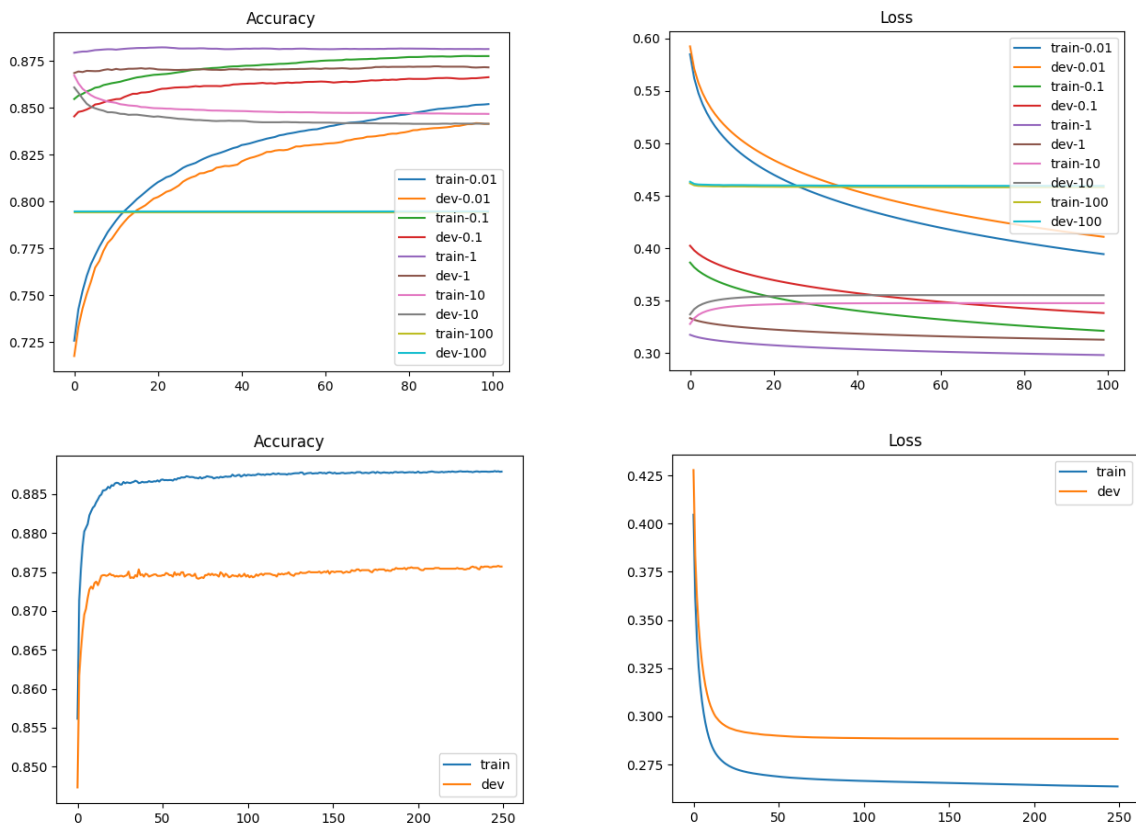


1. (2%) 請比較實作的 generative model 及 logistic regression 的準確率，何者較佳？請解釋為何有這種情況？

在 training data，iteration=100 的 logistic regression 下，準確率為 0.8873606119251682；而 generative model 為 0.8763017233434707，logistic regression 得到的結果較佳。會有這個情況是因為 logistic regression 可以發掘更多 generative model 沒辦法判斷出來，feature 之間的關係，因而也能得到更高的準確率。

另外，將兩者得結果上傳 Kaggle，generative model 在 testing data 的 accuracy 相較於 logistic regression 低了更多，也更說明了 generative model 可能沒辦法很好地連結 feature 與 output 之間的關係。

2. (2%) 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。接著嘗試對正規項使用不同的權重 (lambda)，並討論其影響。(有官 regularization 請參考 <https://goo.gl/SSWGhf> p.35)



