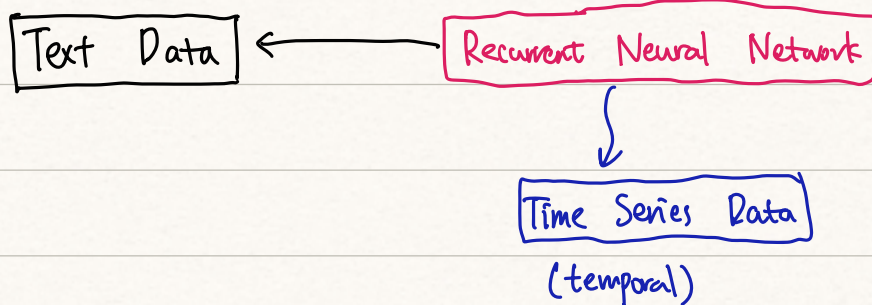
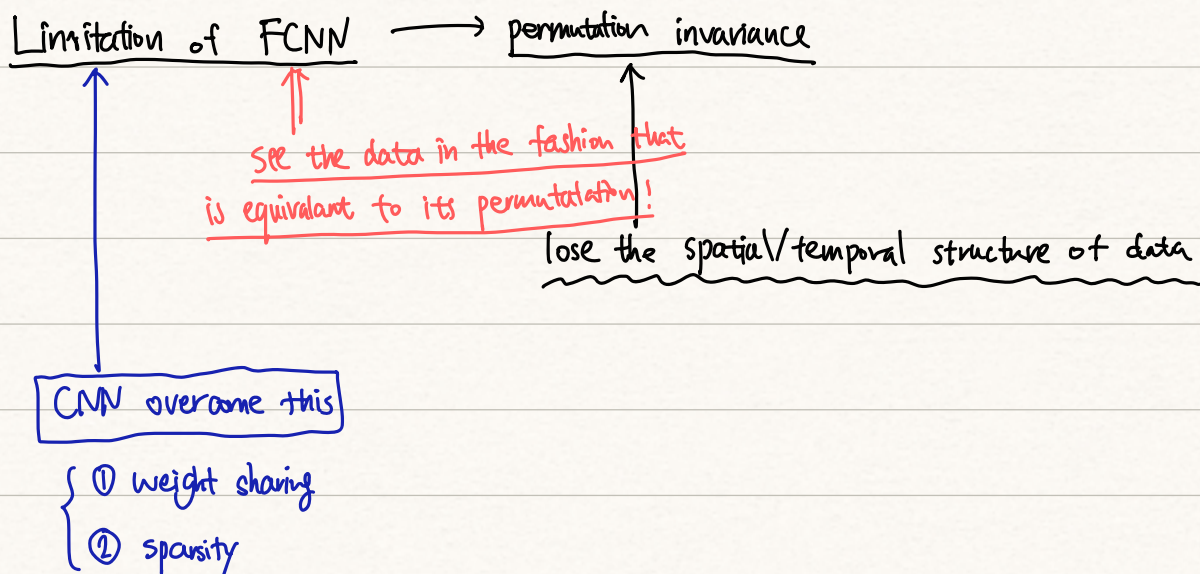


Recap:



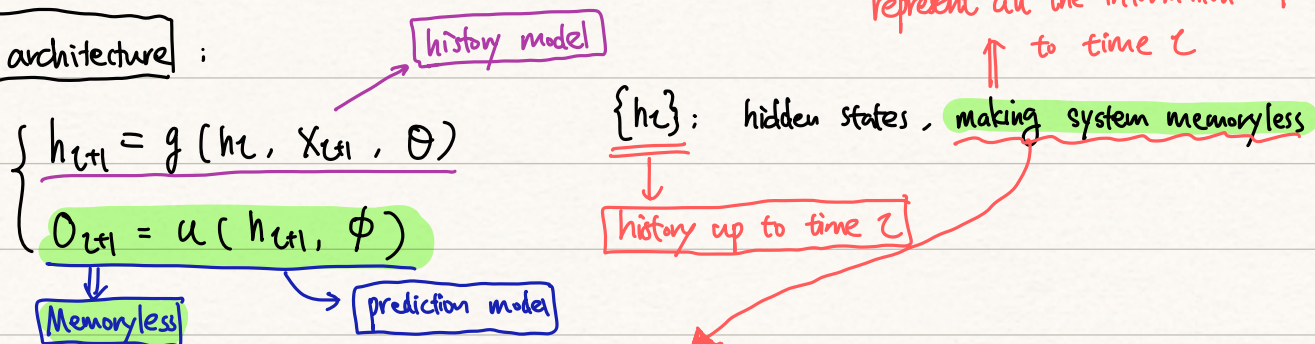
Data: (temporal)

