

10-1 序列資料

<def> 資料前後有關聯性 (有順序的向量資料)

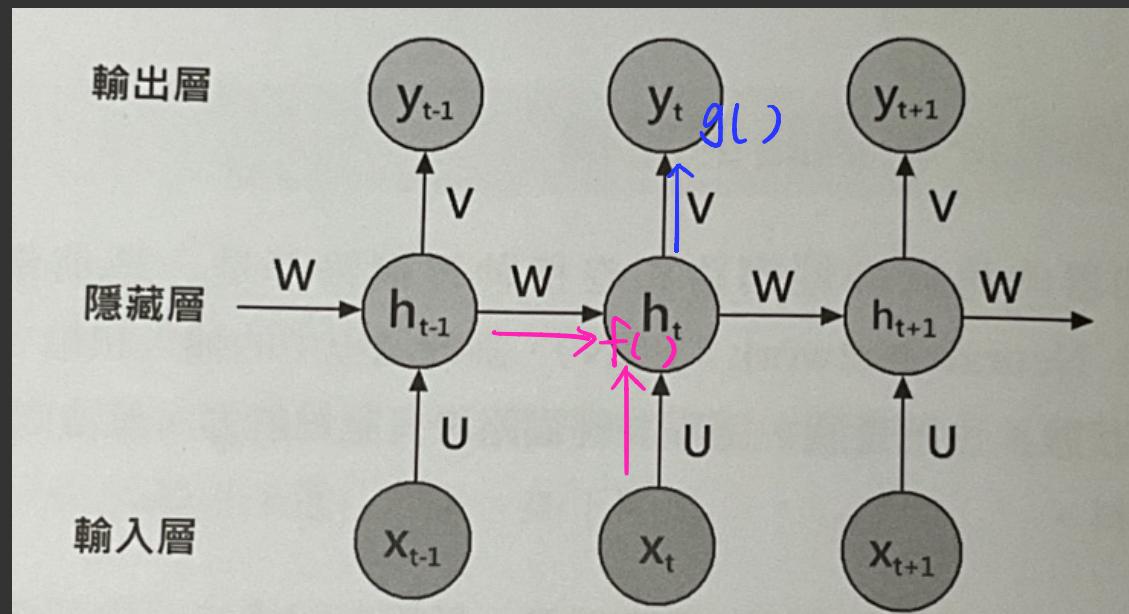
不一定是時間序列

1. 自然語言的句子 (N-gram 模型)

- 二元 (bigram) }
 - 三元 (trigram) }
- 用前 $N-1$ 個單字預測
下一個出現的字

2. 股價 (時間序列)

10-3 循環神經網路(RNN)



W = 時步權重

U = 輸入層權重

V = 輸出層權重

共享權重

使用 W 記錄歷史資料

每一個隱藏層的輸出都作為下一個隱藏層的輸入

$$h_t = f(x_t \cdot U + h_{t-1} \cdot W) \quad \begin{array}{l} \text{過去的記憶} \\ \text{隱藏層的啟動函數} \\ \text{上一個時步} \end{array}$$

