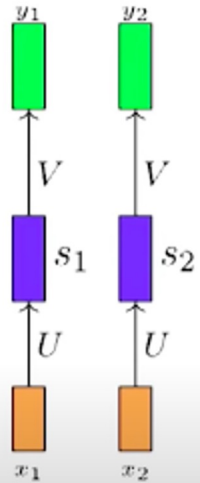
Recurrent Neural Networks

How do we model such tasks involving sequences?

Wishlist

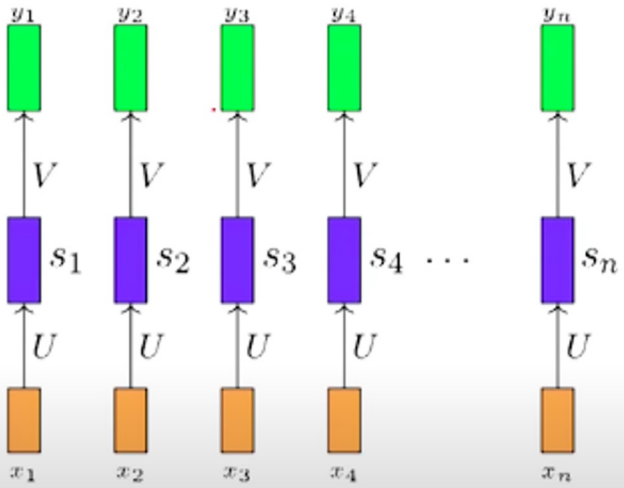
- Account for dependence between inputs
- Account for variable number of inputs
- Make sure that the function executed at each time step is the same
- We will focus on each of these to arrive at a model for dealing with sequences



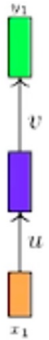
- What is the function being executed at each time step ?

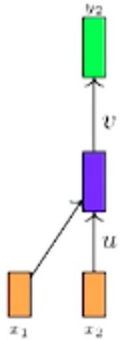
$$s_i = \sigma(Ux_i + b)$$
$$y_i = \sigma(Vs_i + c)$$
$$i = \text{timestep}$$

- Since we want the same function to be executed at each timestep we should share the same network (i.e., same parameters at each timestep)

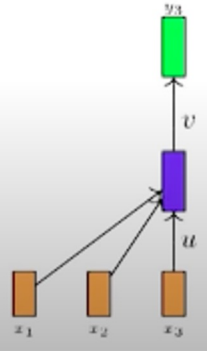


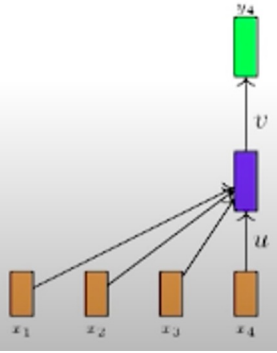
- This parameter sharing also ensures that the network becomes agnostic to the length (size) of the input
- Since we are simply going to compute the same function (with same parameters) at each timestep, the number of timesteps doesn't matter
- We just create multiple copies of the network and execute them at each timestep

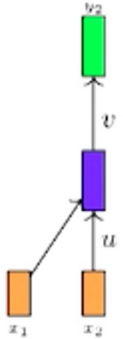


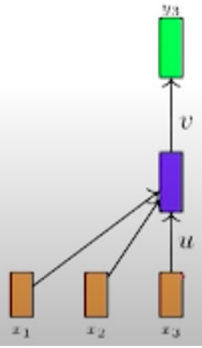


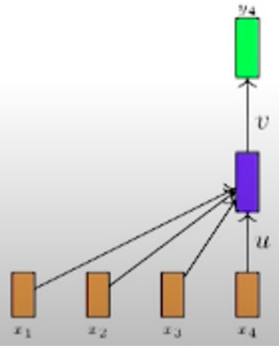
- How do we account for dependence between inputs ?
- Let us first see an infeasible way of doing this
- At each timestep we will feed all the previous inputs to the network







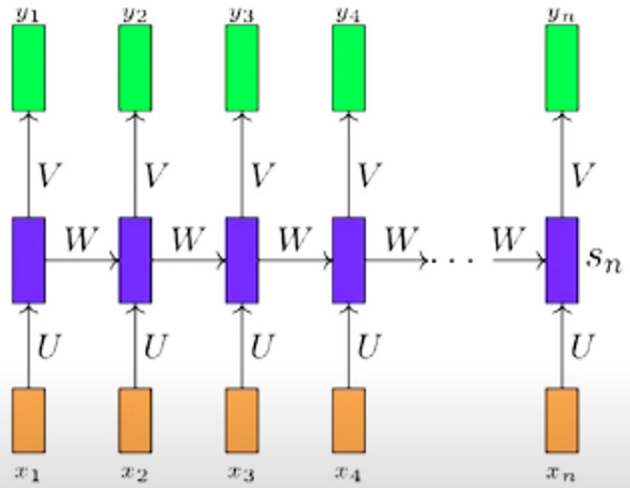




- First, the function being computed at each time-step now is different

$$y_1 = f_1(x_1)$$
$$y_2 = f_2(x_1, x_2)$$
$$y_3 = f_3(x_1, x_2, x_3)$$

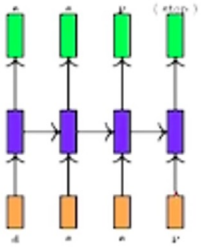
- The network is now sensitive to the length of the sequence
- For example a sequence of length 10 will require $f_1, \ldots, f_{10}$ whereas a sequence of length 100 will require $f_1, \ldots, f_{100}$

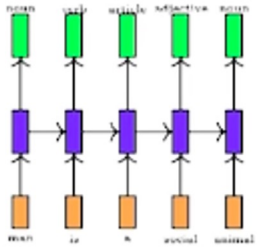


- The solution is to add a recurrent connection in the network,

$$s_i = \sigma(Ux + Ws_{i-1} + b)$$
$$y_i = \sigma(Vs_i + c)$$
$$or$$
$$y_i = f(x_i, s_i, W, U, V)$$

- s_i is the state of the network at timestep i
- The parameters are W, U, V which are shared across timesteps
- The same network (and parameters) can be used to compute $y_1, y_2, \ldots, y_{10}$ or y_{100}





- Let us revisit the sequence learning problems that we saw earlier
- We now have recurrent connections between time steps which account for dependence between inputs