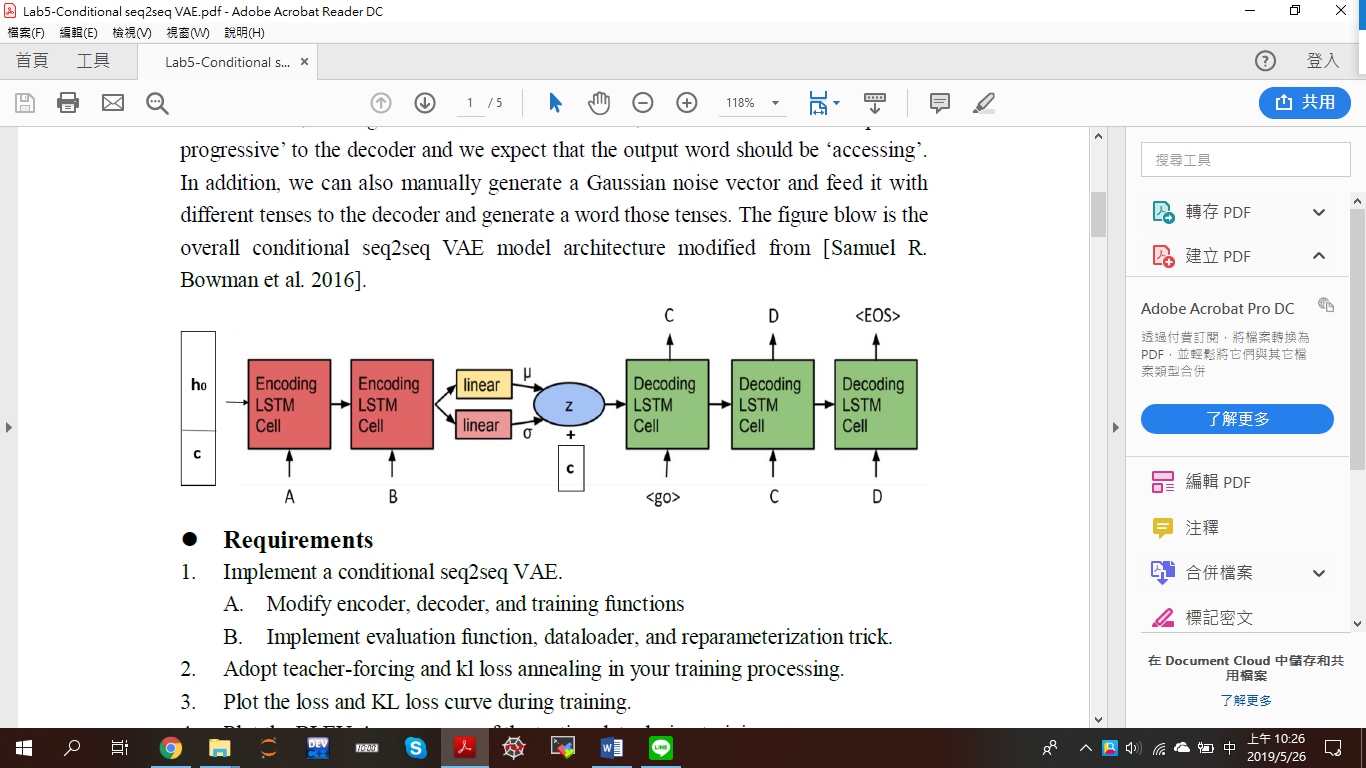
**Lab 5: Conditional Sequence-to-sequence VAE**

陽明大學 不分系二年級 張凱博

1. Introduction

這次lab主要的內容是著重於使用seq2seq的CVAE model去做英文單字的時態轉換，只要輸入一個英文單字和時態的限制，就可以讓那個英文單字輸出成為該時態的單字，主要架構如下:

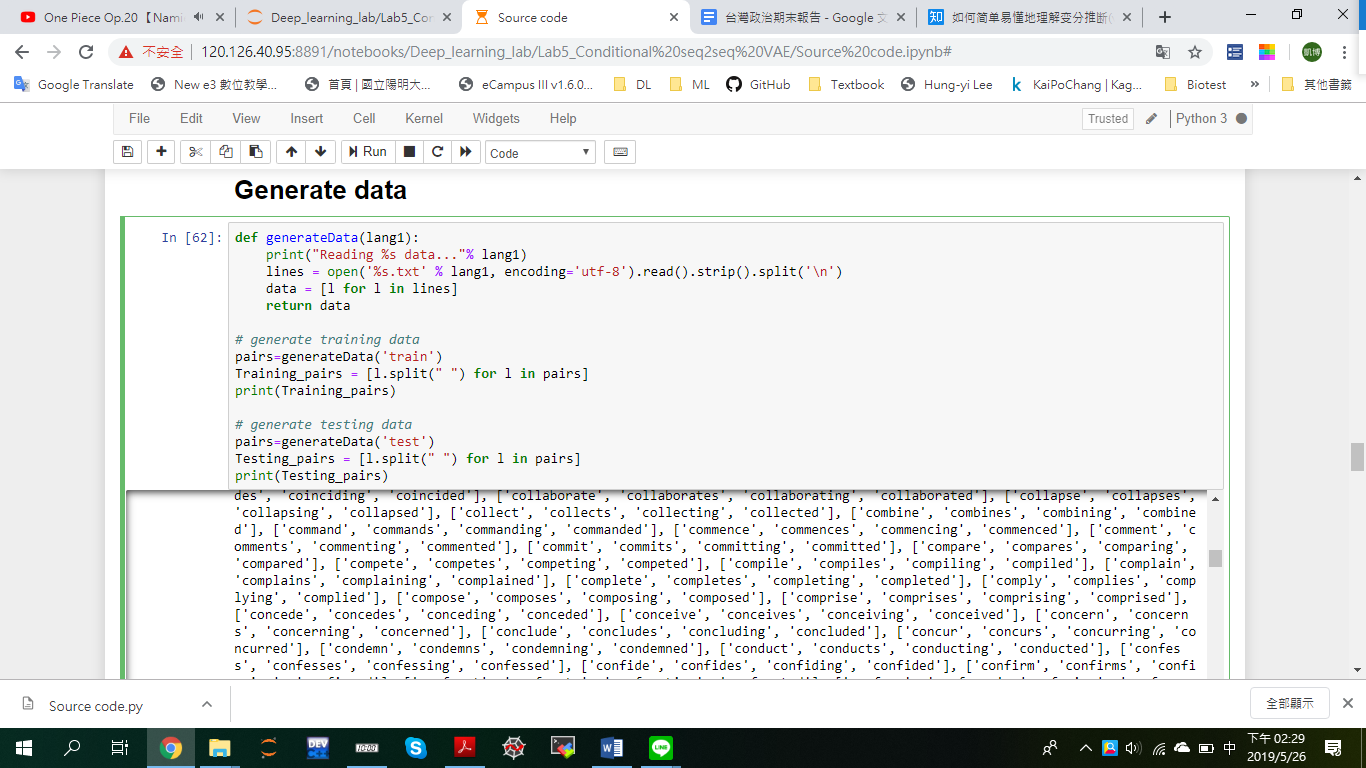


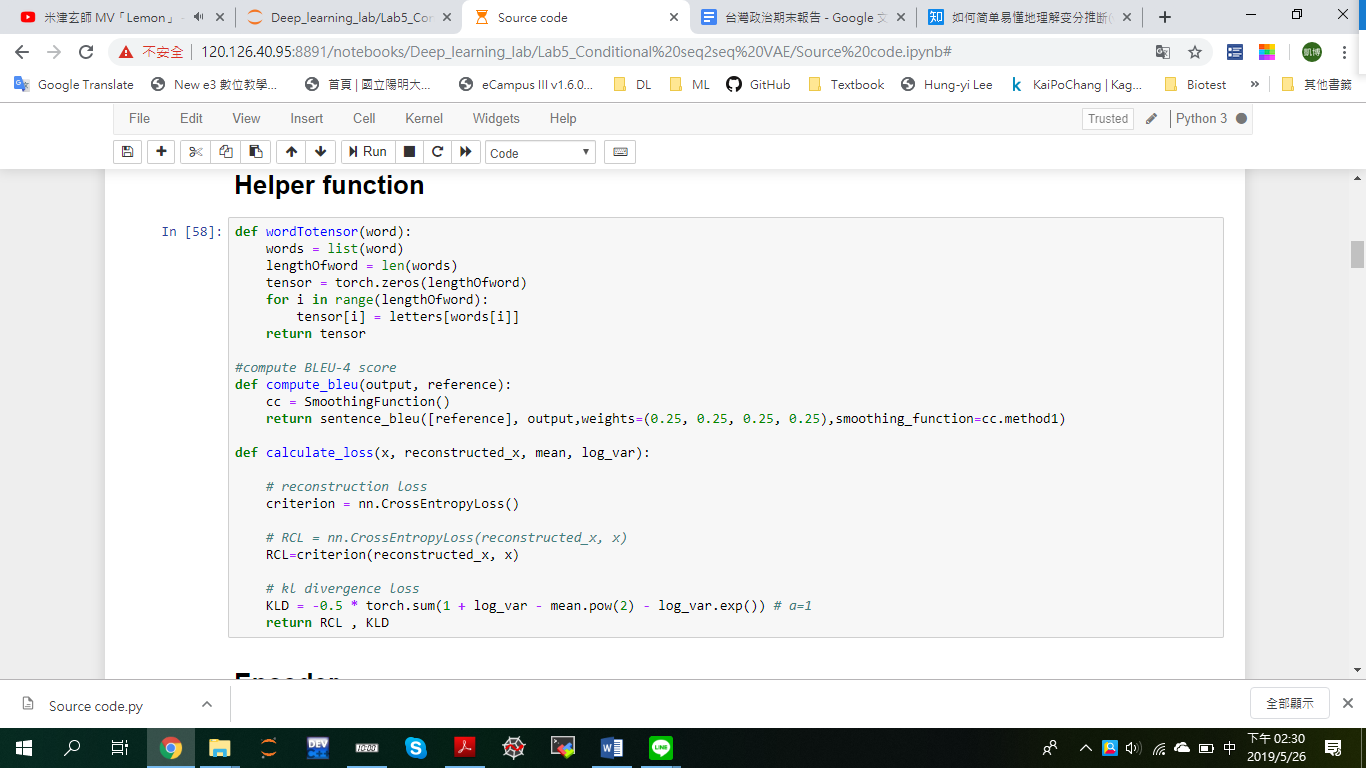
Seq2seq的架構是由一層encoder和另外一層decoder組合起來的，將經過split後的英文單字經過embedding後，將字元一個一個放進GRU之中，把encoder的最後一層輸出計算reparameterization trick以連接decoder，再來進入decoder，其中我們用teacher forcing控制每次在進入gru時的字元是不是輸入正確的字元或預測的字元，最後將每層的output收集起來轉換成相對應的字元即可。