$$y_t = g(h_t \cdot V) \quad \text{輸出層的啟動函數}$$

損失函數

$$E = \sum_{i=1}^t f_e(y_i - t_i)$$

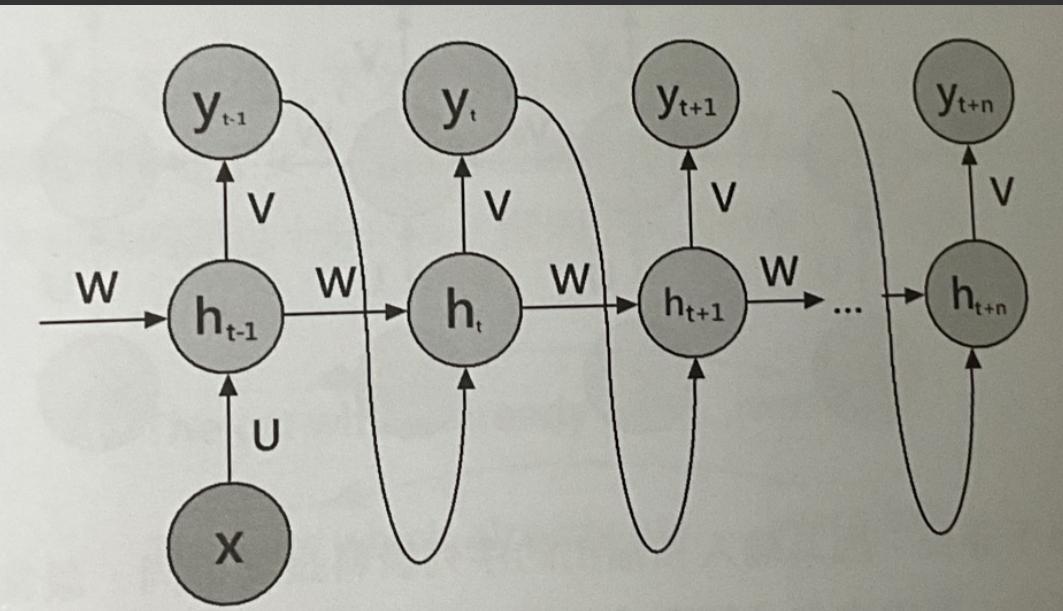
總損失



用 BPTT (透過時間的反向傳播演算法) 來更新權重

種類

- One to Many
- Many to One
- Many to Many
 - 輸入和輸出等長
 - 輸入和輸出不等長

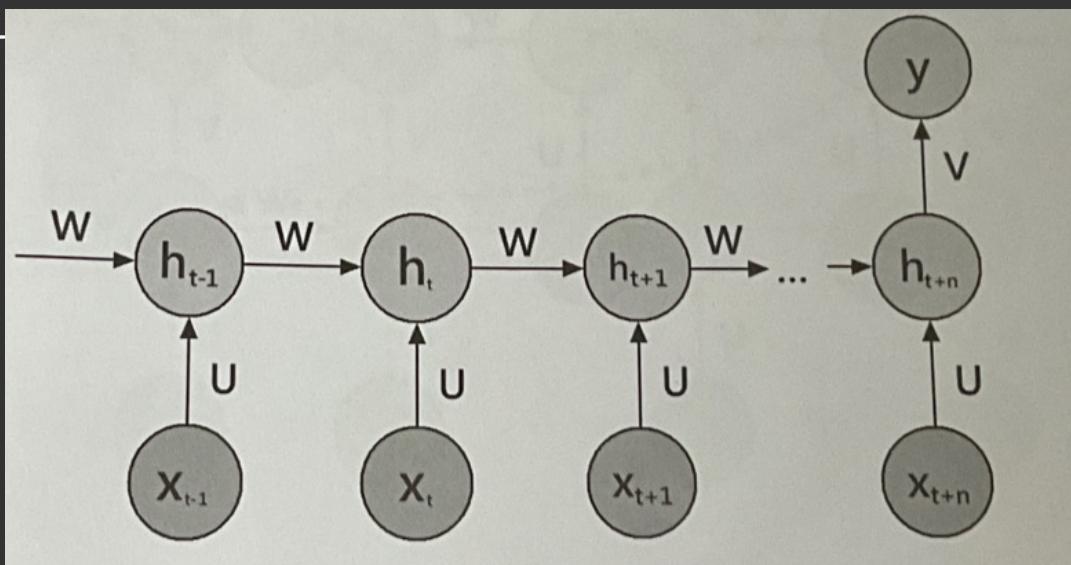


One to Many →

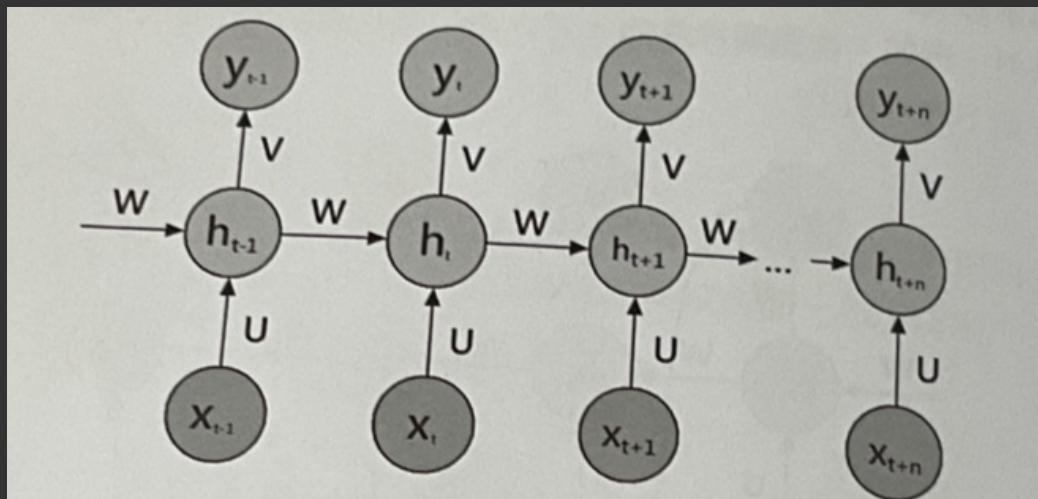
ex: 輸入一張圖片產生
文字評論

← Many to One

ex: 輸入財經新聞
輸出情緒結果(正、負面)



