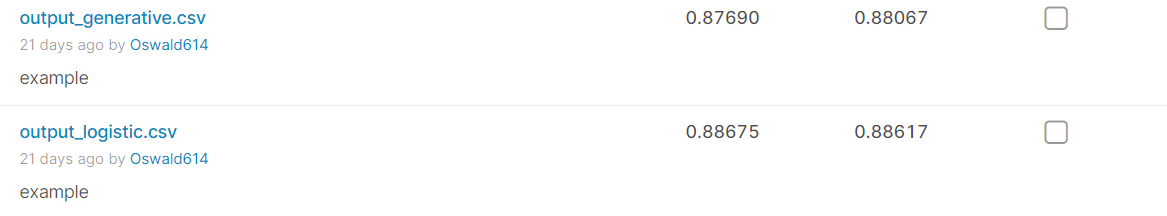
1. (2%) 請比較實作的 generative model 及 logistic regression 的準確率，何者較佳？請解釋為何有這種情況？



Logistic較佳。根據課程講述，在data量少的時候，generative有可能會較佳(因其具有產生關係之特性)，而data量多時，logistic則表現較佳。此次作業，data量約五萬多筆，故logistic表現較佳。

1. (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。(有關 regularization 請參考 <https://goo.gl/SSWGhf> p.35)

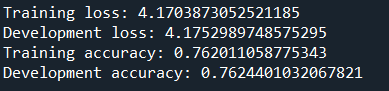
理論上，regularization的目的是防止overfit，故使用regularization會降低training的acc，並提升val的acc。然而，這次在train之acc上卻呈現一微笑曲線，推測是因為此次使用之linear logistic regression並沒有overfit，故結果與理論不相符。

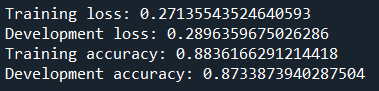
1. (1%) 請說明你實作的 best model，其訓練方式和準確率為何？



我的best model採用將增加feature之2次、3次項，並將與預測較不相干的資料drop掉，且使用Batch size = 1之技巧。經過幾次嘗試，找到最佳的epoch與learning。

1. (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。

未使用feature normalization

使用feature normalization

使用feature normalization會使的不管是在loss function或者是accuracy的部分都表現得更佳。

使用feature normalization可去除特徵之間的相關性 ，使特徵獨立。也可讓所有特徵具有相同的均值和方差 ，讓特徵同分佈。