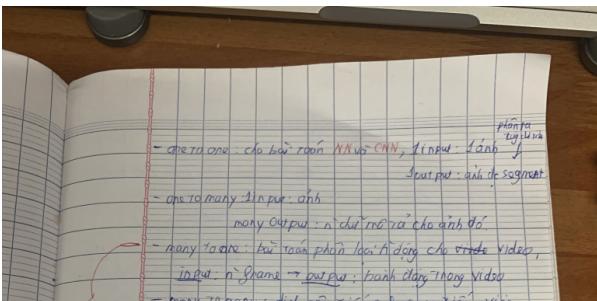
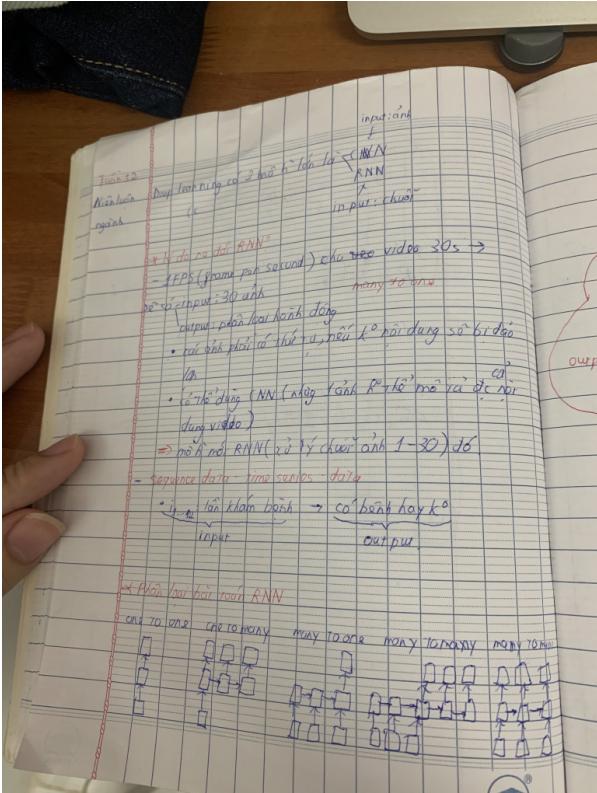


## Week 12

Sunday, March 20, 2022 2:37 PM



\*: Backpropagation Through Time (BPTT)

3 Tham số cần tìm: w, u, v

Để thực hiện gradient descent: tính đạo hàm của  $\frac{\partial L}{\partial \theta_0}$ ;  $\frac{\partial L}{\partial \theta_1}$ ,  $\dots$ ,  $\frac{\partial L}{\partial \theta_n}$

f) Dao hàm với V:

$$\frac{\partial L}{\partial V} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial V}$$

f) Vợ Uvà` M

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial Y} \times \frac{\partial Y}{\partial S_{30}} \times \frac{\partial S_{30}}{\partial W} = \frac{\partial S_{30}}{\partial W} + \frac{\partial S_{30}}{\partial S_{30}} \times \frac{\partial S_{30}}{\partial W}$$

$$Do: S_{30} = g (W \times S_{29} + V \times x_{30}) \text{ có } S_{29} \text{ là } E \text{ và } W$$

Opdung C7:  $(f(x) \circ g(x))' = f'(x) \circ g(x) + f(x) \circ g'(x)$

$$= \sum_{i=0}^{20} \frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial s_{30}} * \frac{\partial s_{30}}{\partial s_i} * \frac{\partial s_i}{\partial W}$$

$$V = \prod_{j=1}^{29} \frac{\partial s_{j+1}}{\partial s_j} \text{ và } \frac{\partial s_i^2}{\partial W} \text{ là đạo hàm của } s_i \text{ với } W$$

khi  $s_{i-1}$  là constant với  $W$ .

