309513061 電信甲 簡詠倫

● RNN and Transformer for News Classification

1.1 Text Preprocessing

1. 我使用 Keras 提供的套件作為 tokenizer。選用 Keras 的原因為在分詞時套件會自動將文字轉換為小寫，加速執行時間。如果用空白鍵來分詞會有標點符號沒有分開的情形，影響訓練正確率。因此用較複雜的分詞模型較能提高模型的精確度。

2. 由於每筆訓練資料的文字長度不一，為使轉換為同一大小之向量，我們即需要 PAD 來填充文字過短的資料。而若在測試時出現原本訓練詞袋沒有的詞，則用 UNK 的索引替代它。

3. 我先使用 Keras 提供的套件進行分詞，並計算出訓練資料新聞標題中最長的長度為 9，因此我使用 PAD 將每個新聞標題的長度都補到 9。再使用史丹佛研究團隊所訓練的詞向量模型將每個詞轉換為 200 維的向量以進行訓練。

1.2 RNN

2. 我使用的是史丹佛研究團隊所訓練的 GloVe 詞向量模型。使用 pretrained 的詞向量模型做 embedding 的好處是，能將兩個屬性相近的詞，轉換為距離相近的向量，因此效果會比隨機初始的向量還要好。

3. 先將訓練資料的每個新聞標題 padding 到長度為 9 個詞，再將每個詞 embedding 成 200 維的向量來訓練。使用的模型為課堂上所提到的 Bidirectional LSTM 模型，並連上全連結層，LSTM 部分的 dropout rate 定為 0.3，全連結層的 activation function 為 tanh，輸出端則是用 softmax。全連結層 units 皆為 128，我的模型使用 5 層全連結層能得到不錯的效果，太多層的模型反而效果變差，因此我的模型全連結層定為 5 層。

1.3 Transformer

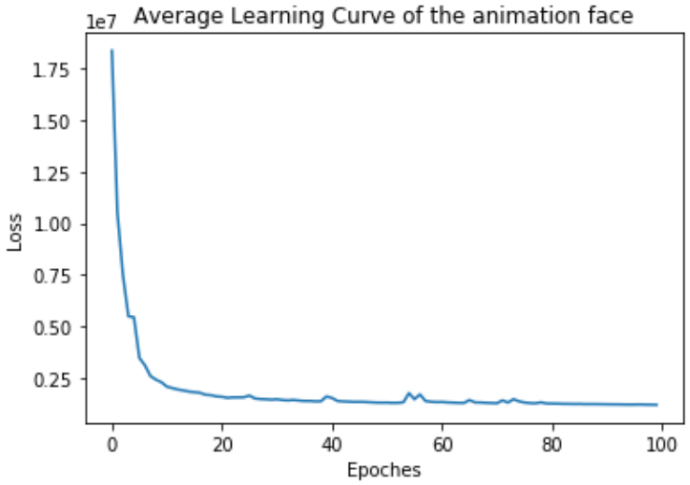
3.

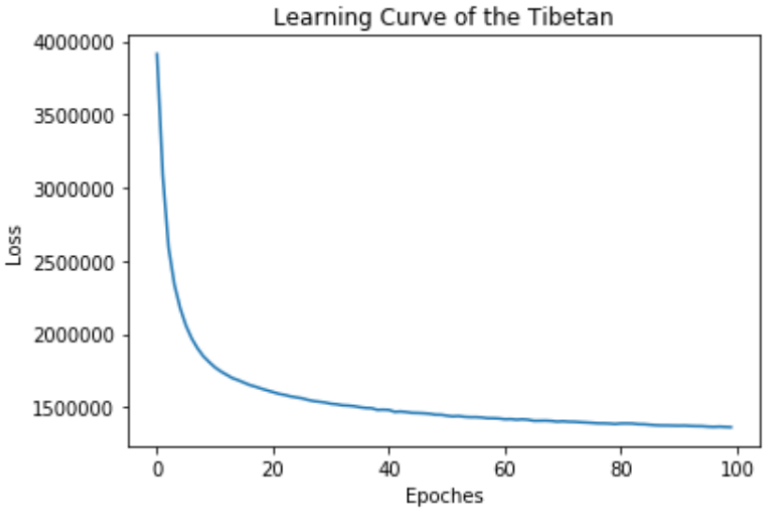
我的 Transformer 在 encoder 的部分，將 Multi-head Attention 層的 head 數當作超參數來設置，其後按照 Transformer 的架構接上 add &

norm 層，最後再接上 FFN 層，並搭配 dropout 及 add & norm。完成 encoder 的結構，而重複執行的次數 N 也是一個超參數。由於是分類模型，我並沒有配置 decoder，而是直接接上 3 層全連結層，activation function 為 tanh，同樣也搭配 dropout，並在最後一層使用 softmax 來求出分類結果。超參數的選擇上，我的 head 數選 12，N 為 3 訓練出的結果是可以超過 Basic baseline 的，但要達到 advanced baseline 則需再改善 model。

1. Variation Autoencoder for Image Classification

1.



