,.\, (\partial \hat{y}/\partial H_3) \,.\, (\partial H_3/\partial U)$

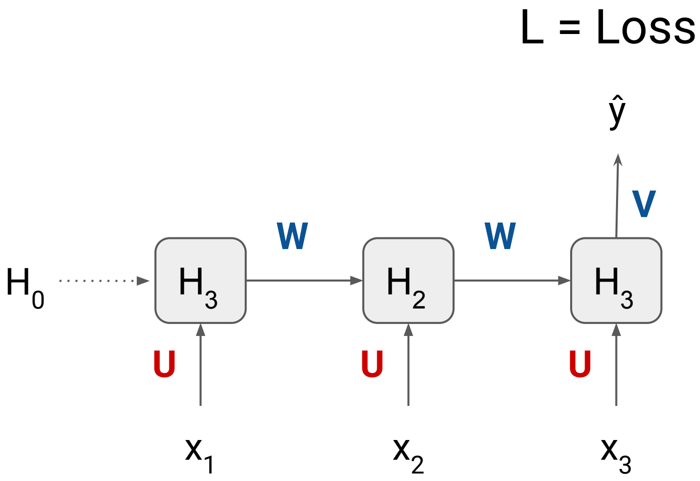
  $= (\hat{y} - y) \,.\, V \,.\, ?$

L = Loss



Backward Propagation

- $$\partial L/\partial U = (\partial L/\partial \hat{y}) \, . \, (\partial \hat{y}/\partial H_3) \, . \, (\partial H_3/\partial U)$$
  $$= (\hat{y} - y) \, . \, V \, . \, (\partial g(z_3)/\partial z_3) \, [\, x_3 + (\partial W H_2/\partial U) \,]$$



max_len = 7

max_len = 7

s_1 = [43, 96, 2, 78, 43]

max_len = 7

$s_1 = [\ 43, 96, 2, 78, 43\ ]$

$s_{1p} = [\ 43, 96, 2, 78, 43, 0, 0\ ]$ (after padding)

max_len = 7

$s_1 = [\ 43, 96, 2, 78, 43\ ]$

$s_{1p} = [\ 43, 96, 2, 78, 43, 0, 0\ ]$ (after padding)

$s_2 = [\ 11, 51, 9, 52, 6, 1, 75, 29\ ]$

$max\_len = 7$

$s_1 = [43, 96, 2, 78, 43]$

$s_{1p} = [43, 96, 2, 78, 43, 0, 0]$ (after padding)

$s_2 = [11, 51, 9, 52, 6, 1, 75, 29]$

$s_{2p} = [11, 51, 9, 52, 6, 1, 75]$ (after truncation)