CVAE最大的特點就是在encoder的input和decoder的input的後方，concatenate一個限制條件，這個特點也是VAE不存在的，所以我們可以藉由條件限制生成我們像要的目標，使得目標物的生成精確度上升。

1. Implementation details
2. Describe how to implement model
3. Dataloader

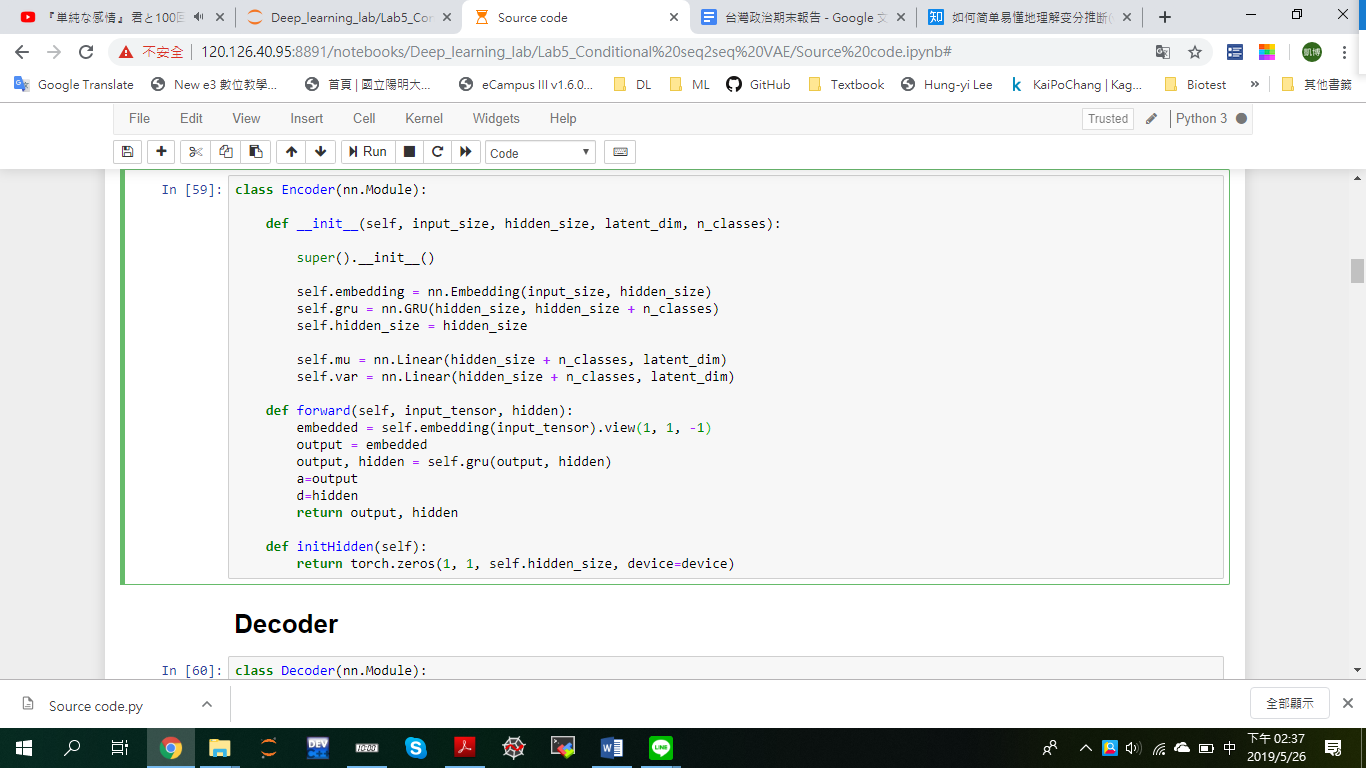
利用generateData read training pair和testing pair，把data一行一行讀出來存在list裡面，再利用word2tensor的這個function將word轉為以字元為單位的index tensor，完成dataloader，若要implement，則將一個字一個字的送到CVAE的forward裡面即可。