Many to Many

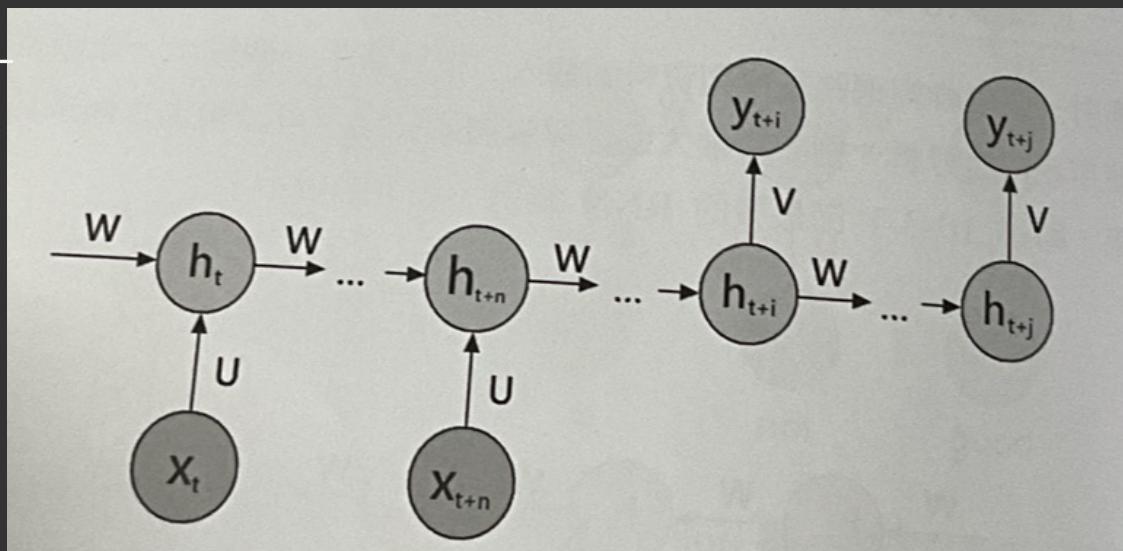


輸出和輸入等長

ex: 判斷輸入是否為人名

← 輸出和輸入不等長

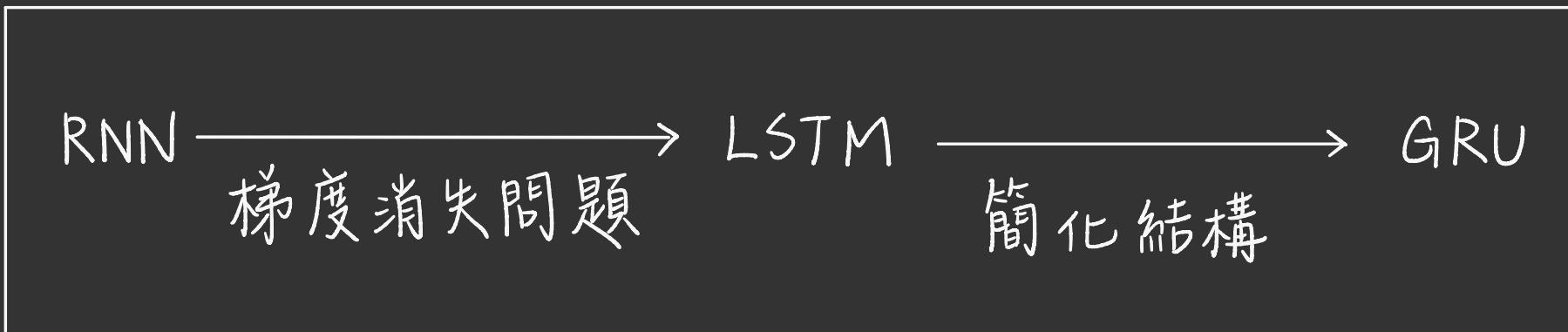
