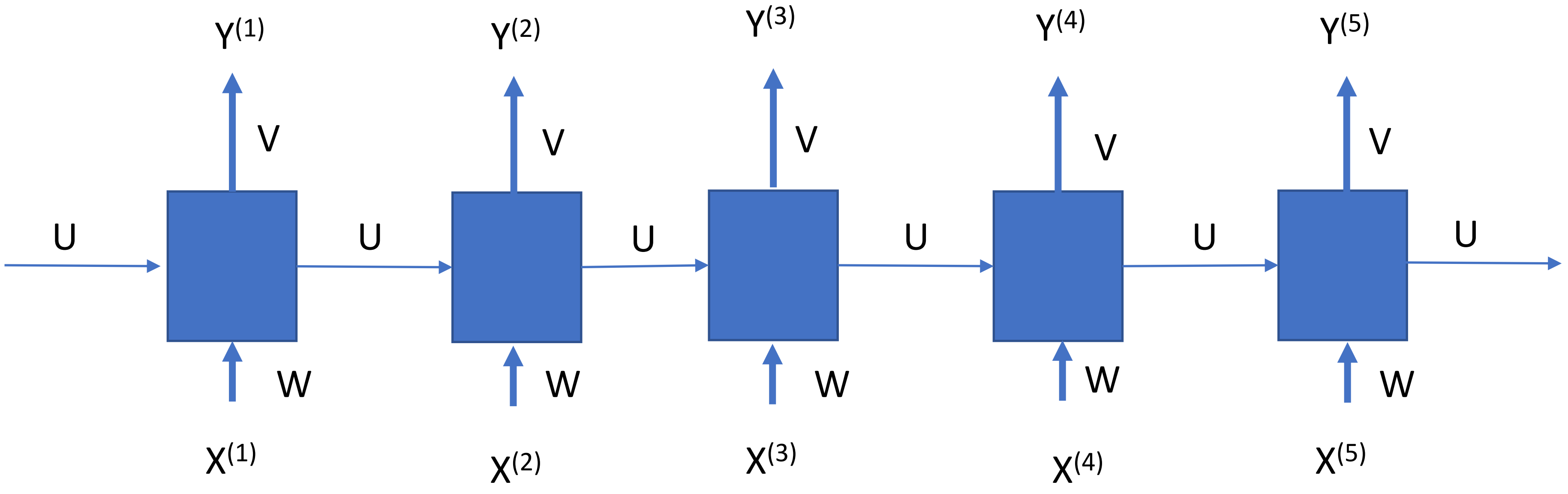
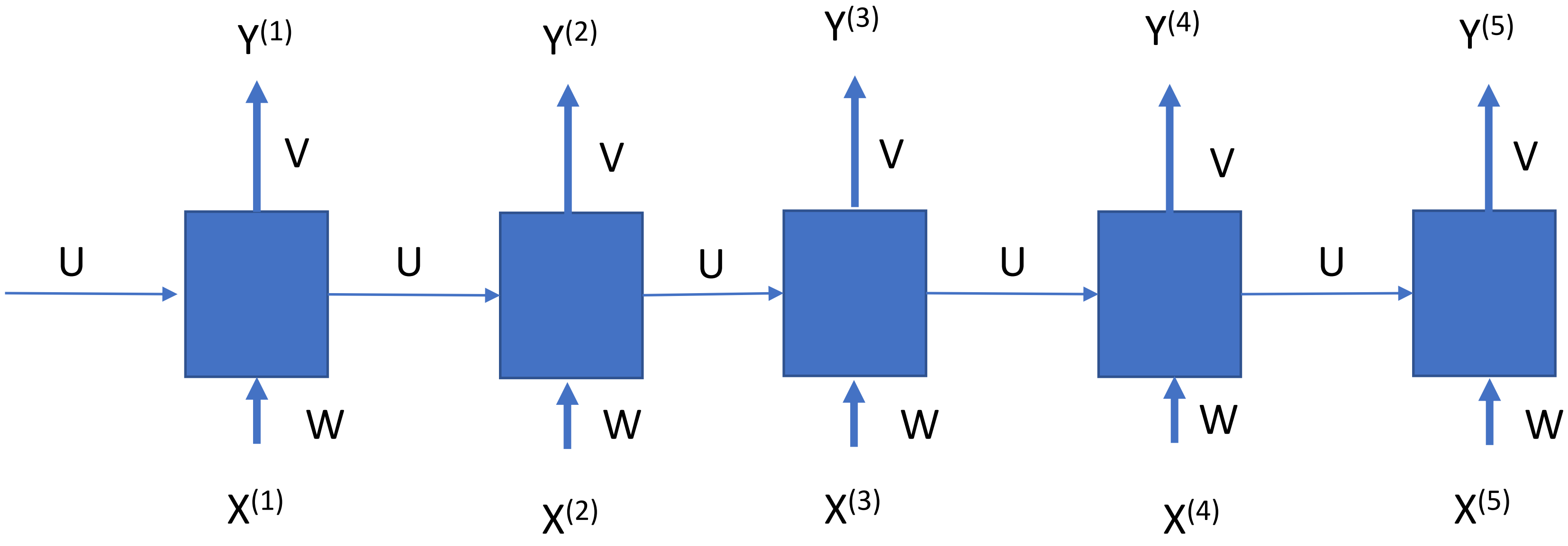
Backpropagation in RNN

Backpropagation through Time

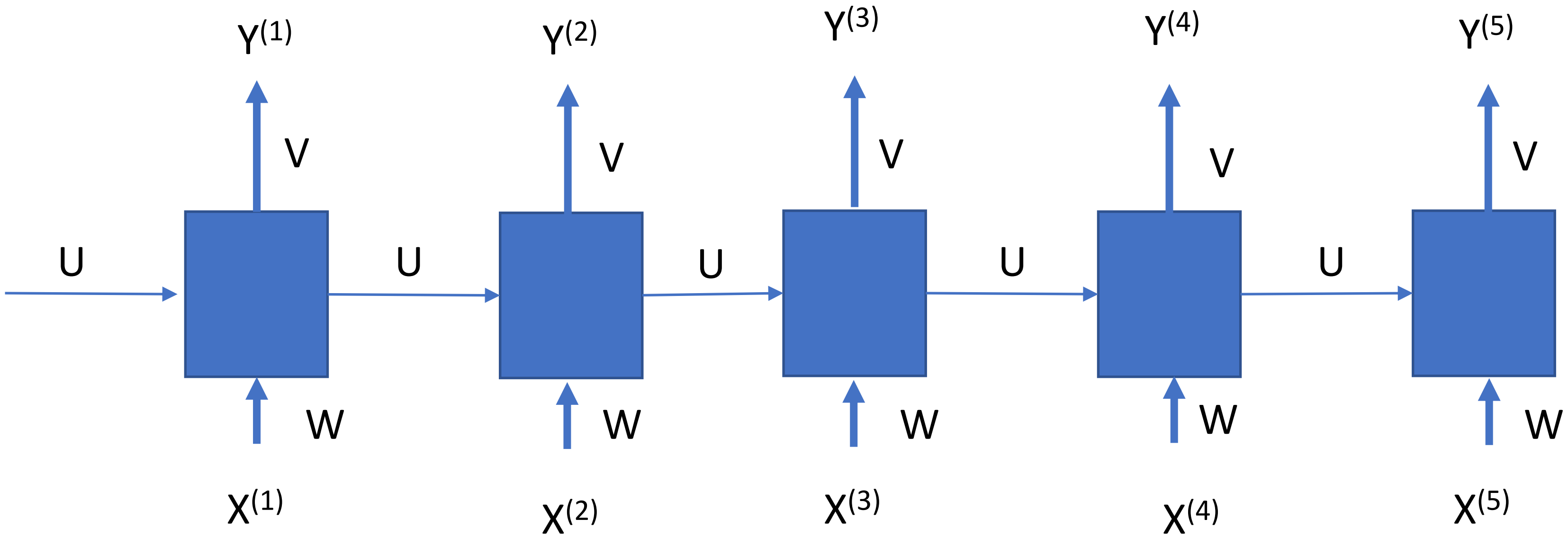


Backpropagation through Time



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right) \qquad \overline{y^t}^T = softmax\left(\overline{h^t}^T V\right) \qquad E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