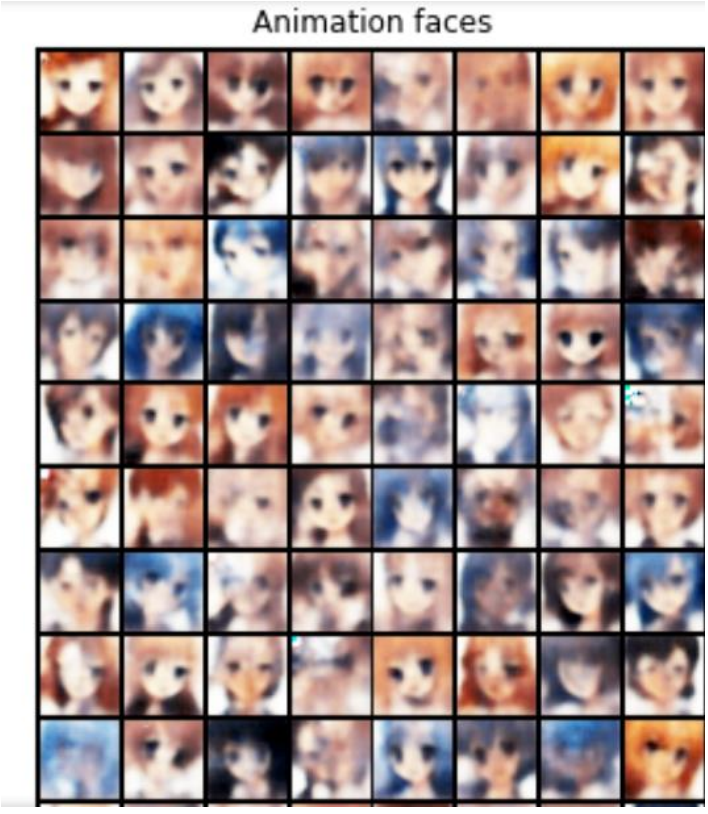

Learning Curve of the Tibetan
4000000
3500000
3000000
2500000
2000000
1500000
Loss
0
20
40
60
80
100
Epoches


Animation faces

Tibetan

2.

Animation faces



Tibetan

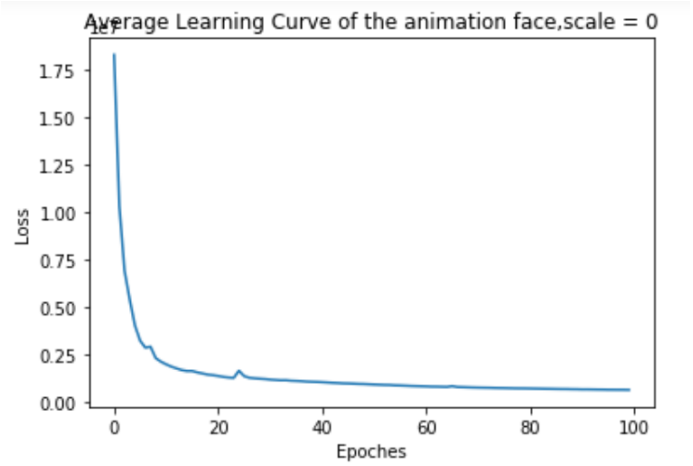
3.

Animation faces

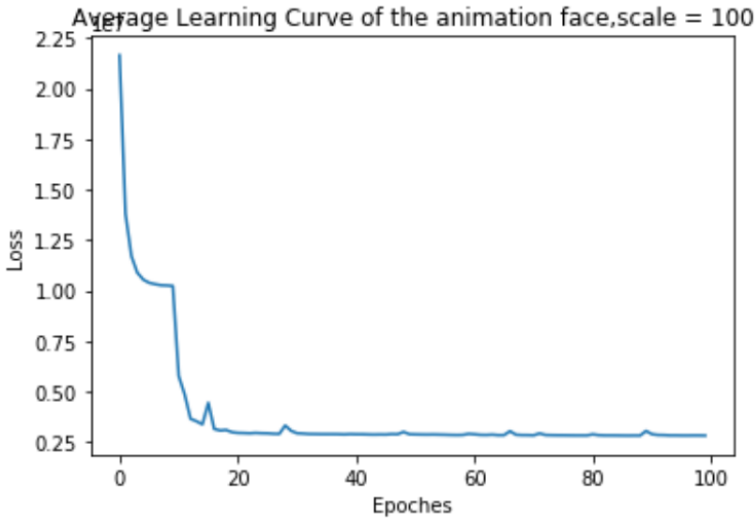
Tibetan

4.