ex: 翻羽譯



梯度消失問題

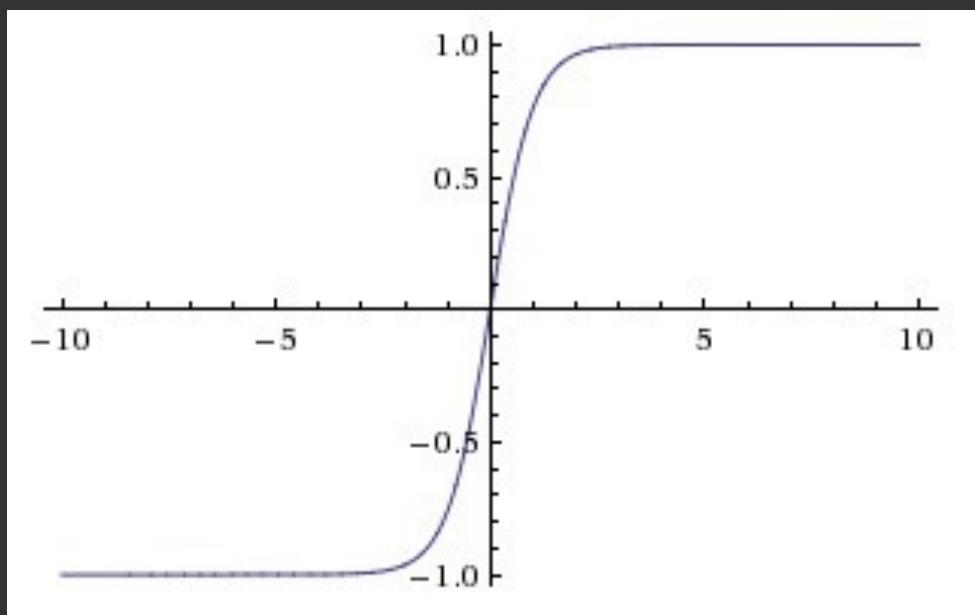
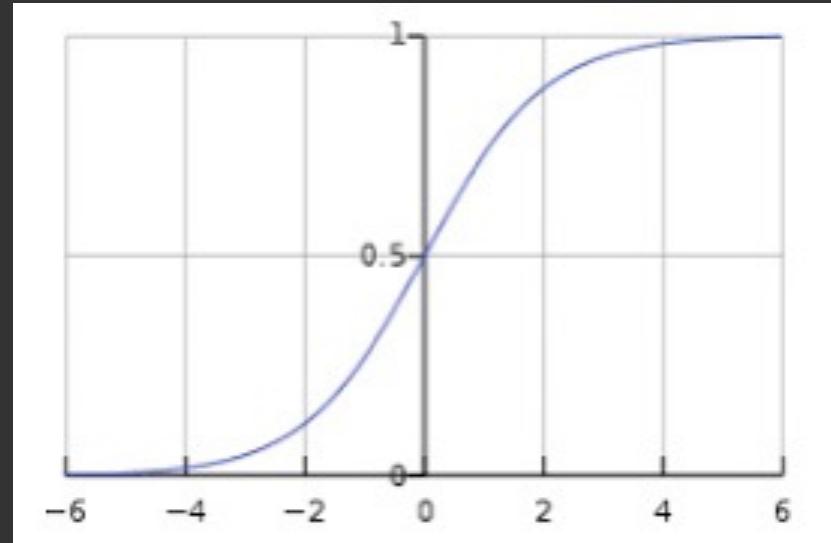
- ∴ 間隔太長時步而無法將資訊連接起來
- ∴ RNN 只擁有短期記憶

