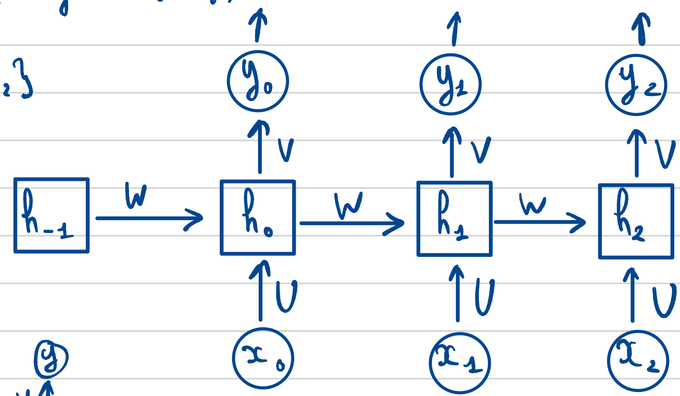
Exercice n°1 : RNN "Elman"

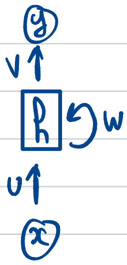
$h_t = \tanh(Ux_t + Wh_{t-1})$ et $y_t = \text{softmax}(Vh_t)$

1) séquence d'entrée : $\{x_0, x_1, x_2\}$

représentation déroulée :



représentation non-déroulée :



2) $y_2 = \text{softmax}(Vh_2)$

$= \text{softmax}(V(\tanh(Ux_2 + Wh_1))$

$= \text{softmax}(V(\tanh(Ux_2 + W(\tanh(Ux_1 + W(\tanh(Ux_0)))))$

Exercice n°2 : implantation Pytorch d'un RNN

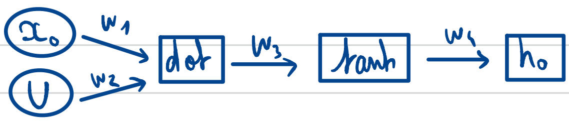
1)
```
class myRNN(nn.Module):
    def __init__(self, in_size, h_size, out_size):
        super().__init__()
        self.U = nn.Linear(in_size, h_size, bias = FALSE)
        self.V = nn.linear(h_size, out_size)
        self.W = nn.linear(h_size, h_size)

    def forward(self, x):
        h = torch.zeros(self.h_size)
        T = x.size(1)
        for t in range(T):
            h = torch.relu(self.U(x[:,t]) + self.W(h))
        y = torch.softmax(self.V(h, dim=1))
        return y
```

2) B=5, T=13, d=7, nb_class = 3, model = myRNN(d, 10, nb_class)
input = torch.randn(B,T,d), out = model(input), print(out.size()) = [5, 3) (B, nb_class)

Exercice n°3 : Backprop Through Time (BPTT)

- gradient $\frac{\partial h_0}{\partial U}$



pass forward :

$w_1 = x_0$, $w_2 = U$

$w_3 = w_1 w_2$

$w_4 = \tanh(w_3)$

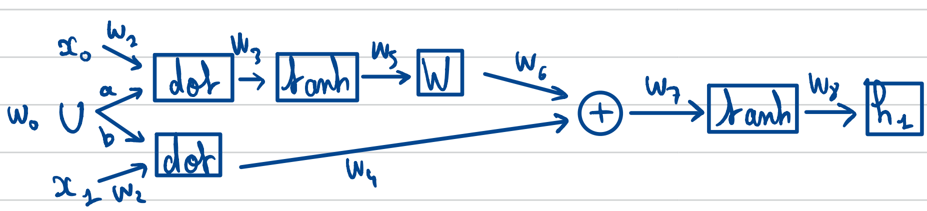
mode reverse :

$\overline{w_2} = (1-w_4^2)\, w_1 = \text{"}dU\text{"} = (1-h_0^2)\, x_0$

$\overline{w_3} = \overline{w_4} \times \frac{\partial w_4}{\partial w_3} = 1 \times (1-w_4^2) = 1-w_4^2$

$\overline{w_4} = \frac{dw_4}{dw_4} = 1$ "seed"

- gradient $\frac{\partial h_1}{\partial U}$:



pass forward :

$w_0 = U$, $w_1 = x_0$, $w_2 = x_1$

$w_3 = w_0 w_1$

$w_4 = w_0 w_2$

$w_5 = \tanh(w_3) = h_0$

$w_6 = W w_5$

$w_7 = w_4 + w_6$

$w_8 = \tanh(w_7) = h_1$

mode reverse : branche $\overline{w_0}^a$

$\overline{w_0}^a = \overline{w_3}\, w_1$

$\overline{w_3} = \overline{w_5}\,(1-\tanh^2(w_3)) = \overline{w_5}\,(1-h_0^2)$

$\overline{w_5} = \overline{w_6}\, W$

$\overline{w_6} = \overline{w_7} \times 1$

$\overline{w_7} = 1 - \overline{w_8} = 1 - h_1^2$

$\overline{w_8} = 1$ "seed"

mode reverse : branche $\overline{w_0}^b$

$\overline{w_0}^b = \overline{w_4}\, w_2$

$\vdots$

$\overline{w_4} = \overline{w_7} \times 1$

$\overline{w_7} = 1 - h_1^2$

$\overline{w_8} = 1$

$$\overline{w_0} = \overline{w_0}^a + \overline{w_0}^b = \text{"}dU\text{"}$$
$$= \overline{w_3}\, w_1 + \overline{w_4}\, w_2$$
$$= (1-h_1^2)\, W (1-h_0^2)\, x_0 + (1-h_1^2)\, x_1$$
$$= (1-h_1^2)\left(x_1 + W(1-h_0^2)\, x_0\right)$$

Généralisation pour h_t :

$$\frac{\partial h_t}{\partial U} = (1-h_t^2)\left(x_t + W \frac{\partial h_{t-1}}{dU}\right)$$