

1.請比較你實作的generative model、logistic regression 的準確率，何者較佳？

generative model :  $(0.85884(\text{public}) + 0.85431(\text{private}))/2 = 0.85657$

logistic regression :  $(0.85823(\text{public}) + 0.85640(\text{private}))/2 = 0.85732$

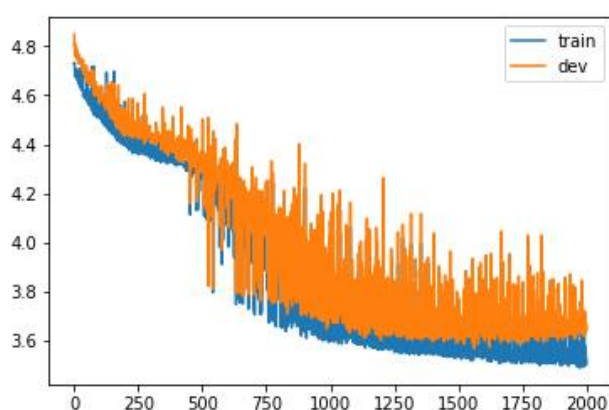
幾結果看來logistic regression的準確率略高一些。

2.請說明你實作的best model，其訓練方式和準確率為何？

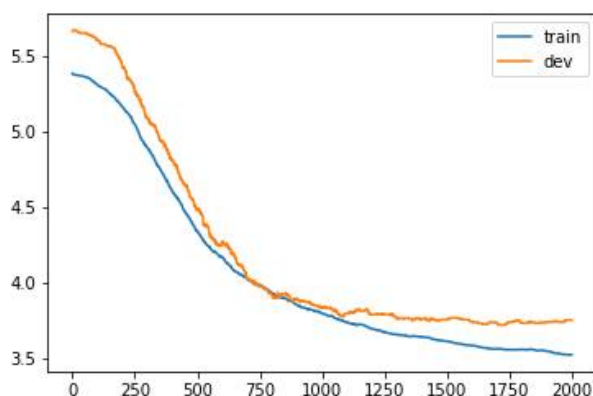
我使用的best model 為logistic regression，正確率為 $(0.85823(\text{public}) + 0.85640(\text{private}))/2 = 0.85732$ ，gradient descent方法為adam，並在訓練過程有配合使用cross validation、feature normalization以防止overfitting，並且觀察validation set之準確率作early stopping。

3.請實作輸入特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

Without feature normalization:



With feature normalization:



以logistic regression並採用相同的gradient descent方法以及相同的iteration，可以看到若使用normalization可以使loss下降得更為平滑、穩定，讓最後訓練出的model可以有比較好的performance。

4.請實作logistic regression 的正規化(regularization)，並討論其對於你的模型準確率的影響。

使用regularization： $(0.85382 + 0.85442) / 2 = 0.85412$

未使用regularization： $(0.85394 + 0.85503) / 2 = 0.85449$

由於我在這次作業中有使用early stopping所以似乎regularization並未起到太大的作用。

5.請討論你認為哪個attribute 對結果影響最大？

以model中各個參數來看是capital\_gain影響力最大，weight為1.9918