Backpropagation through Time



$$\frac{\partial E^t}{\partial V}$$

$$\frac{\partial E^t}{\partial W}$$

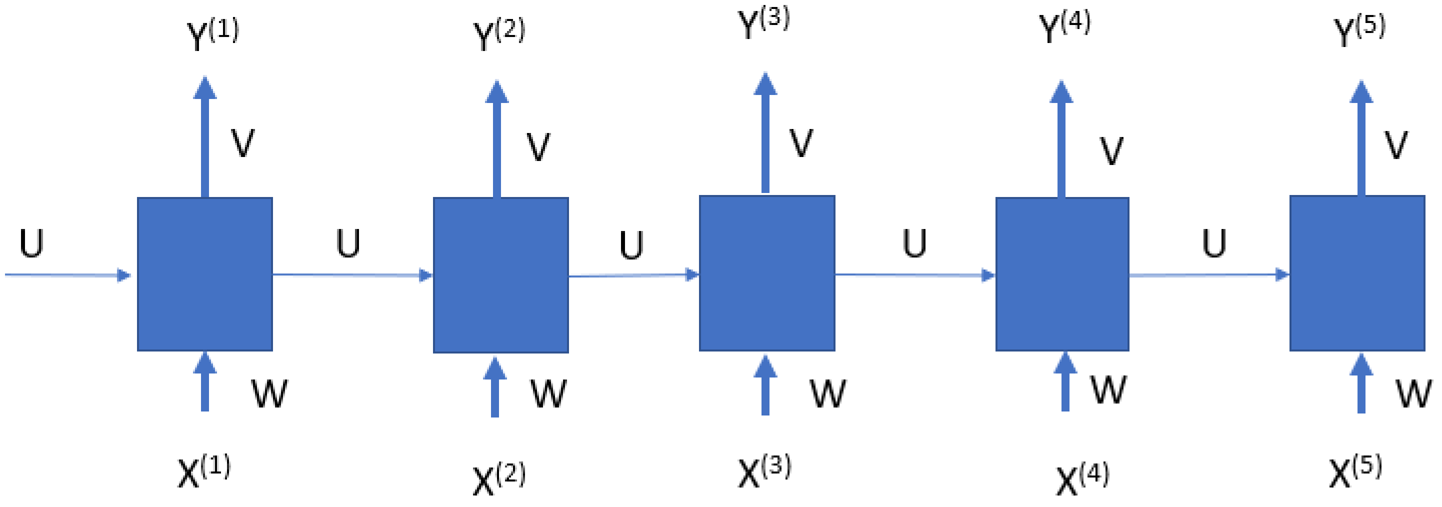
$$\frac{\partial E^t}{\partial U}$$

$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

Estimating The Parameter V Matrix



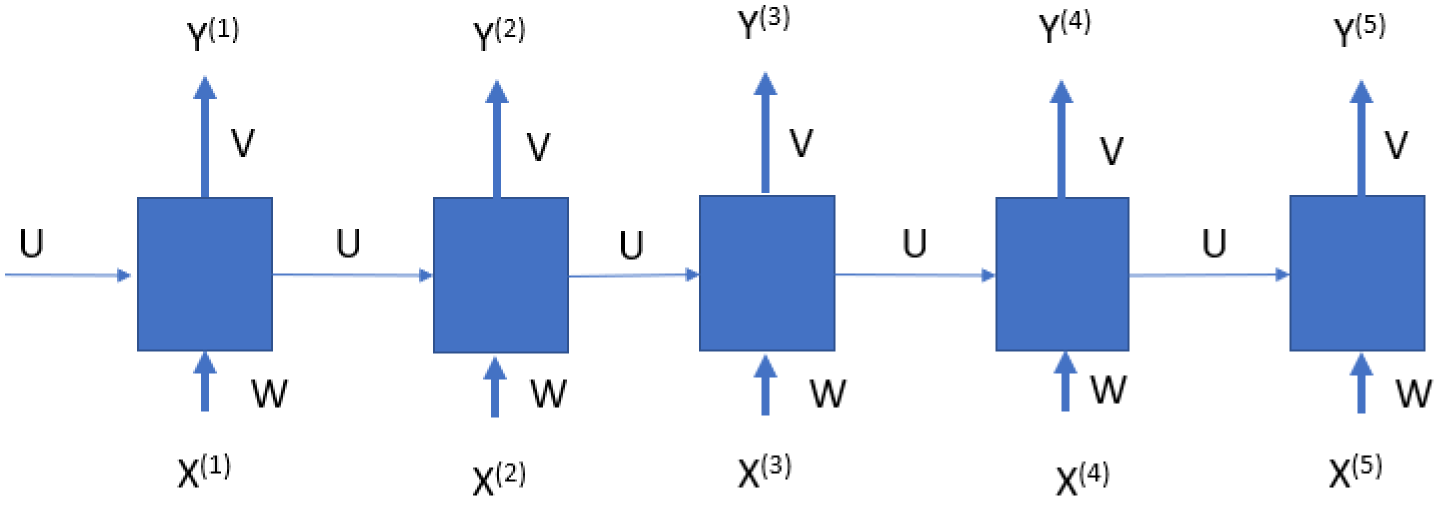
$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

Estimating The Parameter V Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

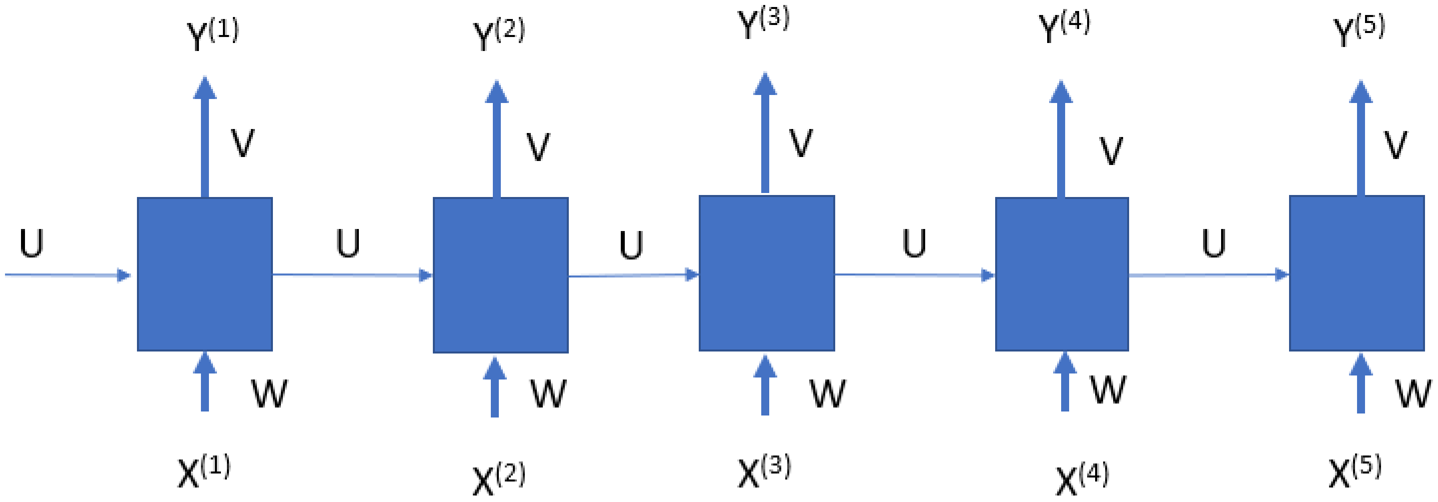
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\overline{y^3}^T = softmax\left(\sigma\left(\overline{x^3}^T W + \sigma\left(\overline{x^2}^T W + \sigma\left(\overline{x^1}^T W + \overline{h^0}^T U\right) U\right) U\right) V\right)$$

