**Problems with simple RNN:**

Let’s take a many-many(same length) RNN, now here we can see that output we are getting at time=4 depends more on xi4(current input), as compare to xi1

But in real world, there are requirements where the the latter result dependent a lot on earlier input, that means yi4 is more dependent on xi1 as compare to xi4, but since in simple RNN yi4 is direct connected to xi4, it will be more dependent on xi4 as compare to xi1, this is **first problem.**

**Another problem** is even we consider the small dependency of yi4 on xi1, but since there will be long chain derivatives so there will be the problem of gradient vanishing and gradient exploding.

So to overcome this LSTM was invented.













