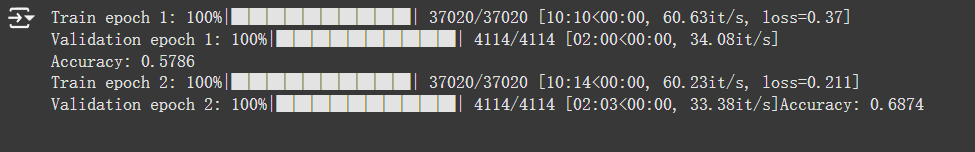
Assignment 2: Arithmetic as a language

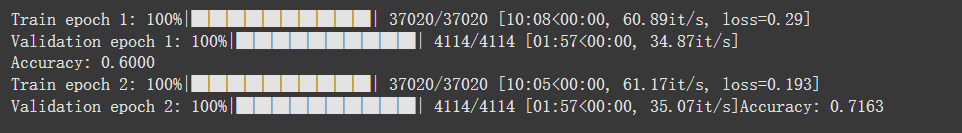
NYCU 313581016 丁驥

●What impact does using different learning rates have on model training?

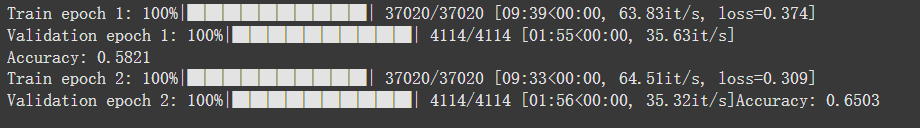
Learning rate：0.005



Learning rate：0.001

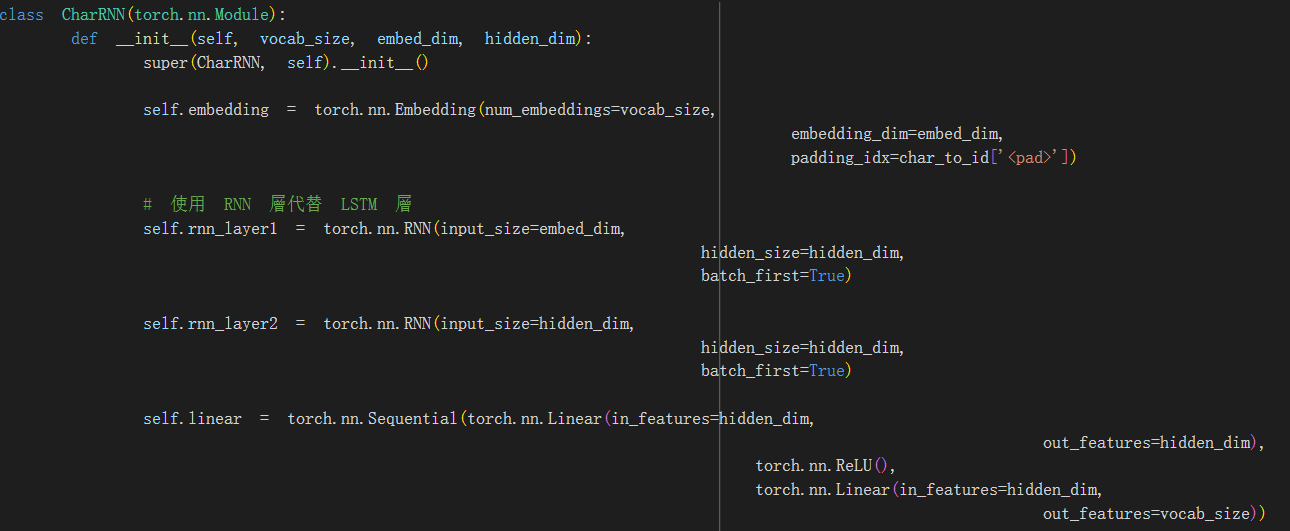


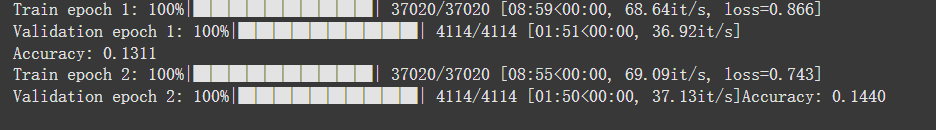
Learning rate：0.0001

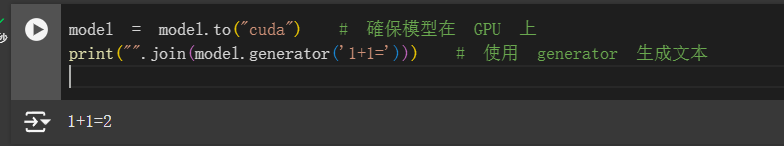


說明：我們可以發現當Learning rate 越低不一定可以收斂到越好的結果，一般來說前期learning rate 可以設大一點，再慢慢地減低，讓位置收斂，但在這個比較簡單的情況下，可以維持原本預設的 learning rate 就有不錯得結果，我想可以藉由我說明前期先維持原本的0.001 後期再改成0.0001，應該會有不錯的效果。

● If you use RNN or GRU instead of LSTM, what will happen to the quality of your answer generation? Why?

