

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

答：

我實作了 Joint Gaussian distribution 的 generative model。此種機率模型在高維度空間中也十分好表達成數學式，且其函數完全可以簡單的由兩個參數：期望值 \mathbf{m} 與共變異矩陣 Σ 決定，故其訓練方式非常直觀：直接將機率模型中的參數用樣本平均 $\hat{\mathbf{m}}$ 和樣本共變異矩陣 $\hat{\Sigma}$ 代入，再將機率分布函數代入 loss function 就可以得出分辨兩個 class 的函數。

在使用助教給的 feature set 實驗後，此方法在 training set 上的分辨正確率約 84.25%

2.請說明你實作的 discriminative model，其訓練方式和準確率為何？

答：

我實作了 logistic regression，使用 sigmoid function 將 feature 線性組合之後的值從 $(-\infty, \infty)$ 打到 $(0,1)$ 。其訓練方式為 batched Gradient descent，同時試著使用 feature normalization 和 Adam 來加速收斂。並且試著使用 regularization 且觀察其效果。

使用這方法得出的正確率約為 85.2~85.6%

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。

答：

輸入特徵標準化在 logistic regression 中對模型準確率理論上不會有任何影響。

假設一組 $M \times N$ 維的 feature data X 有經過標準化(一個行向量為一組 feature)，則有：

$$\tilde{X}_{m,n} = \frac{(X_{m,n} - \hat{X}_n)}{\sigma_n}$$

考慮此原始資料的 logistic regression 最佳解 w, b 有以下關係式：

$$\begin{aligned} w, b &= \operatorname{argmin}_{w, b} \sum_m C(\sigma(w\tilde{X}_m + b), Y_m) \\ &= \operatorname{argmin}_{w, b} \sum_m C\left(\sigma\left(w \frac{(X_m - \hat{X})}{\sigma} + b\right), Y_m\right) \\ &= \operatorname{argmin}_{w, b} \sum_m C\left(\sigma\left(\frac{w}{\sigma} X_m + \left(-\frac{\hat{X}}{\sigma} + b\right)\right), Y_m\right) \end{aligned}$$

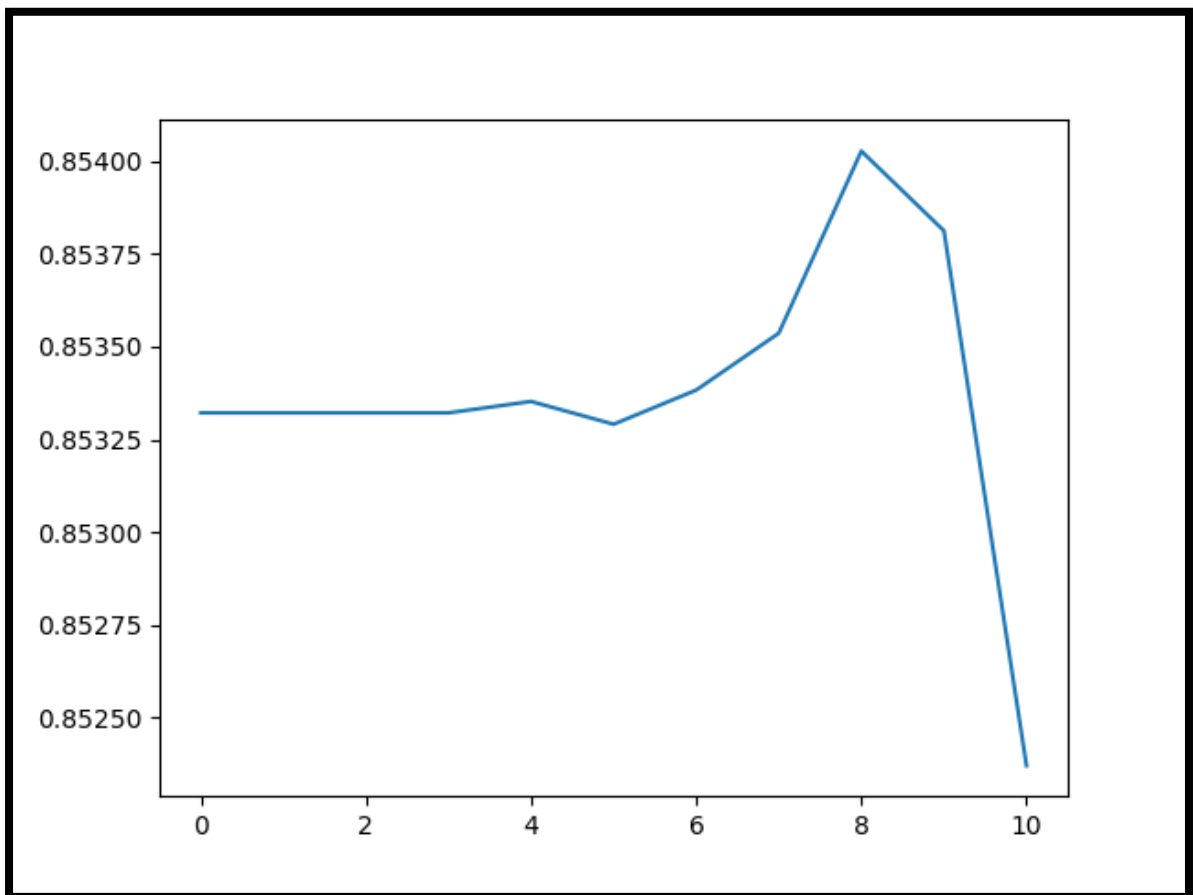
如果令 $\tilde{w} = \frac{w}{\sigma}$, $\tilde{b} = -\frac{\hat{X}}{\sigma} + b$ ，則因為這個關係為線性轉換，故兩者的解空間相同，因此在 \tilde{w}, \tilde{b} 域上的最佳解應和在 w, b 域上的最佳解有對應關係，因此最終得出來的函數也會是同一個。

實作上，標準化加速了 Gradient descent 的效率。也衍伸出了一些實作上的問題例如浮點數運算的誤差等，只要控制好這些變因，輸入特徵標準化對這種簡單 logistic regression 的影響就趨近於零。

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。

答：

以下為實作 regularization 準確率與參數 λ 的關係圖，其中 λ 分別取了 0.01, 0.03, 0.05, 0.1, 0.5, 1, 5, 10, 20, 40, 100 等十個數字，可以發現在這個問題中 λ 取在 20 附近會有較高的準確率。



5. 請討論你認為哪個 attribute 對結果影響最大？

我認為第二個 feature, “fnlwgt” 是影響比較顯著的 attribute，如此判斷的原因是有作了一個實驗將所有 attribute 逐一去除，最後去掉此項的準確率下降得比其他 attribute 顯著的多。