Estimating The Parameter V Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

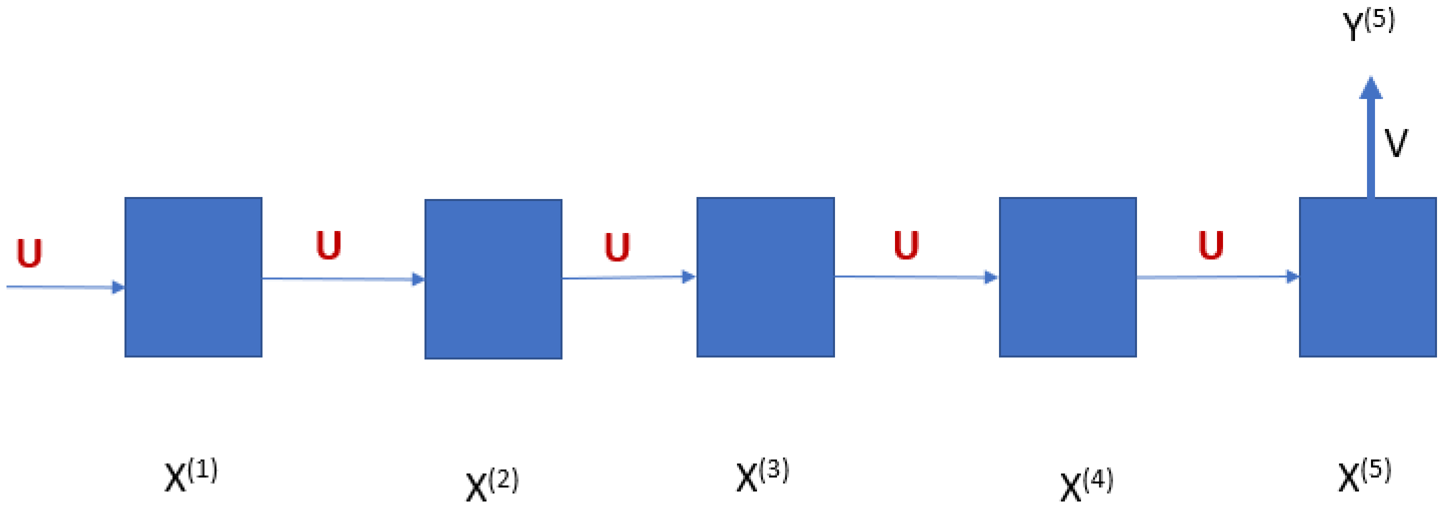
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\overline{y^3}^T = softmax\left(\sigma\left(\overline{x^3}^T W + \sigma\left(\overline{x^2}^T W + \sigma\left(\overline{x^1}^T W + \overline{h^0}^T U\right) U\right) U\right) V\right)$$

It depends only on current time step

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

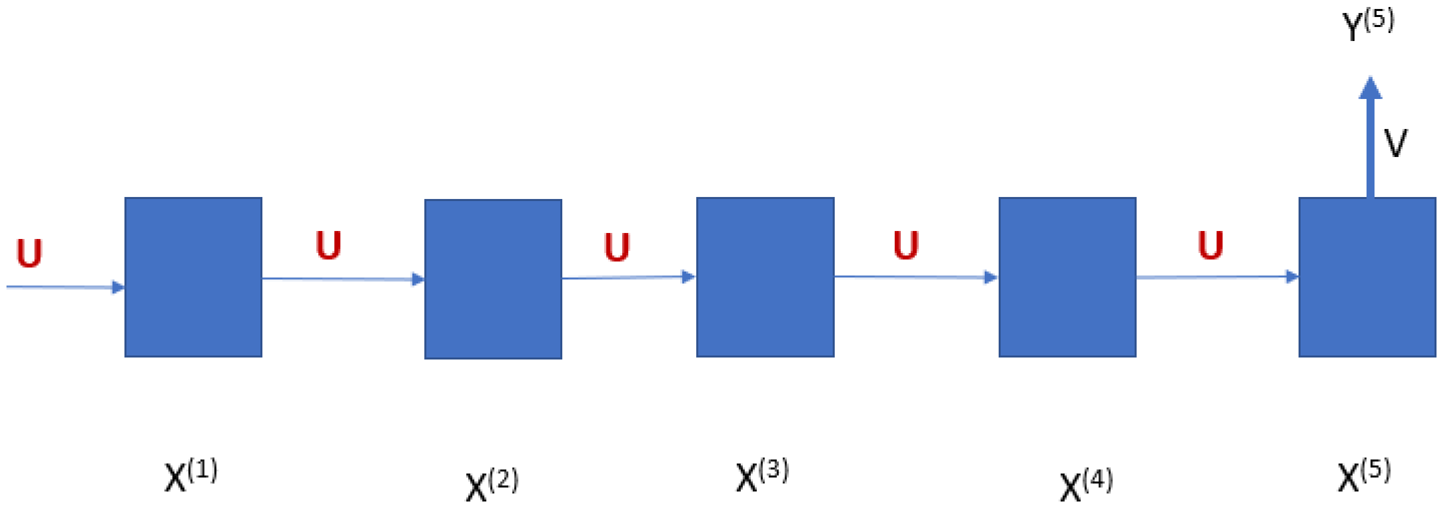
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U}$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

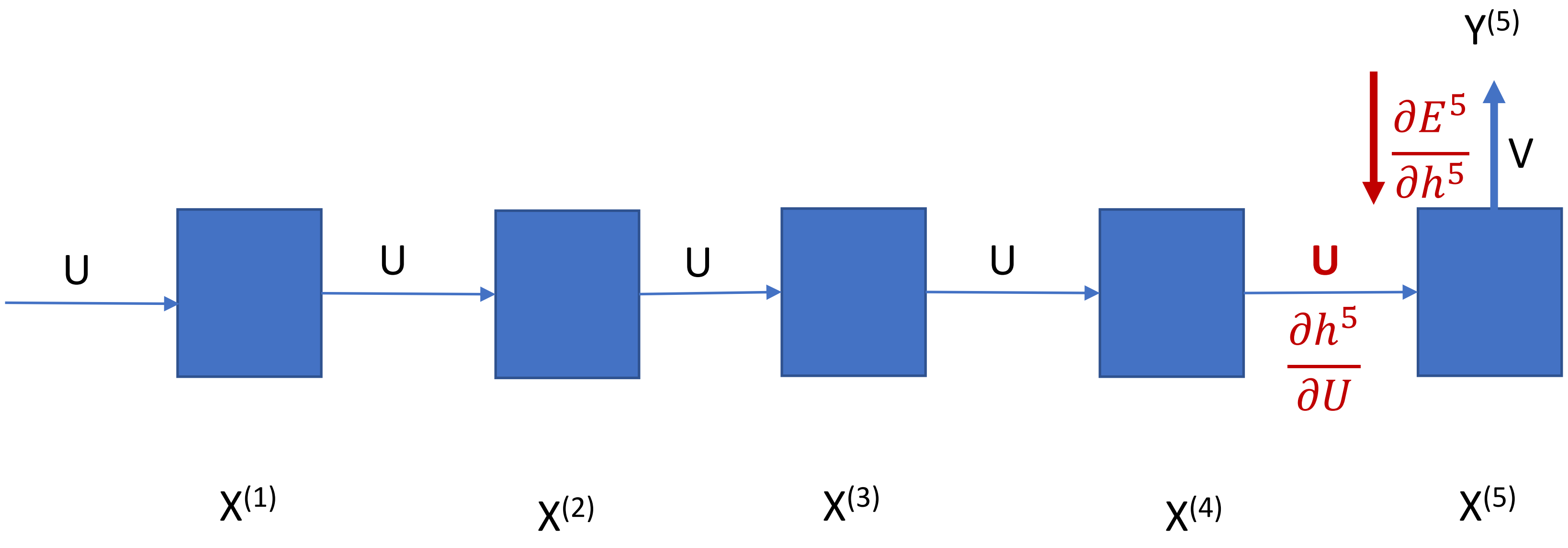
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U}$$

$$\overline{y^3}^T = softmax\left(\sigma\left(\overline{x^3}^T W + \sigma\left(\overline{x^2}^T W + \sigma\left(\overline{x^1}^T W + \overline{h^0}^T U\right) U\right) U\right) V\right)$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