解決方法：LSTM、GRU



Sigmoid 函數 <

值在 0~1 之間

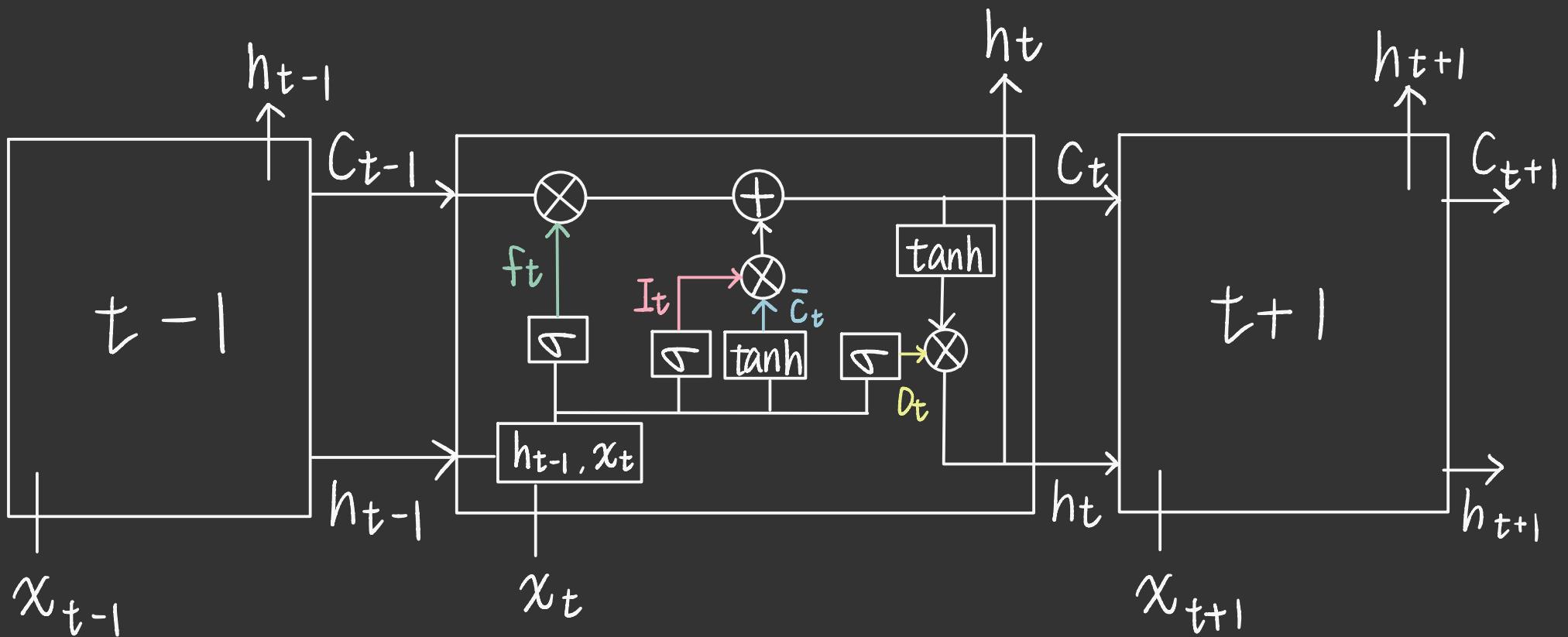


> tanh 函數

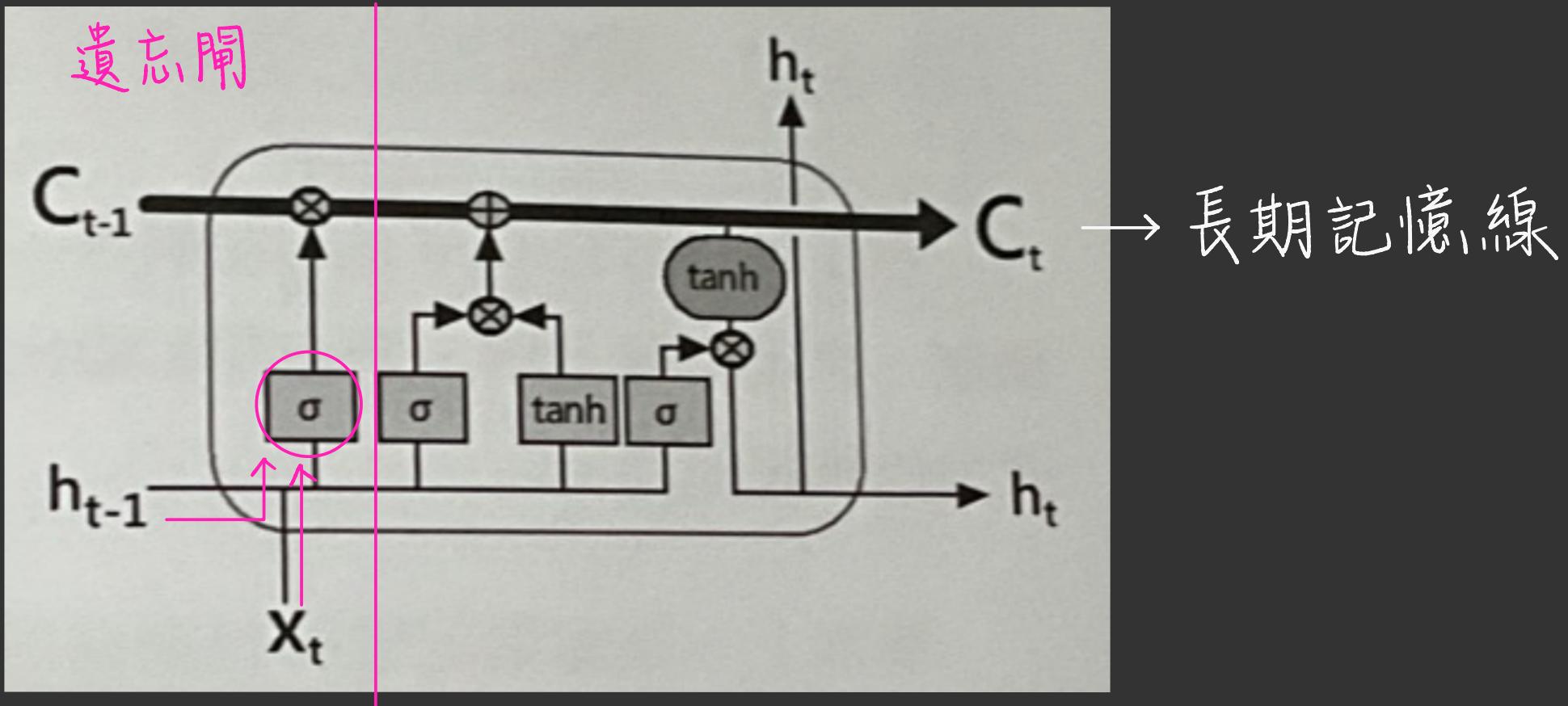
雙曲正切函數

值在 -1~1 之間

10-4 長短期記憶神經網路 LSTM