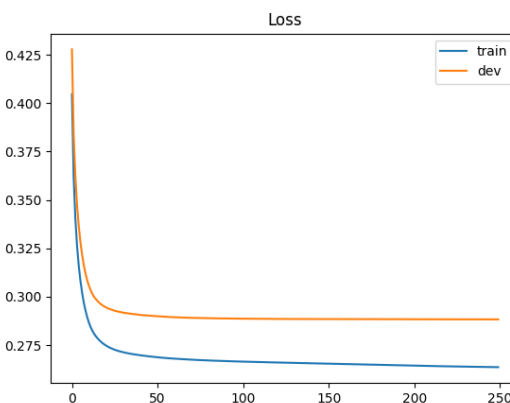
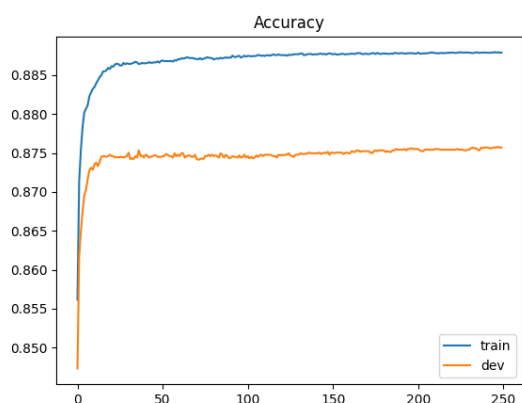
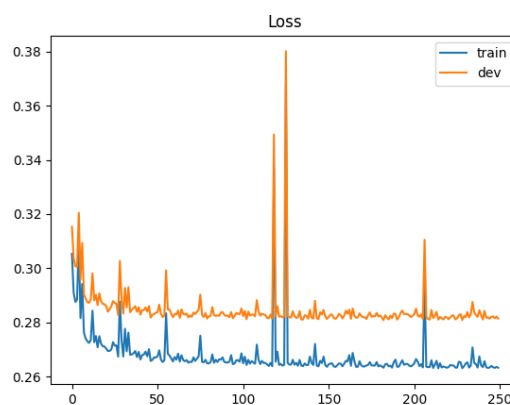
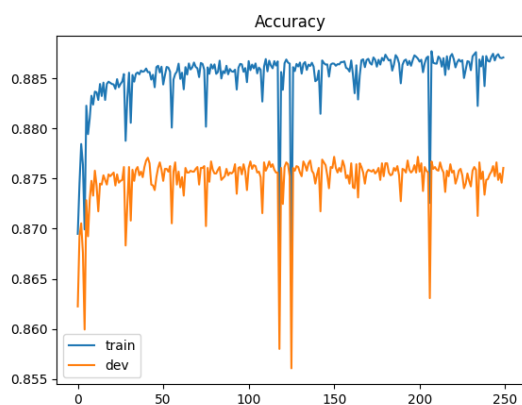
上兩張：regularization，下兩張：沒有 regularization

在測試 weight 的 regularization 下，測試集的 accuracy 都沒有變好，有可能是因為本來就沒有什麼太多很巨大的 weight 誤差。至於對於 $\lambda=[0.01, 0.1, 1, 10, 100]$ ，最好的結果是在 $\lambda=1$ 時出現。

3. (1%) 請說明你實作的 best model，其訓練方式和準確率為何？

我使用 logistic regression，對年齡的區間（0–18 歲、18–25 歲……）做 one hot encoding，再把原本的年齡 feature 刪掉。Normalization 的部分是把每項 feature 減掉 mean 再除以 standard deviation。最後，epoch 為 200、batch size 為 16、eta 為 0.01，並使用 Adagrad 來更新。最後在 Kaggle 上的準確率為 0.89218。

4. (1%) 請實作輸入特徵標準化 (feature normalization)，並比較是否應用此技巧，會對於你的模型有何影響。



上兩張：沒有 normalization，下兩張：normalization

兩者的大致趨勢並沒有什麼差別，但若沒有先做 feature normalization，在 train 的時候比較容易走到錯的地方（因為圖形不那麼對稱跟一致），也造成了上圖中那些 performance 突然變很差的幾次 iteration。