$$\begin{cases} X = \{X(t) : t=1, 2, \dots, t_{max}\} \\ Y = \{Y(t) : t=1, 2, \dots, t_{max}\} \end{cases}$$

Our Oracle (can model the relationship between  $X$  &  $Y$ )

$Y = f^*(X)$

① RNN architecture:

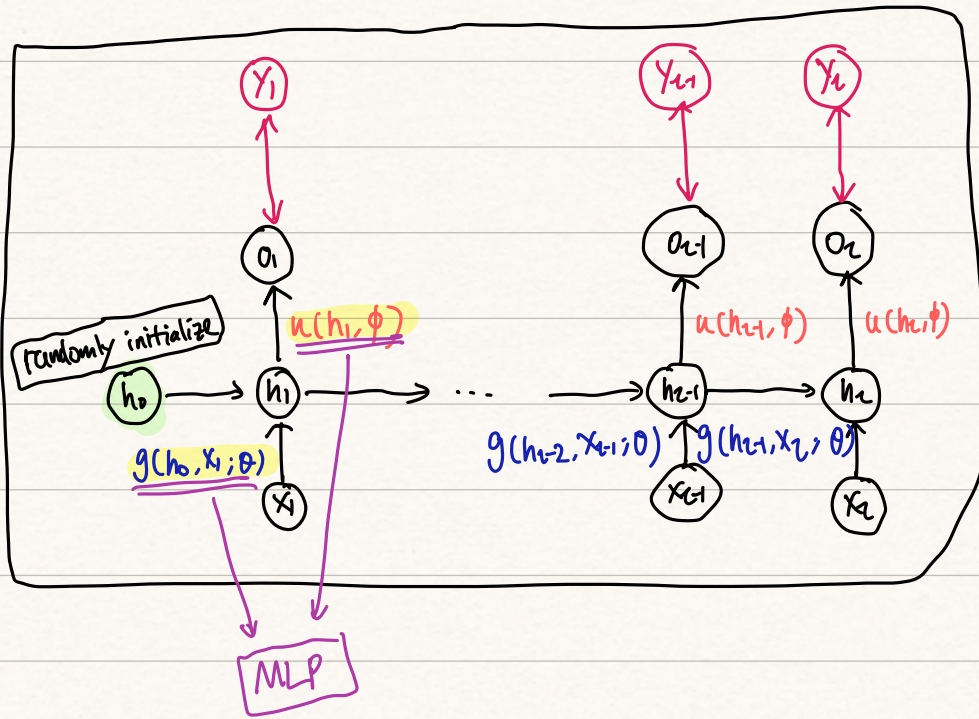


$\Rightarrow$  ARMA model:  $Y_{t+1} = \sum_{t=1}^{t+1} W_t X_t = \sum_{t=1}^t W_t X_t + \underbrace{W_{t+1} X_{t+1}}_{\text{now}}$

$\Rightarrow \begin{cases} \theta \\ \phi \end{cases}$  trainable parameters

## ② How to optimize RNN ?

Training



$$R_{\text{emp}} = \frac{1}{T} \sum_{i=1}^T \ell(o_i, y_i)$$

Diagram

can add temporal structure  
to loss function

different time step, different  
weights

## ③ Observation :

a) Parameters are shared within time for RNN,  
and parameters are shared within space

b) causal structure is preserved (temporal structure)

(Modification)

c) hidden state units can be made large to give the system  
some form of memory

modify the function  $g(\cdot)$  (complicate architecture)

{ GRU  
LSTM