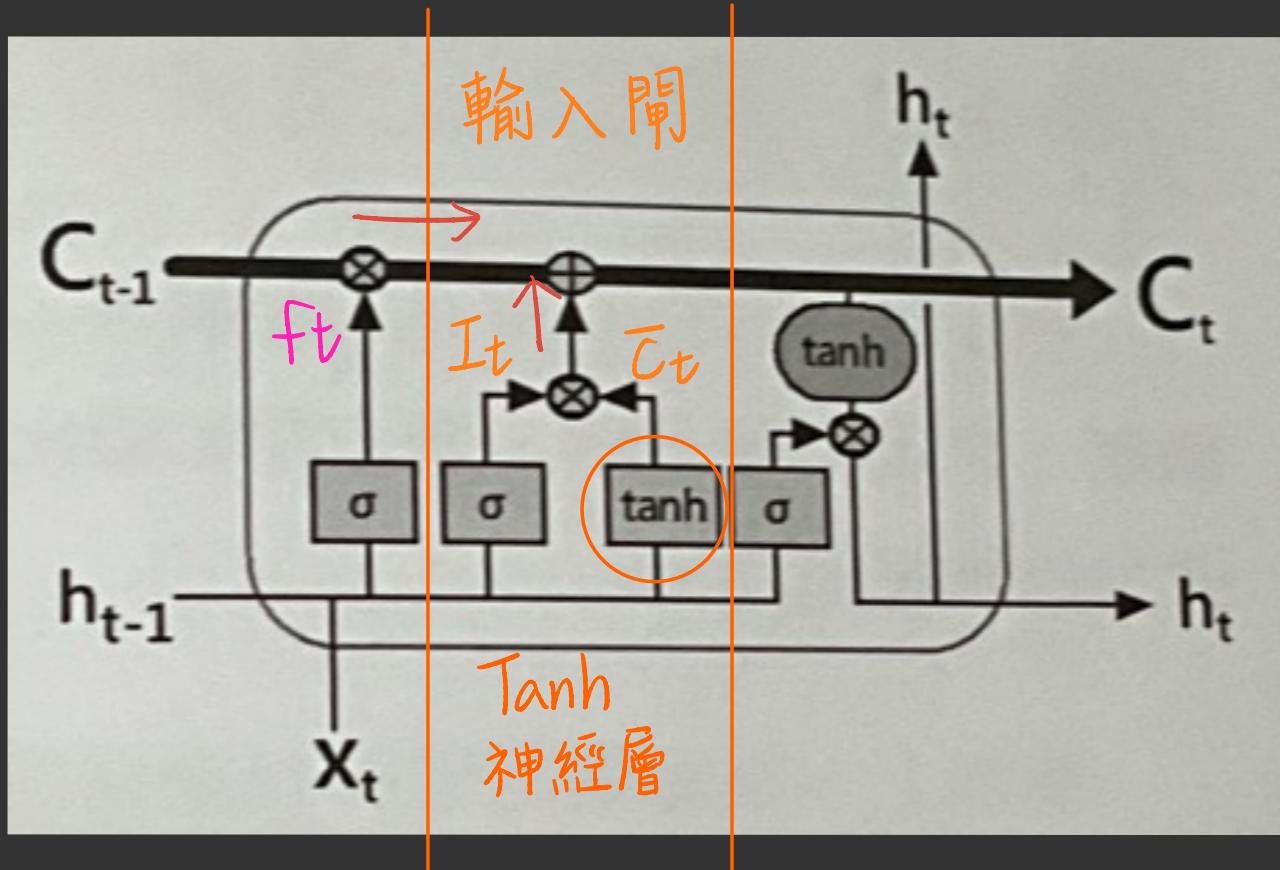
遺忘閘：從長期記憶線中刪除資料



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \Rightarrow \text{輸出 } 0 \sim 1 \text{ 的向量}$$

[] [] []
權重 輸入資料 bias

長短期記憶神經網路 (LSTM)

