ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

BÀI TẬP
MÔN CÁC KỸ THUẬT HỌC SÂU VÀ ỨNG DỤNG

KHOA: KHOA HỌC MÁY TÍNH

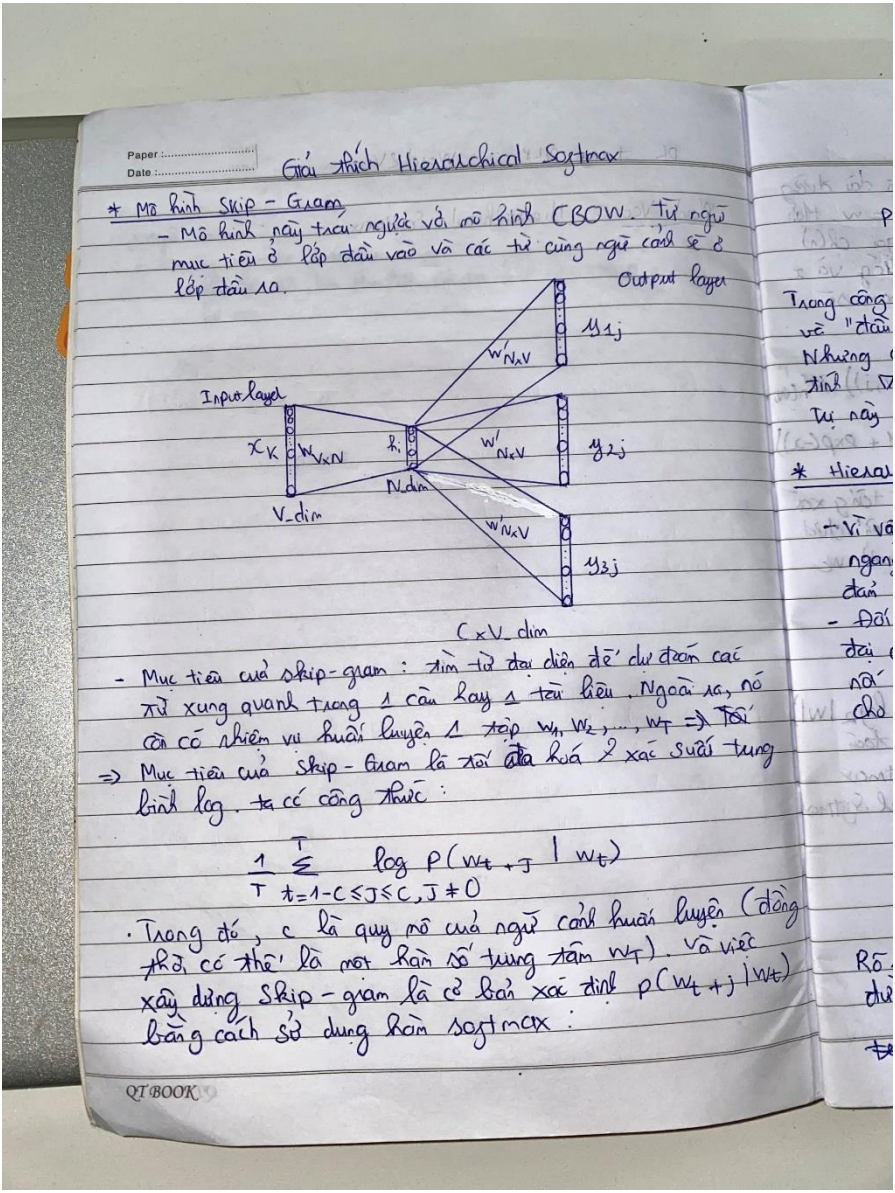
HOMEWORK: **GIẢI THÍCH HÀM HIERALCHICAL SOFTMAX**