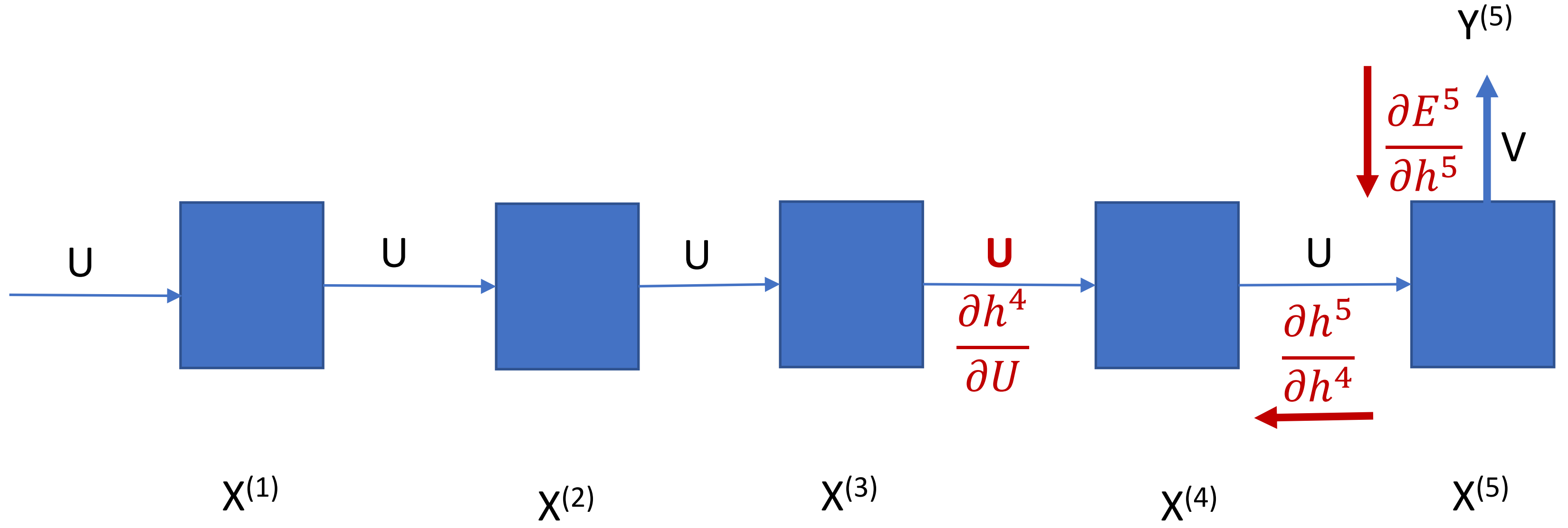
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U}$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

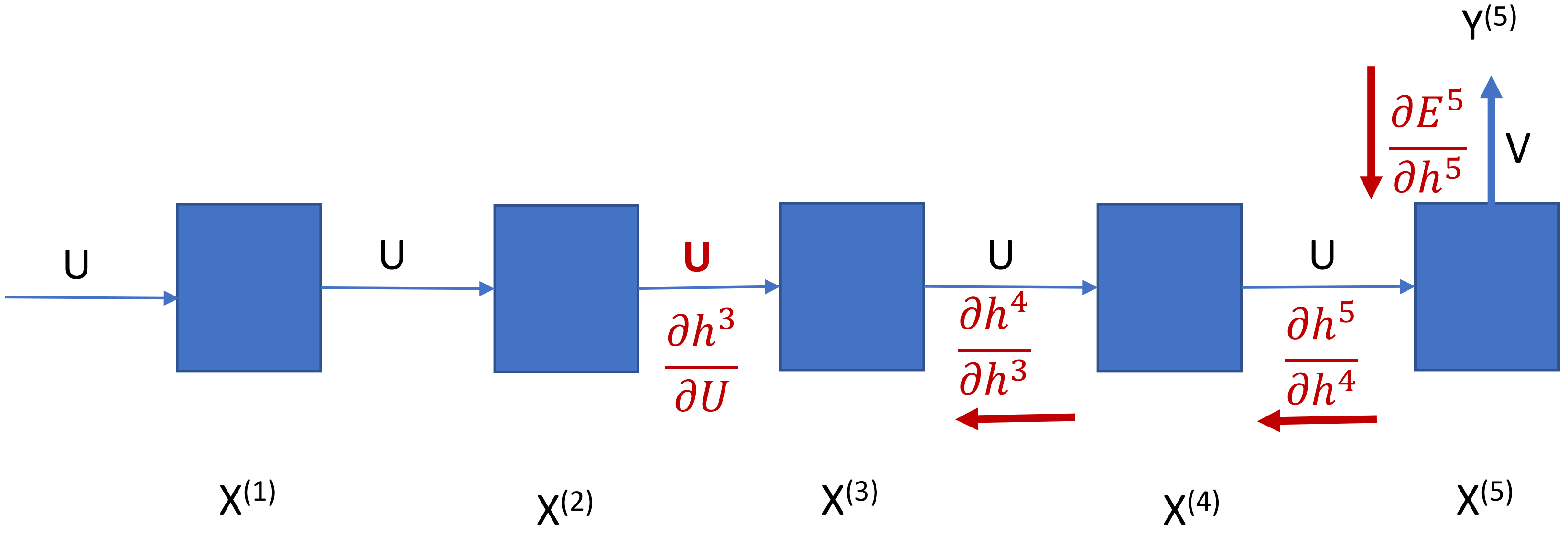
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial U}$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

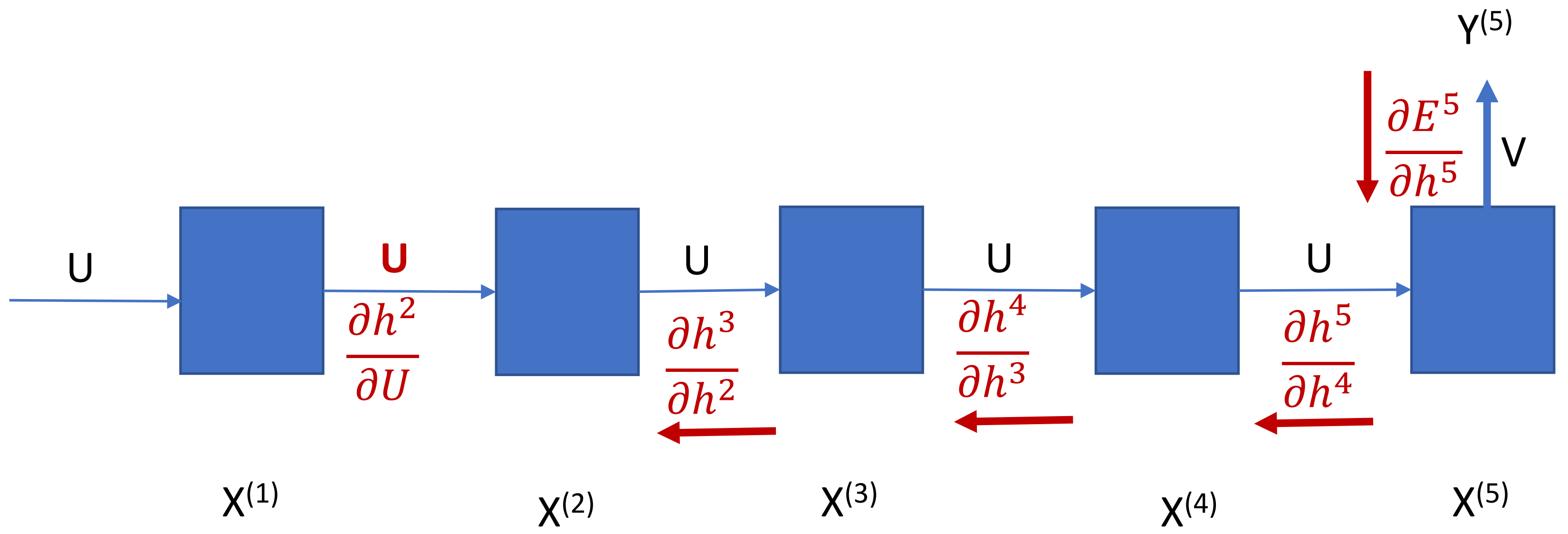
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial U}$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