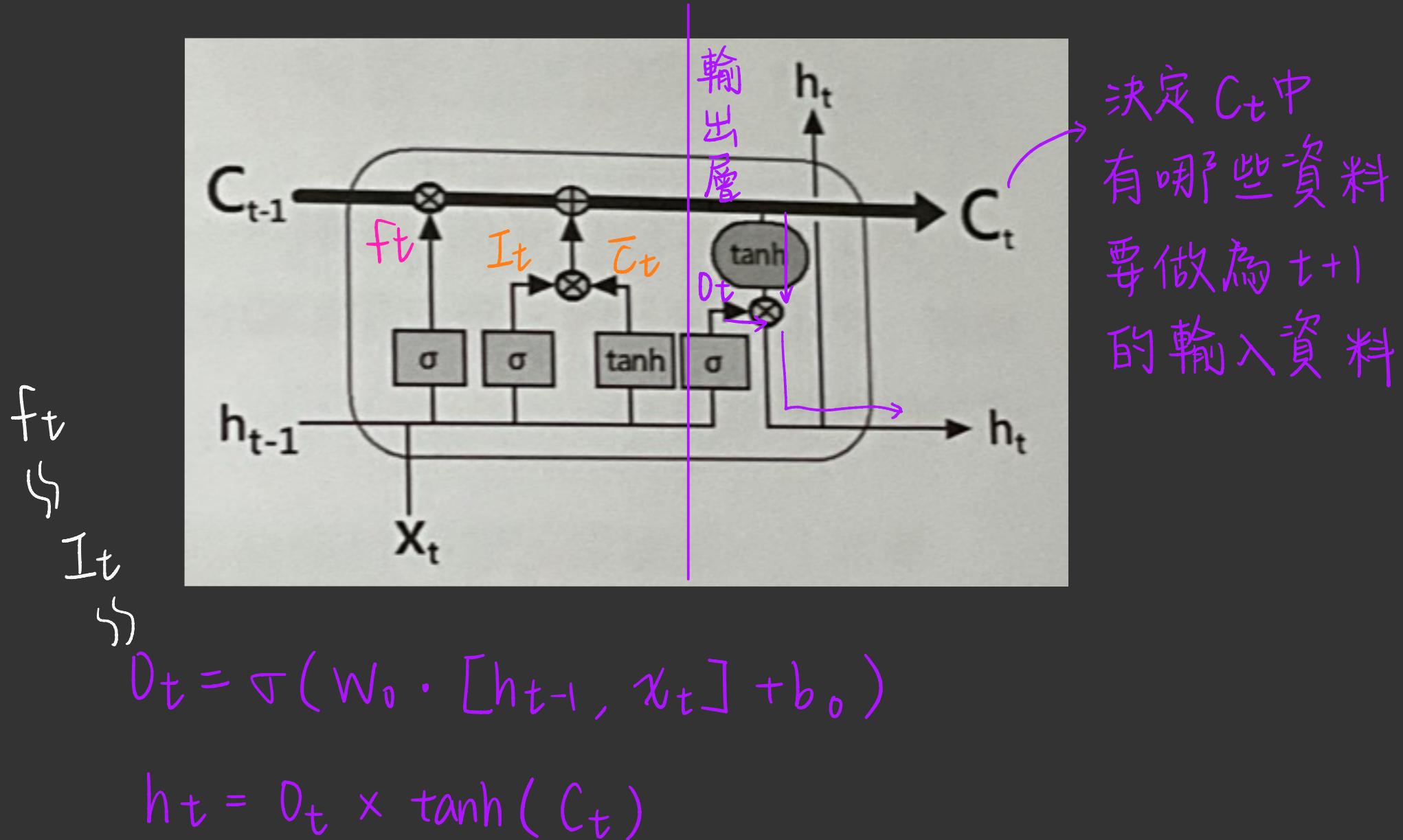


ft (?)

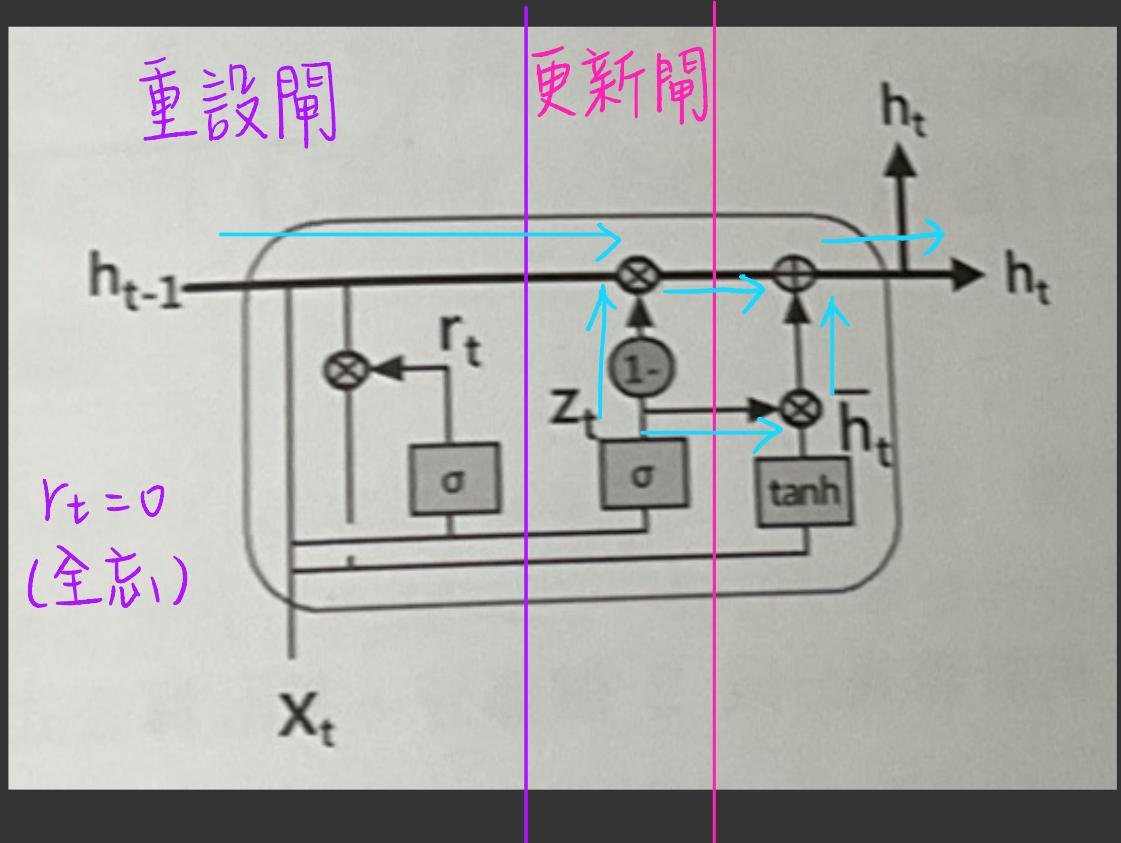
I_t : 決定需不需要更新哪些資料 = $\sigma(w_i \cdot [h_{t-1}, x_t] + b_i)$