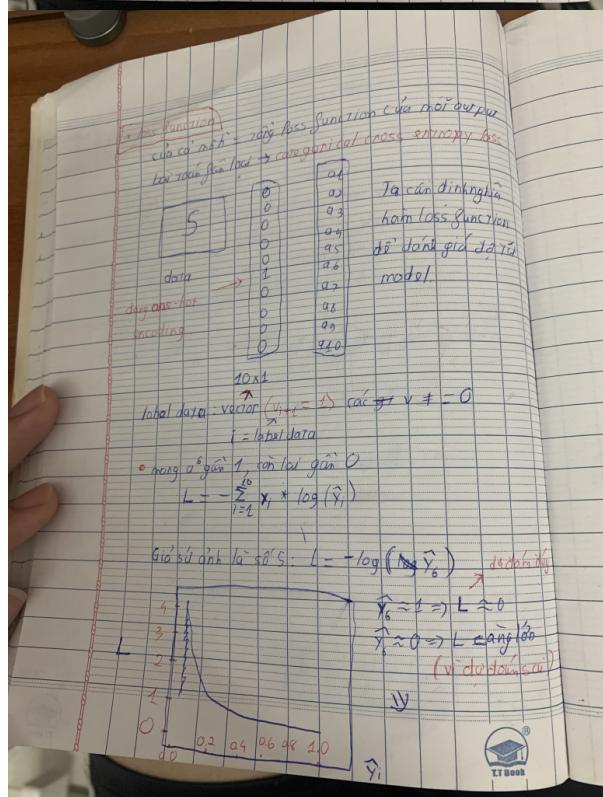
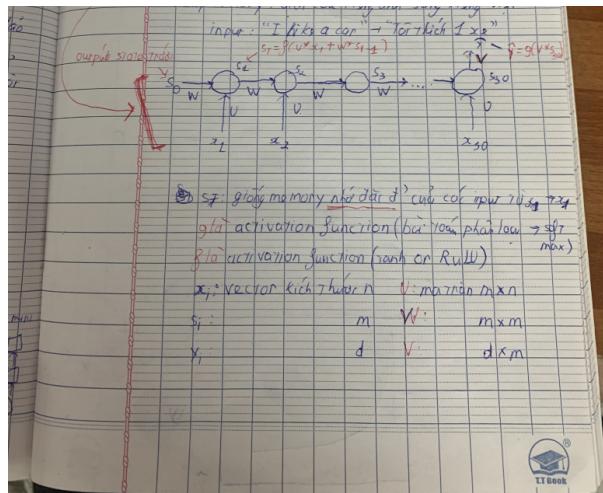
\*: Nhìn vào công thức đạo hàm của W và L

Ta thấy hiện tượng vanishing gradient ở các

State  $\hat{d}_k^{\text{opt}} \hat{x}_{ik} \Rightarrow \hat{a}_k^{\text{opt}} \hat{x}_{ik}^{\text{opt}} = \hat{m}_k^{\text{opt}} \neq \hat{z}_{ik}^{\text{opt}}$  from def.

grain hiến-tưởng nay

→ Khi  $W$  nhỏ hơn 1 thì  $\hat{W}$  tính cóc gradient ở cóc layer đầu tiên sẽ phai nhau tích của rất ngsố bé  $< 1$  nên gradient  $\rightarrow 0$ ,  $\Rightarrow$  bước giảm hết kim số trong gradient descent rõ ràng.



$\Rightarrow$  Bài toán trả Thành bài toán Tìm min L

vai các hì sót nong NN se k'ò doc  
đè nèig

$$\frac{\partial S_{30}}{\partial s_i} = W^{30-i} \times \prod_{j=1}^{29} (1 - s_j^2)$$

$S_i < 1, W < 1 \Rightarrow$  static  $\hat{d} x q: \frac{\partial S_{i0}}{\partial S_i} \simeq 0 \Leftrightarrow \frac{\partial L}{\partial W} \simeq 0 \Rightarrow$  vanishing gradient

## Gradient descent :

giáo sư đồng ý định núi và ban cấm xuống thuyền lung cách nhau  
- Thuyền lung ở đây là n' đ'c d'c tiêu  
chỗ nào cản nhau tại chỗ đó không được nhau thi ra bờ xuống  
Tuy bờ 1 chỗ đến khi cản nhau xung quanh k' có cản đỡ nhau (bằng phẳng)  
→ thuyền lung

# LSTM

long short term memory

- mô hình RNN nào:

$$\text{Đoạn hàm } L \text{ và } W \text{ ở state } i: \frac{\partial L}{\partial W} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial s_0} \times \frac{\partial s_0}{\partial s_i} \times \frac{\partial s_i}{\partial W}$$

$$= \prod_{j=i}^{29} \frac{\partial s_{i+1}}{\partial s_j}$$

- Giả sử activation function: tan

$$\Rightarrow s_t = \tan(V * x_t + W * s_{t-1})$$

$$\frac{\partial s_t}{\partial s_{t-1}} = (1 - s_t^2) \neq W \Rightarrow \boxed{\frac{\partial s_0}{\partial s_i} = W^{30-i} \times \prod_{j=1}^{29} (1 - s_j^2)}$$

$$s_i < 1, W < 1 \Rightarrow \text{nhiều state } \partial x \text{ sao: } \frac{\partial s_0}{\partial s_i} \approx 0 \Leftrightarrow \frac{\partial L}{\partial W} \approx 0 \Rightarrow \text{vanishing gradient}$$