Giảng viên: Nguyễn Duy Khánh

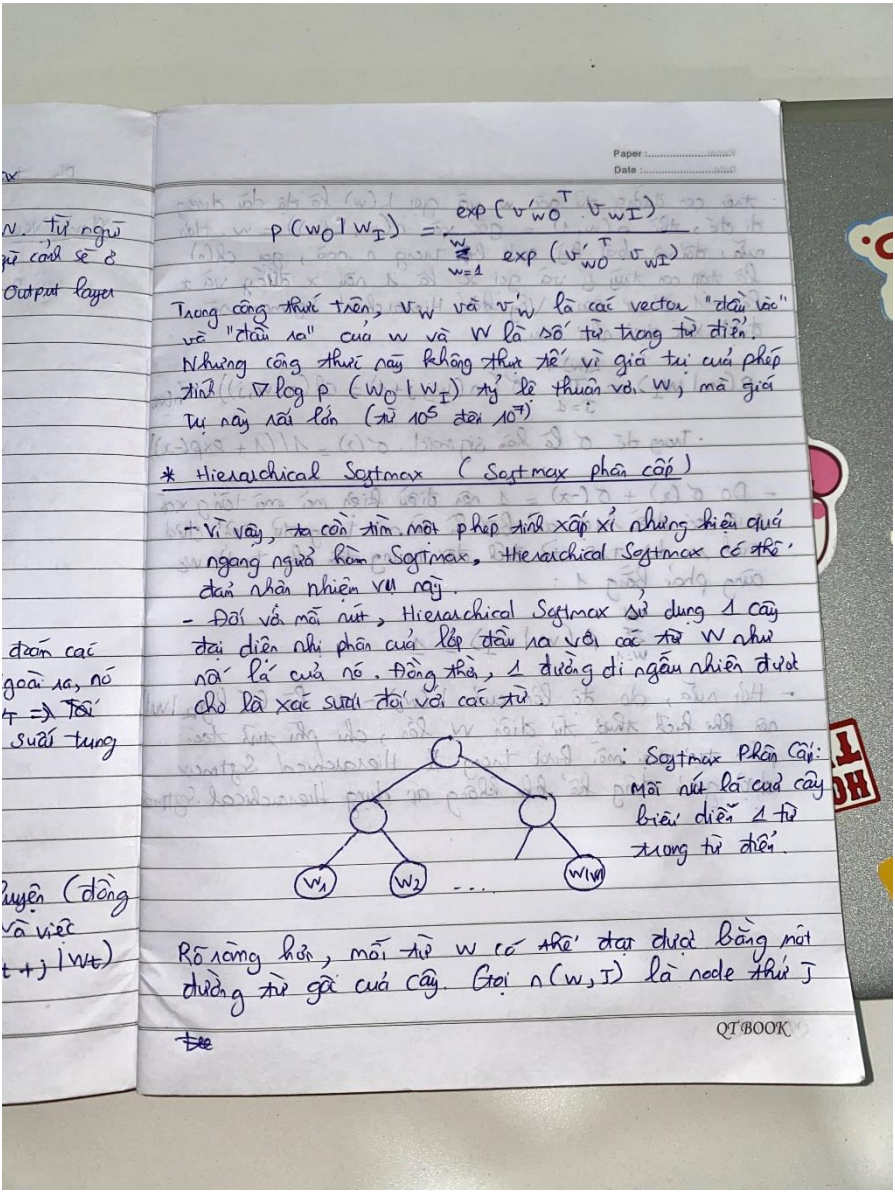
Nhóm thực hiện:

1. Trương Văn Khải– 21520274

Hình 1


Giải thích Hierarchical Softmax
* Mô hình Skip - Gram
- Mô hình này trái ngược với mô hình CBOW. Từ ngữ mục tiêu ở lớp đầu vào và các từ cùng ngữ cảnh sẽ ở lớp đầu ra.
Output layer
Input layer
y_{1j}
y_{2j}
y_{3j}
$W'_{N\times V}$
$W'_{N\times V}$
$W'_{N\times V}$
x_k
$W_{V\times N}$
h_i
N-dim
V-dim
C x V_dim
- Mục tiêu của skip-gram: tìm từ đại diện để dự đoán các từ xung quanh trong 1 câu hay 1 tài liệu. Ngoài ra, nó còn có nhiệm vụ huấn luyện 1 tập $w_1, w_2, \ldots, w_T$ =>
=> Mục tiêu của skip-gram là tối đa hoá ? xác suất tương ứng log. có công thức:
$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\le j\le c, j\ne 0} \log p(w_{t+j} \mid w_t)$$
. Trong đó, c là quy mô của ngữ cảnh huấn luyện (đồng thời có thể là một hàm số tương tâm w_t). Và việc xây dựng Skip-gram là cơ bản xác định $p(w_{t+j}|w_t)$ bằng cách sử dụng hàm softmax:
QT BOOK

Hình 2



$$p(w_O \mid w_I) = \frac{\exp(v'^{T}_{w_O} v_{w_I})}{\sum_{w=1}^{W} \exp(v'^{T}_{w_O} v_{w_I})}$$

Trong công thức trên, v_w và v'_w là các vector "đầu vào" và "đầu ra" của w và W là số từ trong từ điển. Nhưng công thức này không thực tế vì giá trị của phép tính $\nabla \log p(w_O \mid w_I)$ tỷ lệ thuận với W, mà giá trị này rất lớn (từ 10^5 đến 10^7)

* Hierarchical Softmax (Softmax phân cấp)