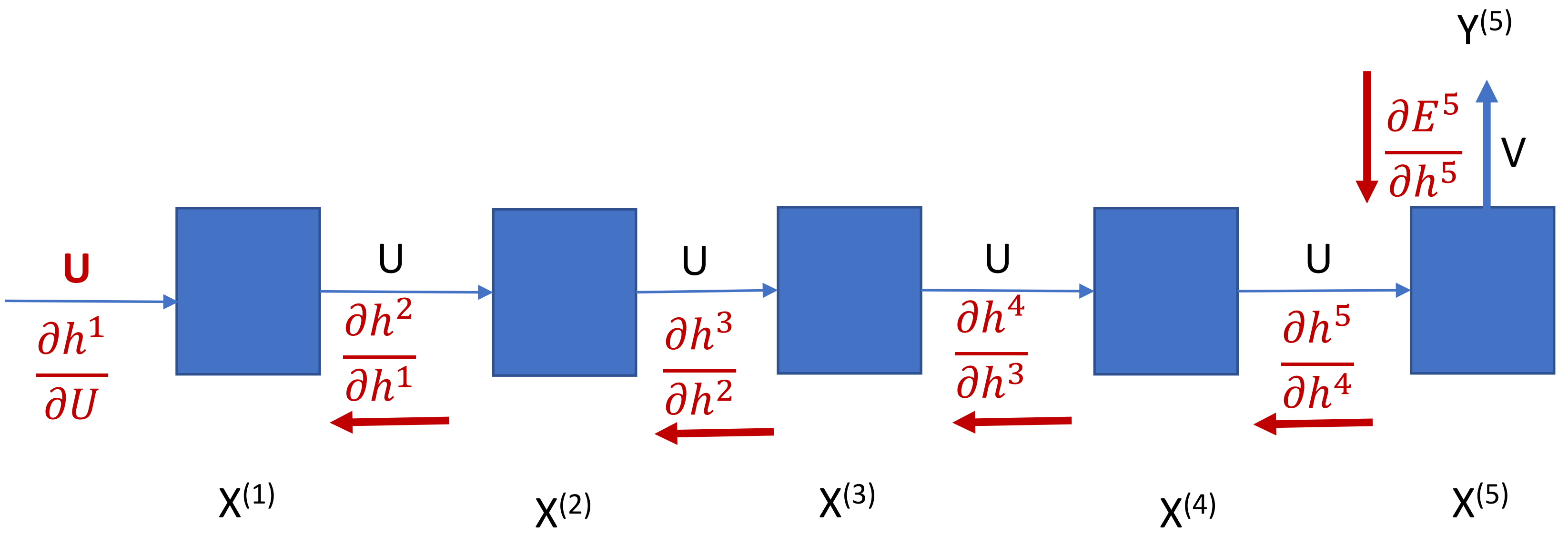
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial U}$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

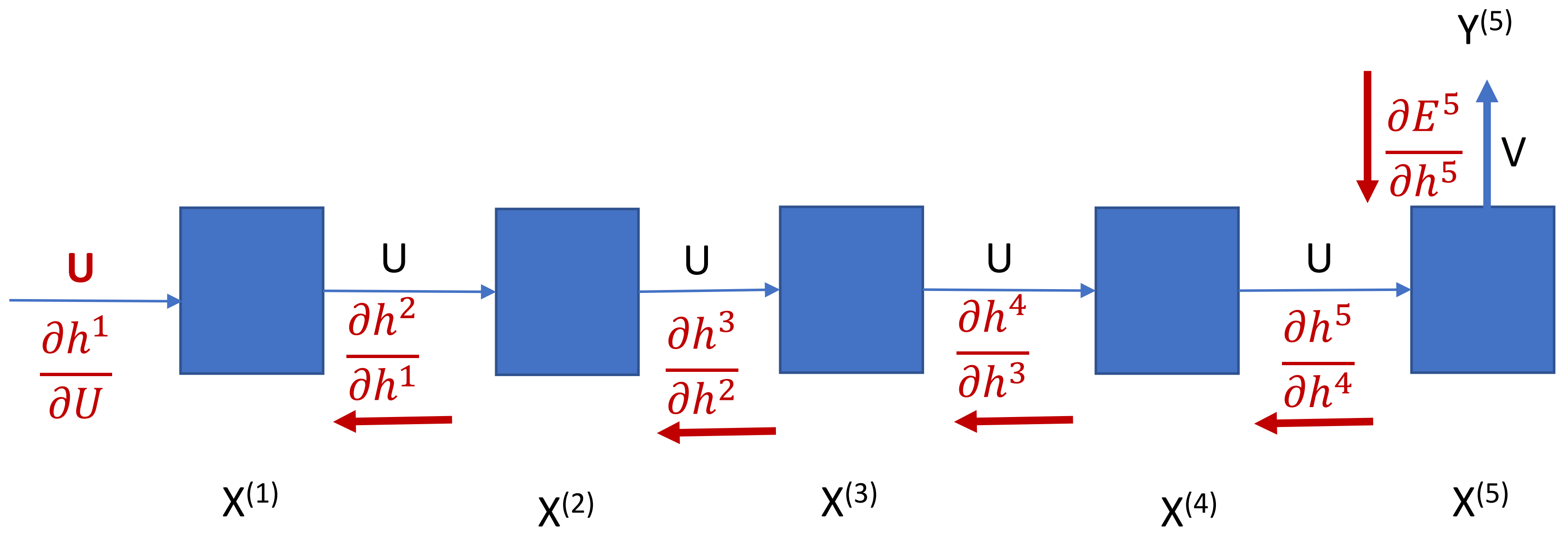
$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial h^1}\frac{\partial h^1}{\partial U}$$

Estimating The Parameter U Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

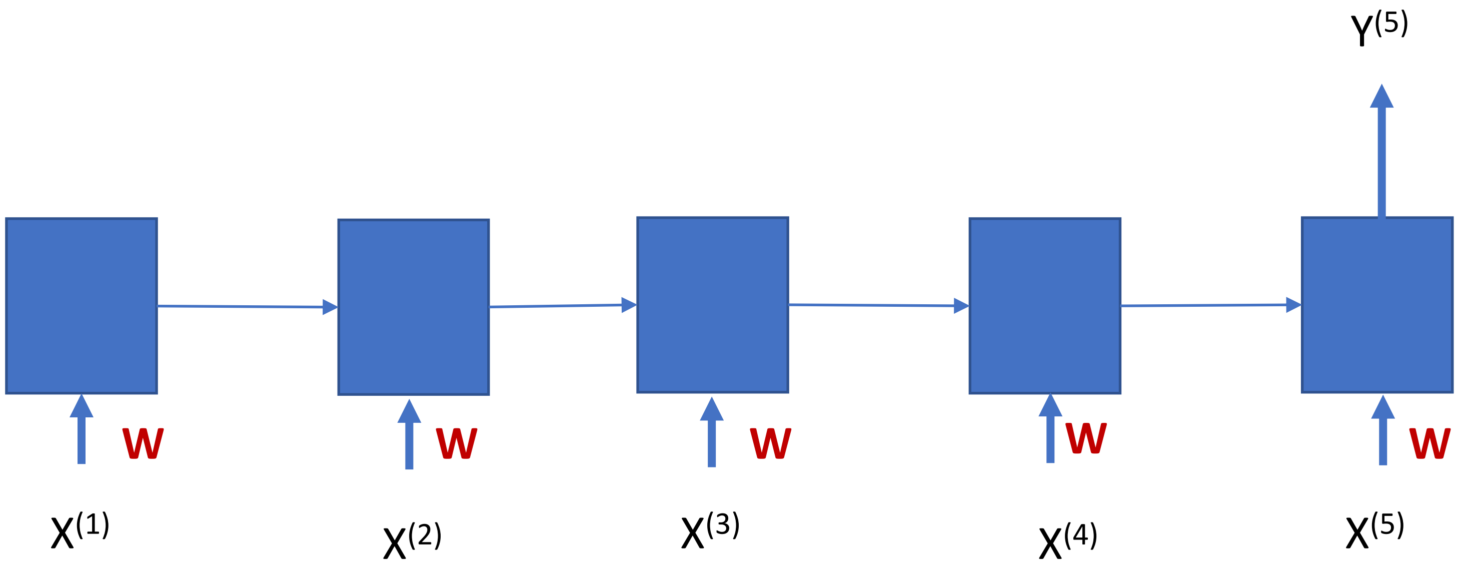
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^5}{\partial U} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial U} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial h^1}\frac{\partial h^1}{\partial U}$$

$$= \frac{\partial E^t}{\partial y^t}\sum_{i=1}^{t}\frac{\partial y^t}{\partial h^i}\frac{\partial h^i}{\partial U}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

