學號：r09922065　系級：資工碩二 姓名：張庭逸

執行方式：確保資料跟程式碼都在同個路徑，執行

python <code.py> <sol_name.csv>即可

1. (1%) 請比較說明generative model、logistic regression兩者的異同為何？再分別列出本次使用的資料中五個分得正確/不正確的sample，並說明為什麼如此？

Condition	Kaggle	
	Public	Private
Generative model	**0.83894**	**0.83048**
Linear regression	**0.85393**	**0.84989**

根據跑出來的數據，推測Generative model因為對資料已經有預估的分佈（Gaussian），因此會對一些資料有錯誤的描述跟預期，而logistic regression比較沒有這個問題，因此會有較好的表現。

2. (1%) 請實作兩種feature scaling的方法 (feature normalization, feature standardization)，並說明哪種方法適合用在本次作業？

Condition	Kaggle	
	Public	Private
Normalization	0.84545	0.83871
Standardization	0.85614	0.85579

以上表格為用hw2_best.py，epoch改為300，分別做Normalization及Standardization跑出來的數據，可看出Standardization表現較好。

我認為本次作業較適合Standardization，不只是從數據上分數較好，也因為Standardization可以讓資料較接近常態分佈，讓outlier影響更小，因此較適合在大部分的資料預處理，包含此次作業。

3. (1%) 請說明你實作的best model及其背後「原理」為何？你覺得這次作業的dataset比較適合哪個model？為什麼？

我做最好的 model 是把以及fnlwgt這項代表調查人的ID拿掉，然後把age, capital_gain, capital_loss, hours_per_week這些參數乘上2~5次方，接在原本的feature的後面（新增額外feature），做epoch為1000的logistic regression。原因為，其他參數皆為1,0，只有這些項會有明顯差異，因此放大這些項，接成新的參數，會讓各筆資料有較大差異，logistic regression的表現會更好。

4. (3%) Refer to math problem

1

令x_n屬於的Class為C_{x_n}。

$$P(x_1, x_2, ..., x_N) = \prod_{n=0}^{N} P(C_{x_n}) P(x_n | C_{x_n})$$

$$\begin{aligned} logP(x_1, x_2, ..., x_N) &= \sum_{n=0}^{N} logP(C_{x_n}) + \sum_{n=0}^{N} logP(x_n | C_{x_n}) \\ &= \sum_{n=0}^{N} N_k logP(C_k) + \sum_{n=0}^{N} logP(x_n | C_{x_n}) \\ &= \sum_{n=0}^{N} N_k log\pi_k + \sum_{n=0}^{N} logP(x_n | C_{x_n}) \end{aligned}$$

因為限制在$\sum_{k=0}^{N} \pi_k = 1$的狀況下，要求$logP(x_1, x_2, ..., x_N)$的最大值，因此令$f = (logP(x_1, x_2, ..., x_N))$，$g = \sum_{n=0}^{N} \pi_k$，套用Lagrange multiplier。

$$\nabla f = \lambda \nabla g$$

$$\begin{aligned}
\frac{\partial}{\partial \pi_i} f &= \frac{\partial}{\partial \pi_i}(\sum_{n=0}^{N} N_k log \pi_k + \sum_{n=0}^{N} log P(x_n|C_{x_n})) \\
&= \frac{\partial}{\partial \pi_i} N_i log \pi_i + 0 \\
&= \frac{N_i}{\pi_i}
\end{aligned}$$

$$\frac{\partial}{\partial \pi_i} g = \frac{\partial}{\partial \pi_i} \sum_{n=0}^{N} \pi_k = \frac{\partial}{\partial \pi_i} \pi_i = 1$$

$$\nabla f = \lambda \nabla g$$

$$\begin{bmatrix} \frac{\partial}{\partial \pi_1} f \\ \frac{\partial}{\partial \pi_2} f \\ . \\ . \\ . \\ \frac{\partial}{\partial \pi_K} f \end{bmatrix} = \lambda \begin{bmatrix} \frac{\partial}{\partial \pi_1} g \\ \frac{\partial}{\partial \pi_2} g \\ . \\ . \\ . \\ \frac{\partial}{\partial \pi_K} g \end{bmatrix}$$

$$\begin{bmatrix} \frac{N_1}{\pi_1} \\ \frac{N_2}{\pi_2} \\ . \\ . \\ . \\ \frac{N_K}{\pi_K} \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}$$

可得$\pi_i = \frac{N_i}{\lambda}$，而

$$\sum_{n=0}^{N} \pi_k = \sum_{k=0}^{N} \frac{N_i}{\lambda} = \frac{N}{\lambda} = 1, \lambda = N$$

因此

$$\pi_i = \frac{N_i}{\lambda} = \frac{N_i}{N} \square$$

2

$$\frac{\partial log(det \sum)}{\partial \sigma_i j} = \frac{1}{det \sum} \frac{\partial det \sum}{\partial \sigma_i j} = \frac{1}{det \sum} (-1)^{i+j} |A_{ij}| = \frac{adj \sum}{det \sum} e_j e_i^T = e_j \sum^{-1} e_i^T \square$$