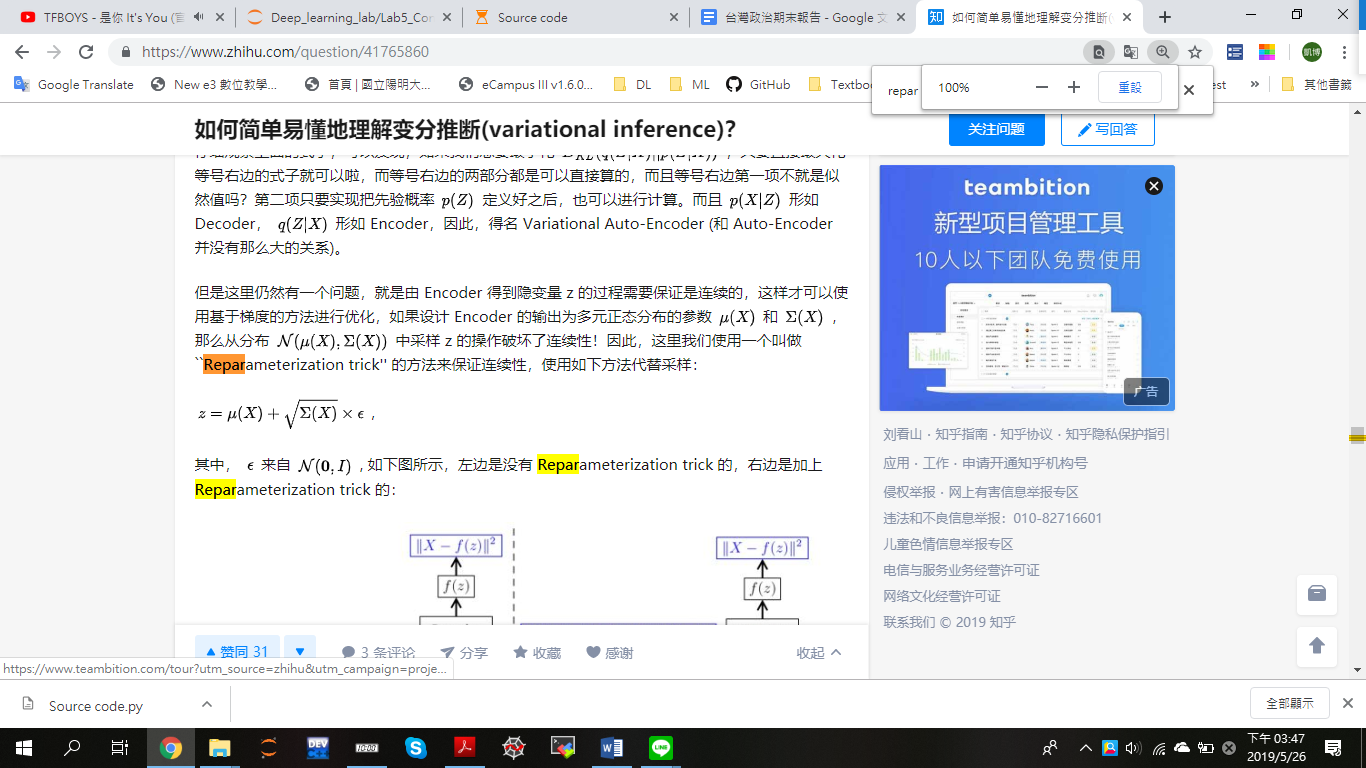
1. Encoder

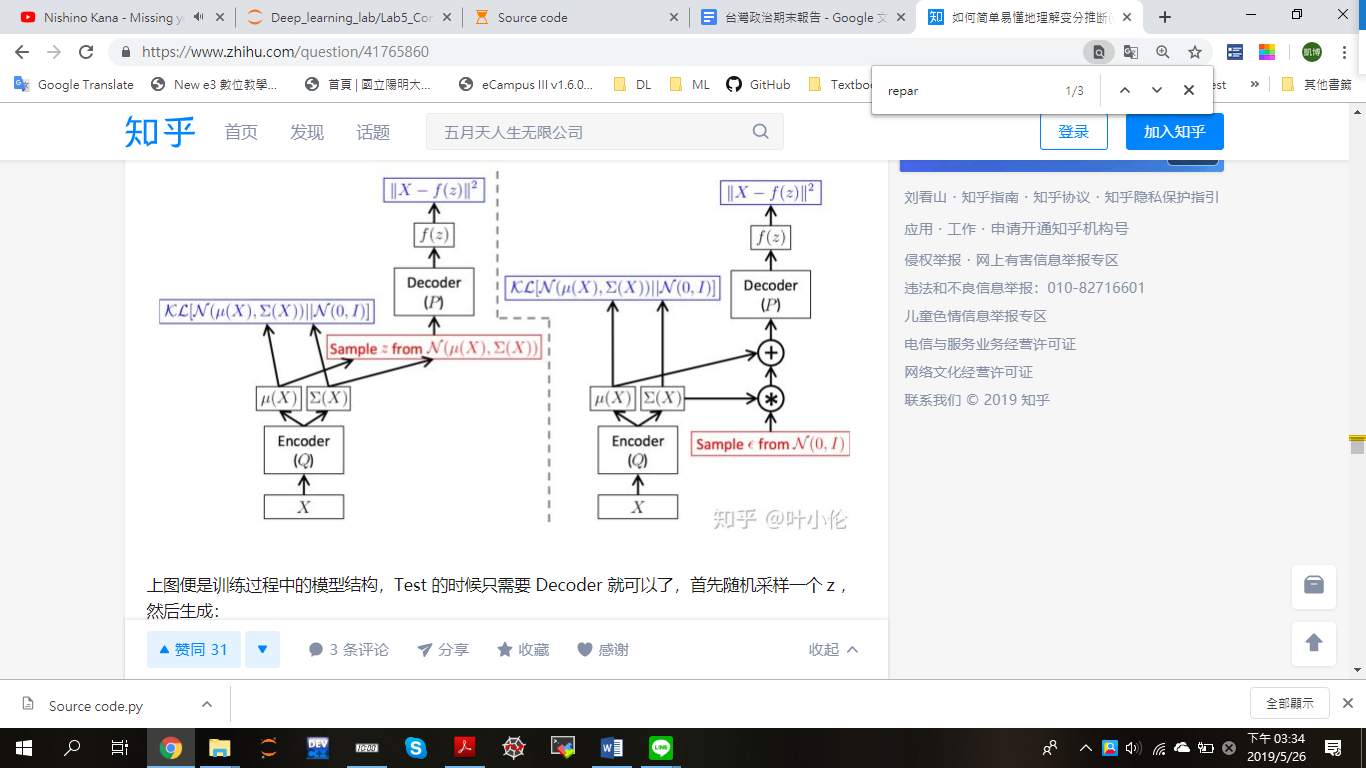
將input word和conditional tense做concatenate後就可以將其輸入到encoder裡面，首先input word會被split成字元的index，經過embedding投影至256的維度後放進GRU，指定變數output為每一層GRU的輸出矩陣，hidden為最後一層的輸出矩陣，encoder最後的return為output和hidden。