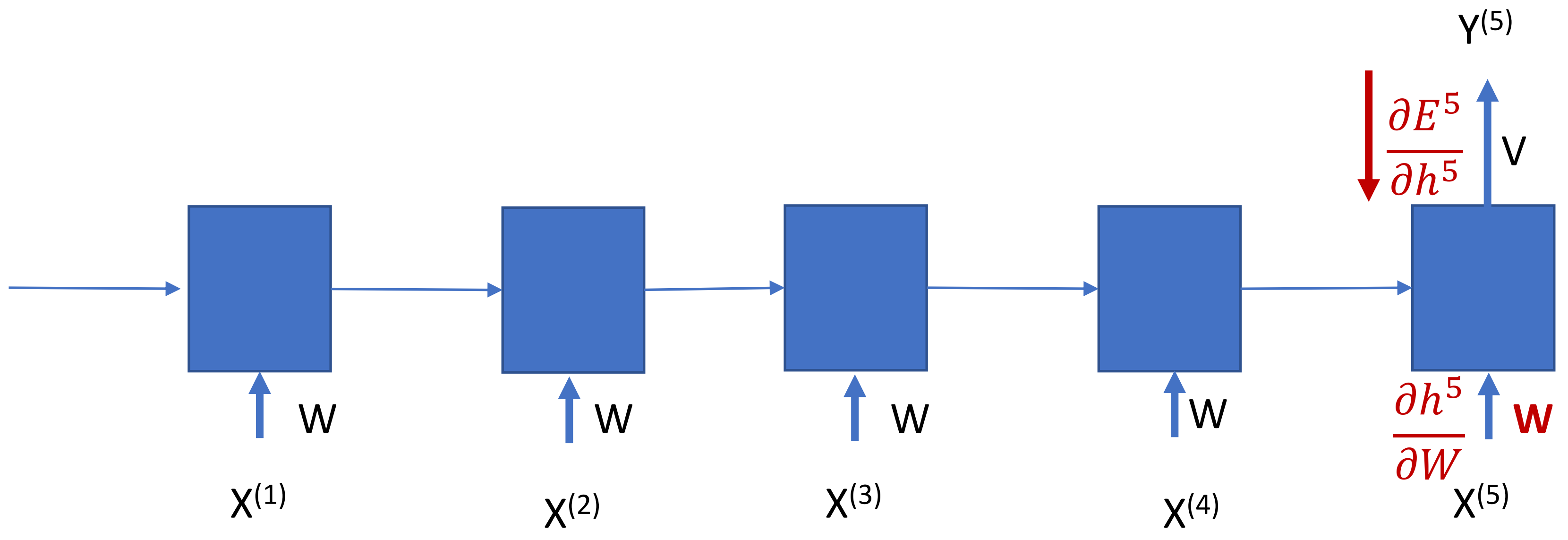
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t}\frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^t}{\partial W}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

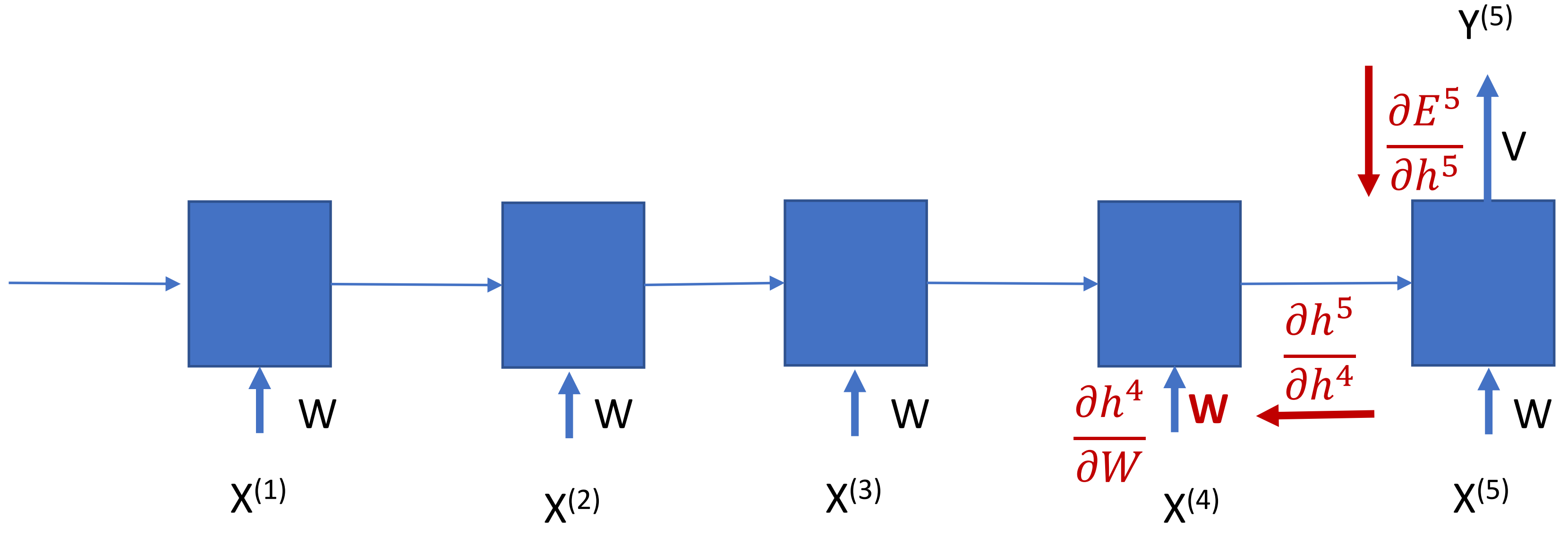
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t}\frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