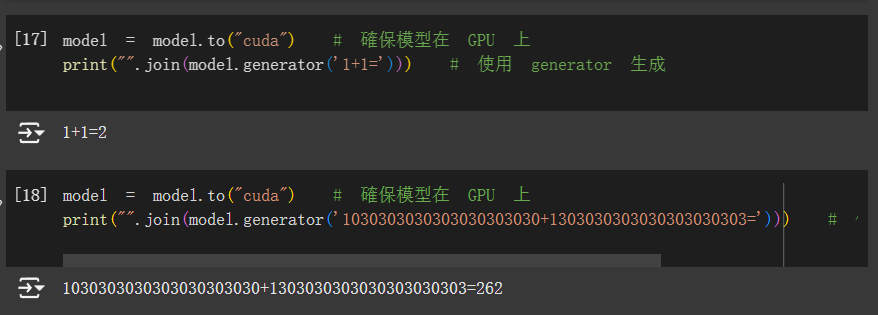






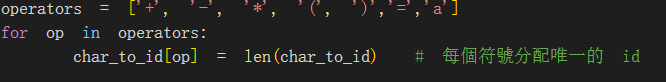
說明：可以看到RNN 在處理長序列時可能會遇到梯度消失或梯度爆炸的問題，這使得它表現不佳，因此在生成更長的答案時，可能無法保持良好的上下文理解，Accuracy只有0.14 左右，在相同 learning rate 條件下，LSTM 的Accuracy有 0.7 左右，差距相當大，就生成的效果來看(1+1=)，其效果跟LSTM輸出的一樣，但我覺得比較可能是因為問題比較簡單。

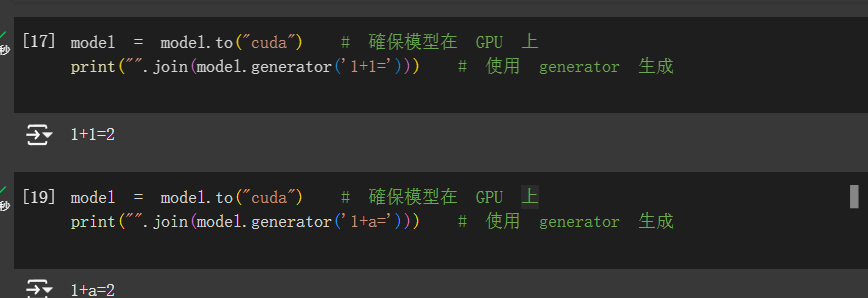
● If we construct an evaluation set using three-digit numbers while the training set is constructed from two-digit numbers, what will happen to the quality of your answer generation?

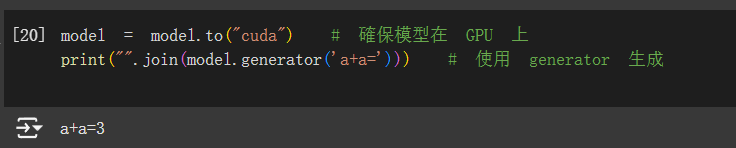


說明：我們嘗試計算一個大數計算來模擬，在 train 的過程中沒有出現過的大數，可以看到生成的結果是，只有三位數，可以推測應該是因為訓練資料位數通常不多，導致訓練的模型朝少位數 overfitting 。

● If some numbers never appear in your training data, what will happen to your answer generation?







說明：首先要先確保char\_to\_id可以轉換，我們這裡加入 “a” 假裝是 16進位的運算，可以確認 a 在訓練過程中完全沒有出現過，可以看到1+a生成的結果輸出2，a+a 的輸出結果是3，雖然不是一個有意義的輸出，但至少不會是輸出一長串內容。

● Why do we need gradient clipping during training?

gradient clipping的作用是避免模型訓練時出現「梯度爆炸」，讓訓練過程更穩定。在深層神經網絡或處理長序列時，有時梯度會變得非常大，導致訓練發散或數值錯誤。為了防止這種情況，gradient clipping會在每次update時，將梯度限制在合理範圍內，這樣可以防止過大的梯度讓模型訓練失控。

執行環境

|  |  |
| --- | --- |
| Running environment | Colab |
| Python version | Colab |

* 本作業利用 ChatGPT 進行程式碼格式調整，並加上註解