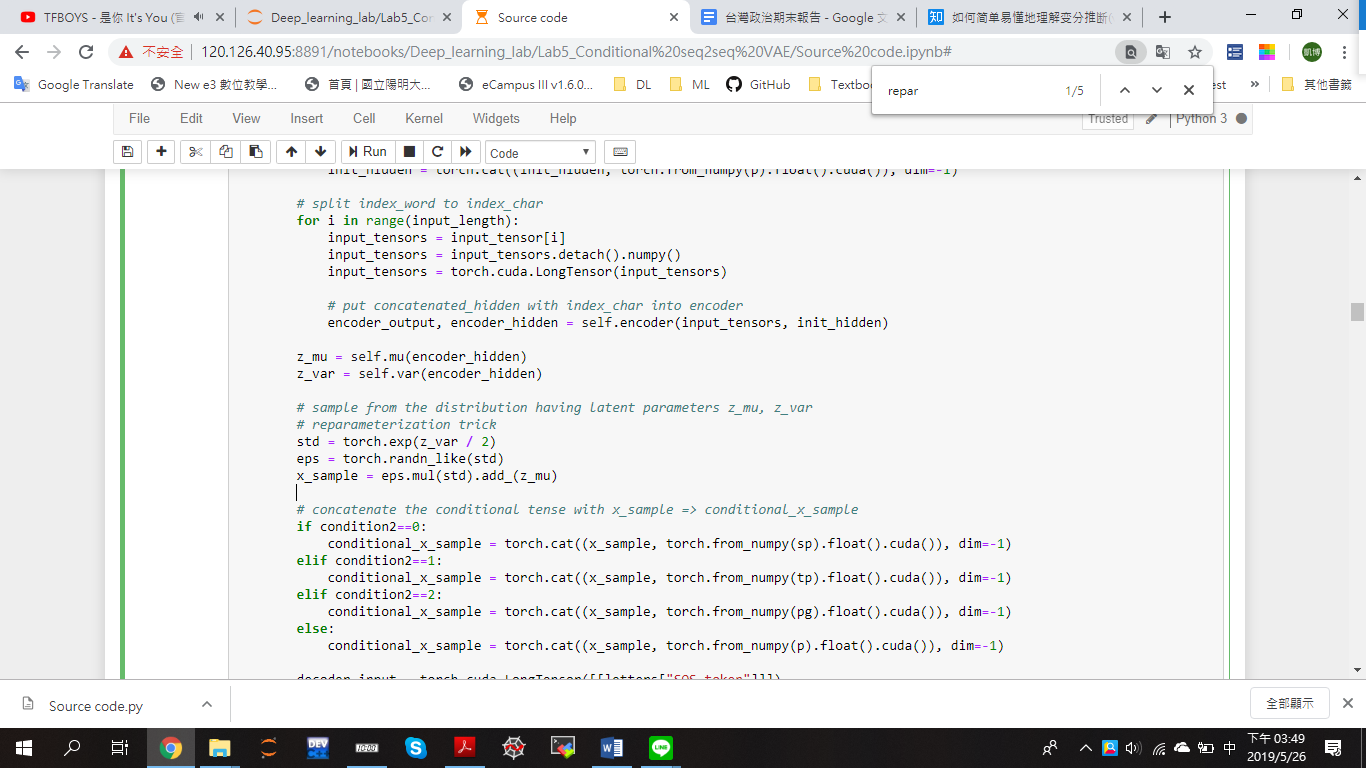


1. Reparameterization trick

Reparameterization trick最主要的目的是要確保從encoder到decoer端的過程中的 hidden (context vector)保有連續的特性，這樣在做backpropagation的微分梯度計算時才不會有問題，以下圖來看，左邊的是沒有經過Reparameterization trick的model，mu(X) 和 Sigma(X)在encoder和decoder端之間屬於離散，但是若經過Reparameterization，