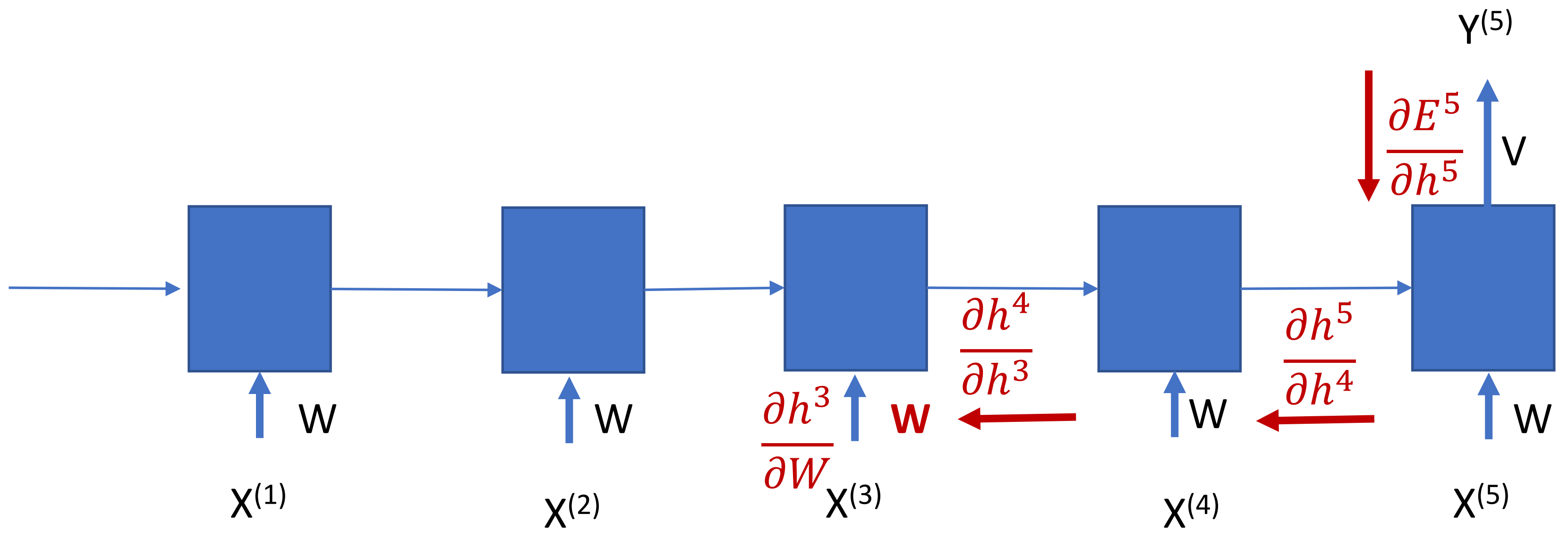
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t}\frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial W}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

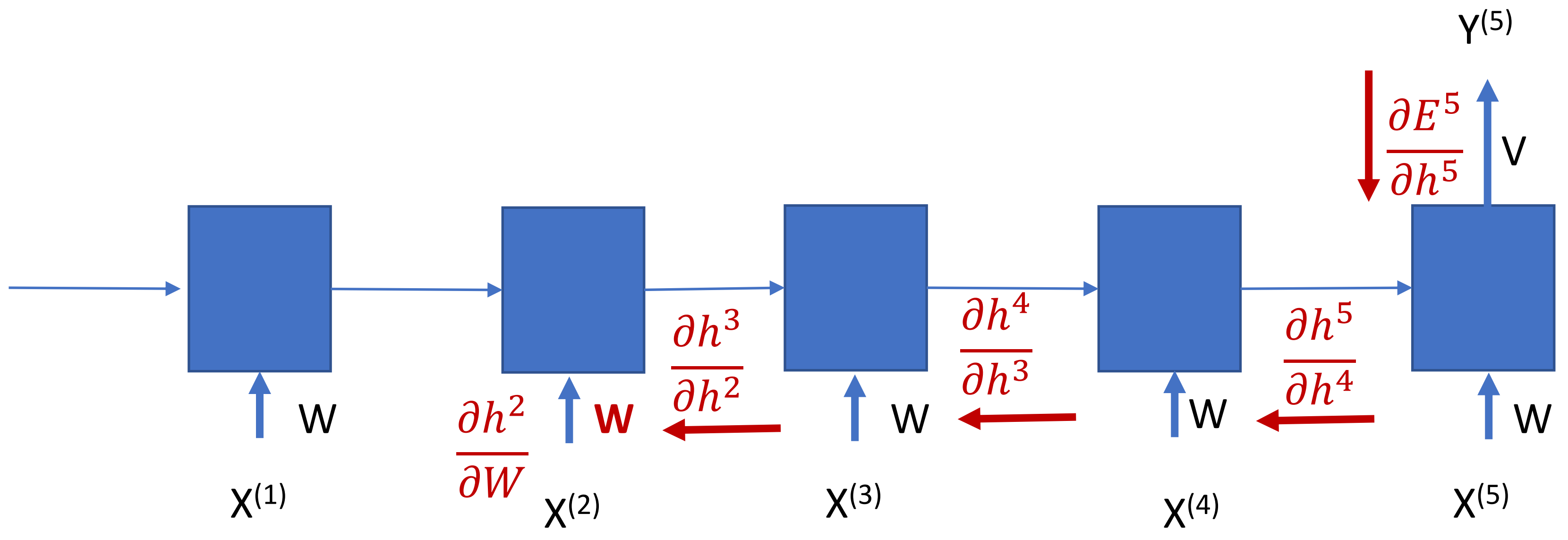
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t}\frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial W}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

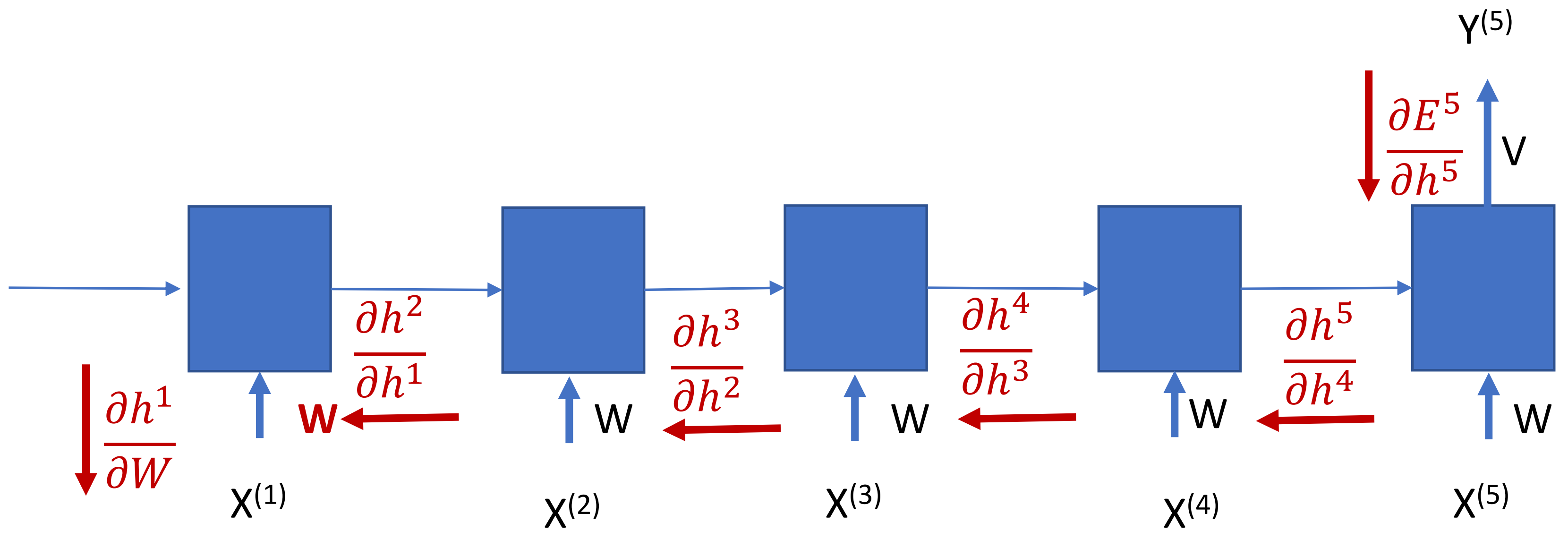
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t}\frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial W}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