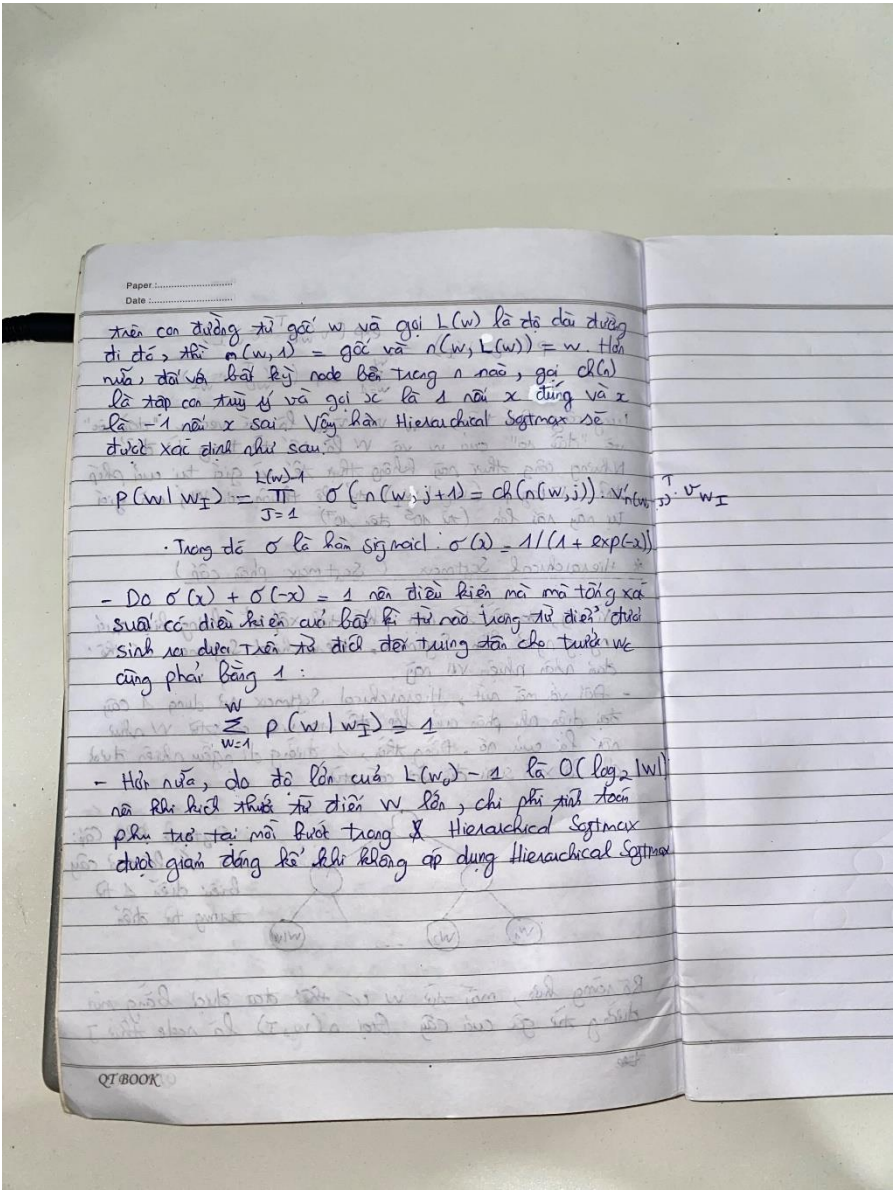
+ Vì vậy, ~~ta~~ còn tìm một phép tính xấp xỉ nhưng hiệu quả ngang ngửa hàm Softmax, Hierarchical Softmax có thể đảm nhận nhiệm vụ này.
- Đối với mỗi nút, Hierarchical Softmax sử dụng 1 cây đại diện nhị phân của lớp đầu ra với các từ W như là lá của nó. Đồng thời, 1 đường đi ngẫu nhiên được cho là xác suất đối với các từ.

Softmax Phân Cấp: Mỗi nút lá của cây biểu diễn 1 từ trong từ điển.

Rõ ràng hơn, mỗi từ w có thể đạt được bằng một đường từ gốc của cây. Gọi $n(w, I)$ là node thứ J

QT BOOK

Hình 3



Paper :
Date :

trên con đường từ gốc w và gọi L(w) là độ dài đường đi đó, thì n(w,1) = gốc và n(w, L(w)) = w. Hơn nữa, đối với bất kỳ node bên trong n nào, gọi ch(n) là tập con tuỳ ý và gọi x là 1 nếu x đúng và x là -1 nếu x sai. Vậy hàm Hierarchical Softmax sẽ được xác định như sau:

$$p(w \mid w_I) = \prod_{j=1}^{L(w)-1} \sigma\left(n(w,j+1) = ch(n(w,j))\right) \cdot v'^{T}_{n(w,j)} \cdot v_{w_I}$$

· Trong đó σ là hàm sigmoid: $\sigma(x) = 1/(1+\exp(-x))$

- Do $\sigma(x) + \sigma(-x) = 1$ nên điều kiện mà mô hình xác suất có điều kiện cuối bất kì từ nào trong từ điển được sinh ra dưới toàn bộ từ điển đều tương tác cho trước w cũng phải bằng 1:

$$\sum_{w=1}^{W} p(w \mid w_I) = 1$$

- Hơn nữa, do độ lớn của $L(w_0) - 1$ là $O(\log_2 |W|)$ nên khi được thước từ điển W lớn, chi phí tính toán phụ trợ tại mỗi bước trong X Hierarchical Softmax được giảm đáng kể khi không áp dụng Hierarchical Softmax

QT BOOK