$$C_t = f_t \times C_{t-1} + I_t \times \bar{C}_t$$

長短期記憶神經網路 (LSTM)



GRU 開門循環單元神經網路



控制資料的保留和更新

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z)$$

$$\bar{h}_t = \tanh(W_h \cdot [r_t \times h_{t-1}, x_t] + b_h)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \bar{h}_t$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r)$$

* r_t : 控制保留多少之前的資料

更新的記憶資料

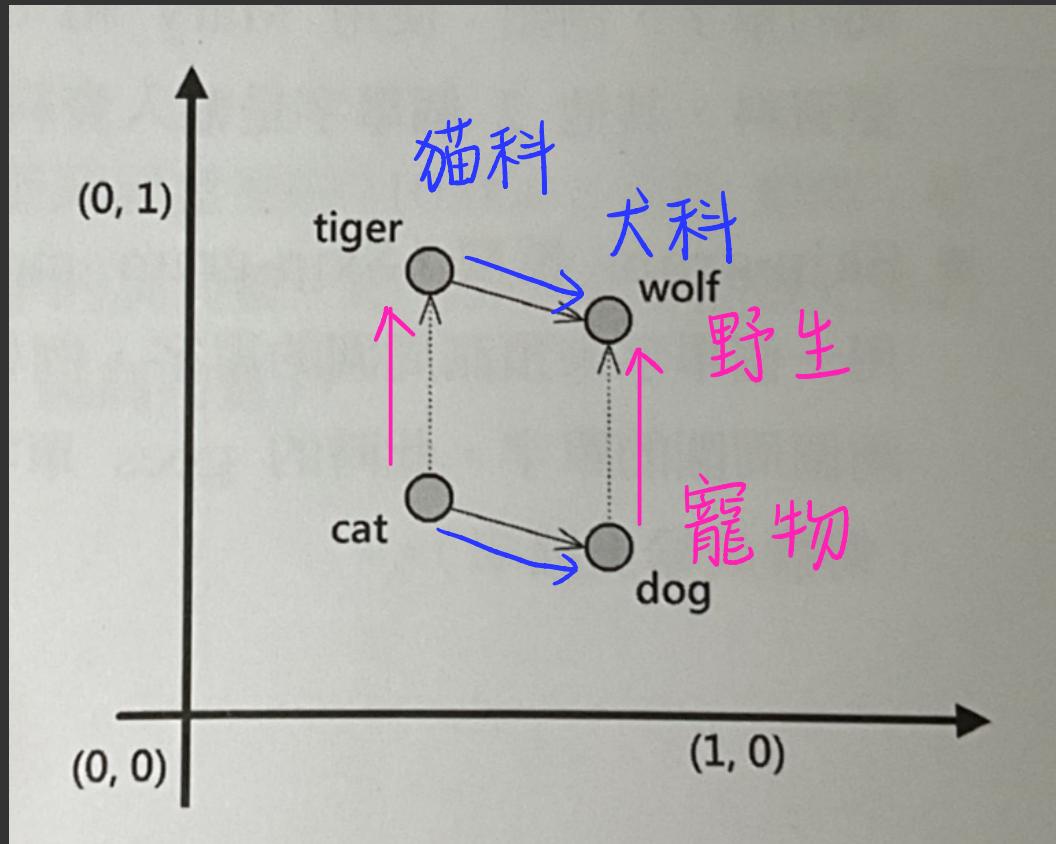
資料向量化

<def> 將資料轉成數值資料的張量

文字資料向量化：

one-hot 編碼	詞向量 (詞嵌入)
使用 code 轉換	建立神經網路自行學習
高維度向量	低維度浮點數向量

詞向量的幾何意義



* 同樣的向量有
相同的關係

單字在句子中的上下文關係

- CBOW 模型：利用上下文預測中心詞。
(輸入)
類似填空題
- Skip-gram 模型：源由 N-gram 模型，
使用中心詞預測上下文
(輸入)