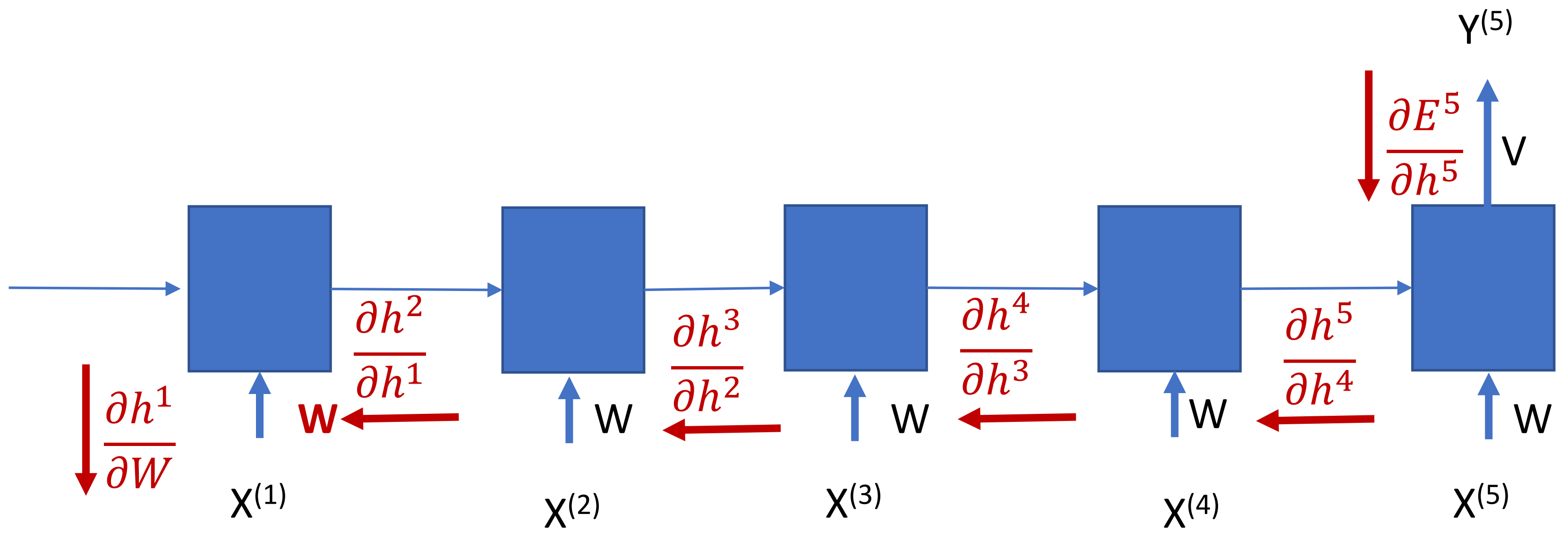
$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t}\frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial h^1}\frac{\partial h^1}{\partial W}$$

Estimating The Parameter W Matrix



$$\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

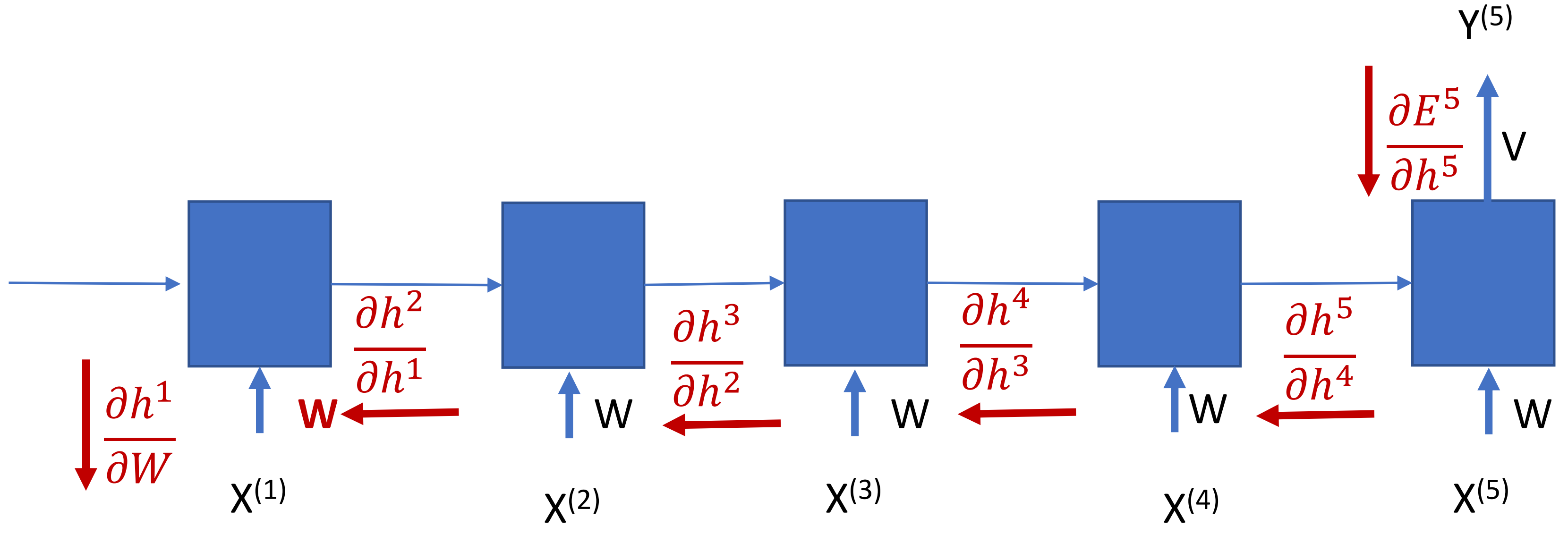
$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t} \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial h^1}\frac{\partial h^1}{\partial W}$$

$$= \frac{\partial E^t}{\partial y^t}\sum_{i=1}^{t}\frac{\partial y^t}{\partial h^i}\frac{\partial h^i}{\partial W}$$

Vanishing/exploding Gradient



$$\boxed{\overline{h^t}^T = tanh\left(\overline{x^t}^T W + \overline{h^{t-1}}^T U\right)}$$

$$\overline{y^t}^T = softmax\left(\overline{h^t}^T V\right)$$

$$E = \sum_t E^t = \sum_t -\acute{y}^t log(y^t)$$

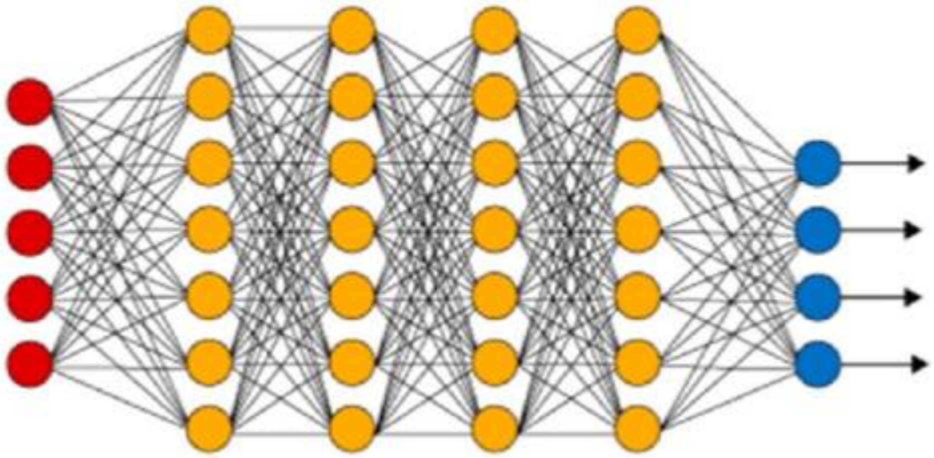
$$\frac{\partial E^t}{\partial V} = \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial z^t}\frac{\partial z^t}{\partial V}$$

$$\frac{\partial E^t}{\partial U} = \sum_{i=1}^{t} \frac{\partial E^t}{\partial y^t}\frac{\partial y^t}{\partial h^t}\frac{\partial h^t}{\partial h^i}\frac{\partial h^t}{\partial U}$$

$$\frac{\partial E^5}{\partial W} = \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\boxed{\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}}\frac{\partial h^3}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\boxed{\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}}\frac{\partial h^2}{\partial W} + \frac{\partial E^5}{\partial y^5}\frac{\partial y^5}{\partial h^5}\boxed{\frac{\partial h^5}{\partial h^4}\frac{\partial h^4}{\partial h^3}\frac{\partial h^3}{\partial h^2}\frac{\partial h^2}{\partial h^1}}\frac{\partial h^1}{\partial W}$$

$$= \frac{\partial E^t}{\partial y^t}\sum_{i=1}^{t}\frac{\partial y^t}{\partial h^i}\frac{\partial h^i}{\partial W}$$

Vanishing/exploding gradient problem may happen in all deep networks



Vanishing/exploding gradient problem may happen in all deep networks