學號：R07946013 系級： 資料科學碩二 姓名：吳泓毅

1. (0.5%) 請比較你實作的generative model、logistic regression 的準確率，何者較佳?

Logistic regression 的準確度比較高，至於兩者再利用HW1 在資料加入noise 達到regularization的效果，結果 logistic regression的準確率反而降了0.01，而generative model 的結果則是提升了0.001；由此推論，兩個模型都不太複雜，所以做regularization反而讓準確率降低。

而 generative model表現比較差還有一種解釋方式是資料的noise其實不是 iid 常態分配的noise。

2. (0.5%) 請實作特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

	Logistic regression	Generative model
normalization	0.85442	0.84361
without normalization	0.76485	0.84410

比起Generative model，Logistic regression 在做了normalization後的表現好很多。Generative model 雖然退步了0.0005，但因為幅度太小可以先當作沒有影響。而打開沒做normalization 的 Logistic regression 會發現，他對所有的測資都預測 Positive；但實際上Logistic regression 是線性模型，準確度不應該受到 normalization影響，所以可以得到的結論是 normalization 可以幫助Logistic regression 的收斂。

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

我使用的是sklearn greadient tree boosting，開100顆tree 每顆的深度不超過2。不論是在training 還是 testing 都達到了準確率0.86以上的成績，而且可能因為是很多顆Tree的關係，overfitting的情況不明顯；原本還打算用nn 的方式做，但怎麼Train 在Training上個準確率一直都沒辦法超過0.86。

4. (3%) Refer to math problem
https://hackmd.io/0fDimqO7RaSCPpD_minSGQ?both

1. $L = \prod_{k=1}^{K} \pi_k^{n_k}$

$$l = \sum_{k=1}^{K} n_k \log \pi_k$$

Let $L(\pi, \lambda) = \sum_{k=1}^{K} n_k \log \pi_k + \lambda(1 - \sum \pi_k)$ ~~=~~ 0

$$\frac{\partial L(\pi,\lambda)}{\partial \pi_k} = 0 \Rightarrow \frac{n_k}{\pi_k} - \lambda = 0 \Rightarrow \pi_k = \frac{n_k}{\lambda} \Rightarrow \pi_k = \frac{n_k}{N}$$

Scanned with CamScanner

---

2. Known: $A^{-1} = \frac{1}{\det(A)} \cdot adj(A)$, $adj(A)_{i,j} = \det(A_{-i,-j})$ where $A_{-i,-j}$ is the block deleting the i-th row, j-th col of A.

$$\det \Sigma = \sum_{k=1}^{m} (-1)^{i+k} \cdot \sigma_{ik} \cdot \det(\Sigma_{-i,-k})$$

$$\therefore \frac{\partial \log \det \Sigma}{\partial \sigma_{ij}} = \frac{1}{\det \Sigma} \cdot \frac{\partial}{\partial \sigma_{ij}} \det \Sigma$$

$$= \frac{1}{\det \Sigma} \cdot (-1)^{i+k} \cdot \det(\Sigma_{-i,-j}) = e_j \Sigma^{-1} e_i^T \ . \ \#$$

3. $L = \prod_{k=1}^{K} \prod_{\substack{i=1 \\ t_{ik}=1}}^{N} N(x_i; \mu_k \Sigma)$

$$l = \sum_{k} \sum_{\substack{i \\ t_{ik}=1}} -\frac{m}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_i-\mu_k)^T\Sigma^{-1}(x_i-\mu_k)$$

Let $\frac{\partial l}{\partial \mu_k} = 0$: $\sum_{\substack{i=1 \\ t_{ik}=1}}^{N} \Sigma^{-1}(\mu - x_i) = 0 \Rightarrow \sum_{\substack{i=1 \\ t_{ik}=1}}^{N} \mu - x_i = 0 \quad \Rightarrow \mu_k = \frac{1}{n_k}\sum_{i=1}^{N} t_{ik}\, x_i$

$$l = \sum_{k} \sum_{\substack{i \\ t_{ik}=1}} -\frac{m}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma^{-1}| - \frac{1}{2}\operatorname{tr}\left[(x_i-\mu_k)(x_i-\mu_k)^T\Sigma^{-1}\right]$$

$\frac{\partial l}{\partial \Sigma^{-1}} = 0$: $\sum_{k} \sum_{\substack{i \\ t_{ik}=1}} \frac{1}{2}\Sigma - \frac{1}{2}(x_i-\mu_k)(x_i-\mu_k)^T = 0$

$$\Rightarrow N\cdot\Sigma - \sum_{i=1}^{N} t_{ik}(x_i-\mu_k)(x_i-\mu_k)^T = 0$$

$$\Rightarrow \Sigma = \frac{1}{N}\sum_{i=1}^{N} t_{ik}(x_i-\mu_k)(x_i-\mu_k)^T$$

$$= \sum_{k=1}^{K} \frac{N_k}{N} \cdot \frac{1}{N_k}\sum_{i=1}^{N} t_{nk}(x_i-\mu_k)(x_i-\mu_k)^T .$$