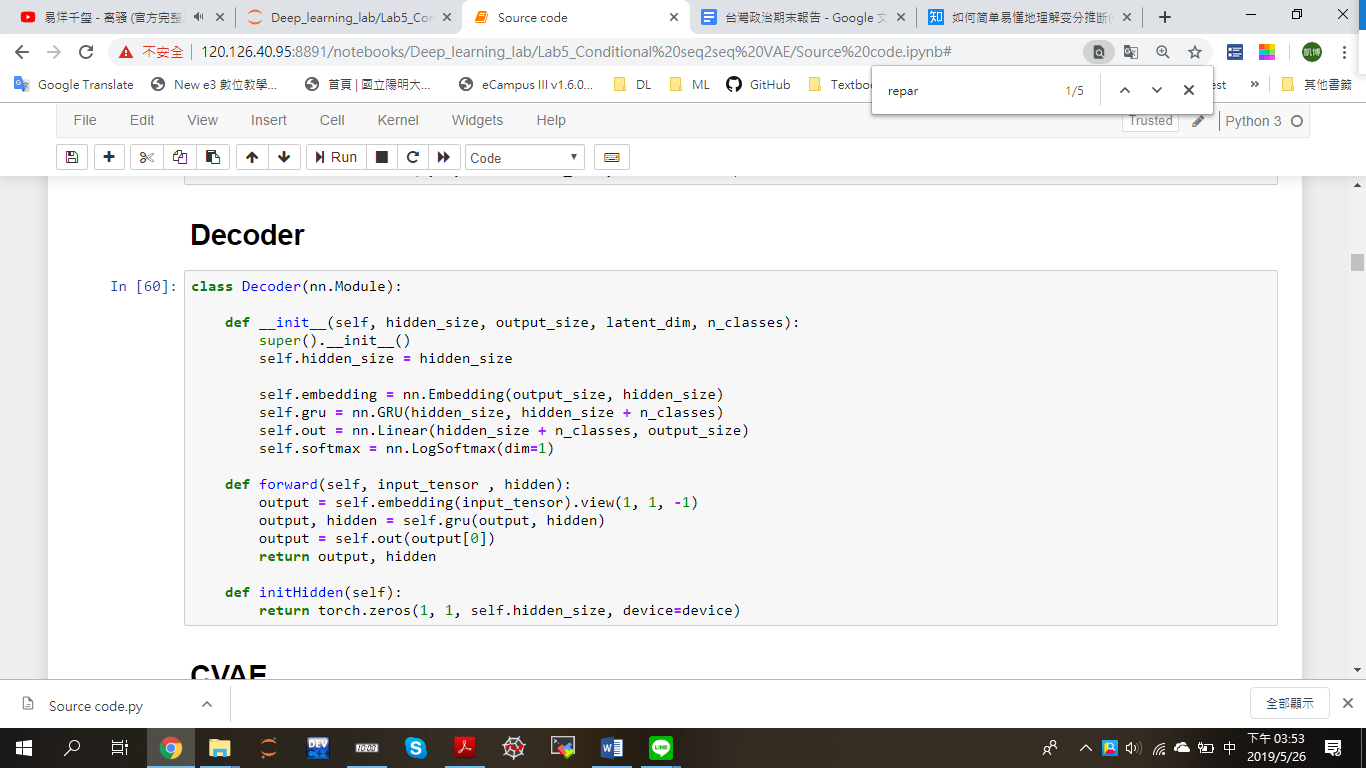
，就可以讓encoder和decoder端之間有連續的關係。





1. Decoder

在decoder的部分，除了一樣要放入和conditional tense concatenate後的hidden外，另外還要再放入SOS或者是GO，來告訴decoder要開始進行decode了，接下來也是經過GRU，和encoder一樣的是指定變數output為每一層GRU的輸出矩陣，hidden為最後一層的輸出矩陣，encoder最後的return為output和hidden，最後用linear的output\_size設定為28。