λ = 0 Animation faces 部分:



λ = 100 Animation faces 部分

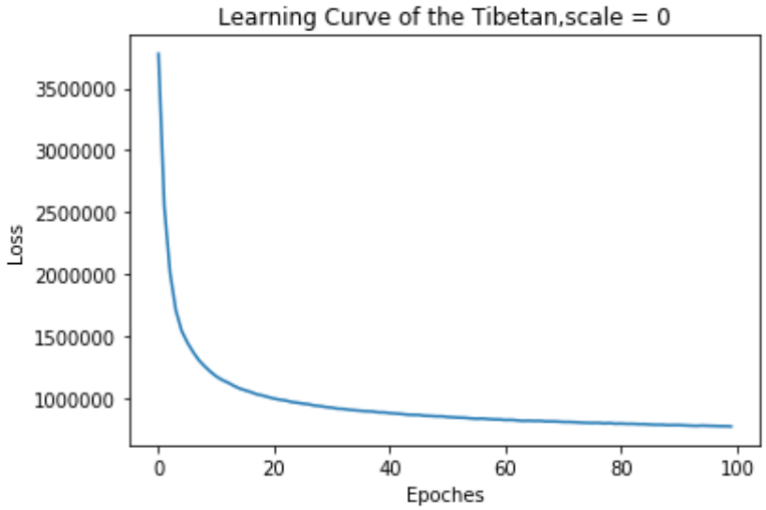


Reconstructed Animation faces,scale = 100

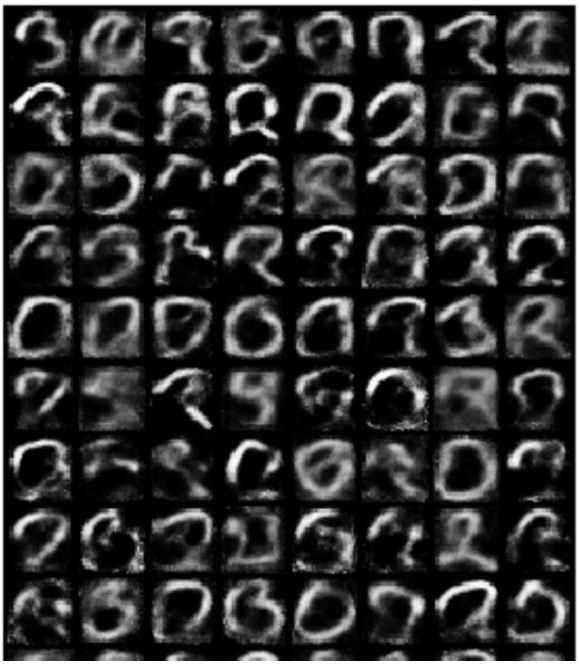
Using the latent codes z to reconstruct , scale = 100

Animation faces interpolation,scale = 100

λ = 0　Tibetan 部分

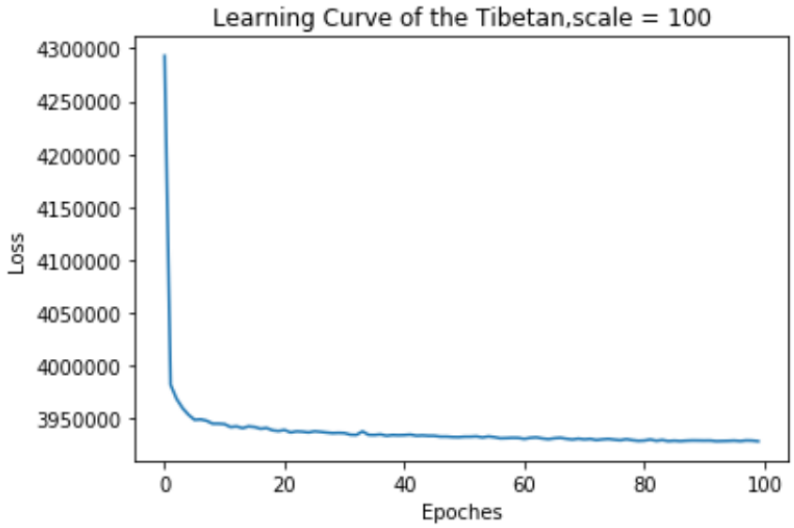


Using the latent codes z to reconstruct , scale = 0



Tibetan interpolation,scale = 0

λ = 100 Tibetan 部分



Reconstructed Tibetan,scale = 100

Using the latent codes z to reconstruct , scale = 100

Tibetan interpolation,scale = 100

由 Learning curve 可發現，scale=100 時收斂較快，但重建圖片的效果明顯不佳。就我自己的理解來說，KL 項有點類似之前

學過的 regularization 項，作為懲罰項防止 over fitting 的發生。若是沒有此項，會造成訓練成效不佳，但此項也不能太大，由此題的結果可看出此結論。