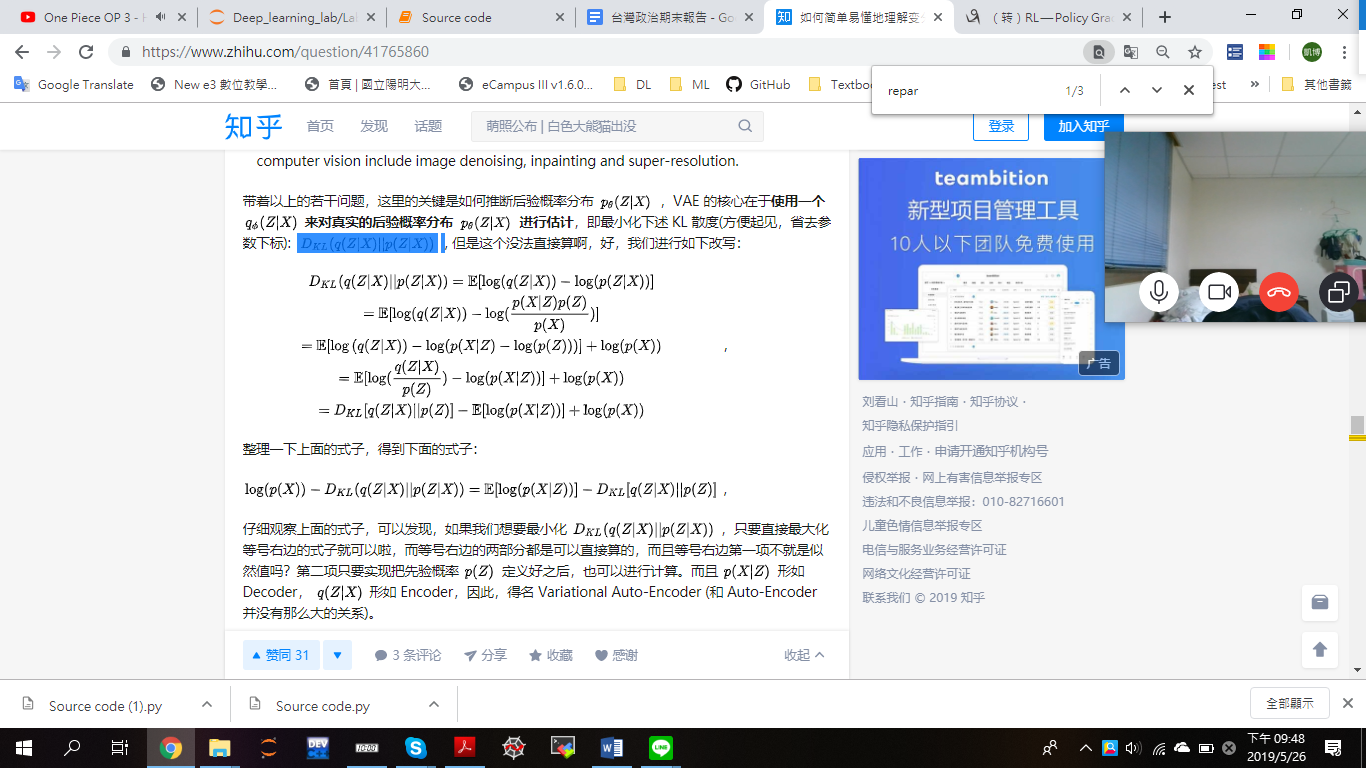


1. Hidden size: 256維
2. condition embedding size: 256維
3. RNN type: GRU
4. Specify the hyperparameters
5. KL weight:

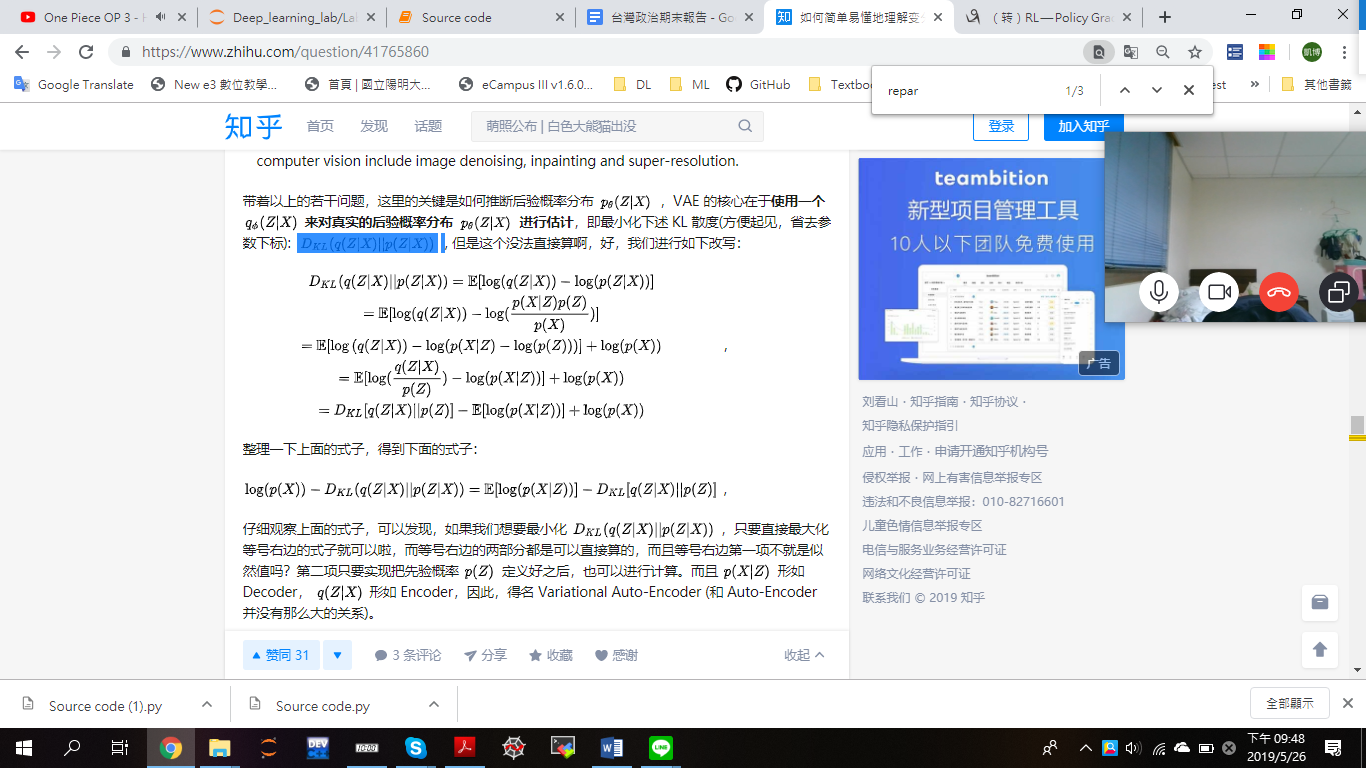
KLD\_weight = 0.01 + 0.01 \* lw，在每一次的forward中，lw都會加1。

1. Learning rate = 0.0001
2. Teacher forcing ratio = 0.5
3. Epochs = 80
4. Vocab\_size = 28 (26個英文字母加SOS和EOS)
5. N\_classes=4
6. Latent\_dim = 256
7. Results and discussion
8. Plot the loss and KL loss curve during training and discuss the results

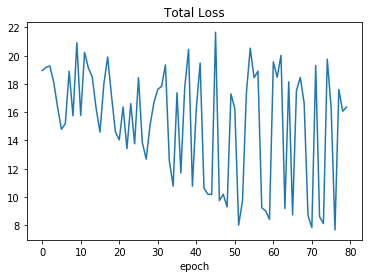
CVAE和VAE的核心在於使用一個q(Z|X)來對真實的posterior probability p(X|Z)進行估計，即最小化q(Z|X)和p(X|Z)的KL divergence: KLD(q(Z|X) || p(Z|X))，但是這個無法計算，所以可以將它寫成:

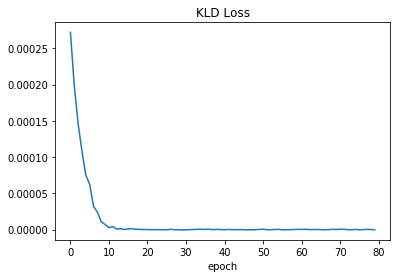


經過簡化:



所以若要最小化KLD(q(Z|X) || p(Z|X))的話，就只要將右邊2項計算出來，第一項就是程式碼中Reconstruction的部分 (Maxium likelihood)，第二項就是KL divergence，由下面2張圖可知，KL divergence loss已趨近於零，如此一來，使total loss上升的就是Reconstruction loss的部分雖然從整體來看loss雖有在下降的趨勢，但是起伏也很大，估計是因為在sampling mu和sigma的過程中(決定prior distribution P(Z) 時)有時會會取到偏離正確值很遠的點，所以才會造成起伏很大的趨勢。





1. Plot the BLEU-4 score of your testing data while training and discuss the result.  
    BLEU，全稱為Bilingual Evaluation Understudy（雙語評估替換），是一種對生成語句進行評估的指標。完美匹配的得分為1.0，而完全不匹配則得分為0.0。在這次的評分中，明顯的分數不是很高，最高的training score為0.128，最高的testing score為0.119，我猜是因為vocab\_size (=28 )太長的原因，所以在encoder和decoder的時候梯度消失導致model無法記住前面的字元，此外，conditional tense的array (1\*4)太小也是個問題，加在word\_index後面無法突顯有condition的作用，改進的方法就是可以用試用attention使得model可以記得較多的東西，或許分數就可以上升了。

