學號：B03502125　系級：　機械四　姓名：倪嘉宏

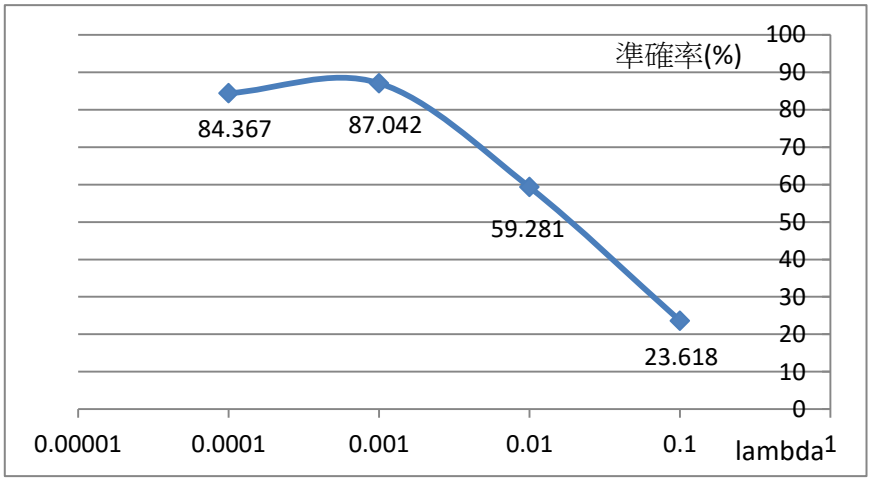
1.請比較你實作的 generative model、logistic regression 的準確率，何者較佳？
答：
以上傳至 kaggle 得到的分數來說，Logistic regression 有著略高的準確率，但兩者的準確率差異不大。
generative model 落在 0.843 左右，而 Logistic regression 則可到達 0.846 的準確率。

2.請說明你實作的 best model，其訓練方式和準確率為何？
答：
我的 best model 是使用套件 Xgboost。經過調整數次的參數後，可達到 0.874 的準確率，且運算時間遠遠少於自己手刻的 generative model 以及 logistic model 。

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：
以我的程式而言，Logistic regression 不使用 normalization 時，準確率約略落在 0.76，而使用 normalization 時，準確率則是 0.84。
Generative model 中，不使用 normalization 時，準確率約略落在 0.75(但是浮動很大)，而使用 normalization 時，準確率則是 0.84。
依照上述兩種情況來看，normalization 對於我的程式有著非凡而不可或缺的助益。

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：
依據不同的 lambda 調整 regularization 的影響力，得到下列圖表。

lambda	準確率
0.1	23.618
0.01	59.281
0.001	87.042
0.0001	84.367
0	84.306



5.請討論你認為哪個 attribute 對結果影響最大？

以 Logistic Regreesion 跑出參數以後，觀察參數數值較大及較小的，合理推論就是對於結果影響較大的 attribute。以下是依此方法所抓出的參數。

正向最大參數	係數	負向最大參數	係數
capital_loss	6.097448	Separated	-1.53319
hours_per_week	2.351087	12th	-1.56106
Federal-gov	2.349883	Wife	-1.5794
age	1.587537	Other-relative	-1.59925
fnlwgt	1.587537	Black	-1.60233
HS-grad	1.449586	Asian-Pac-Islander	-1.69276
Some-college	1.271715	Own-child	-1.74682
sex	0.886843	Assoc-acdm	-1.77824
Preschool	0.777065	9th	-1.81326
capital_gain	0.598463	Unmarried	-2.40301

從這些資料可以觀察到很多很有趣的現象。

不難想像工時長的人與薪水高的人之間的關聯，另外年齡大的人普遍較有可能有高年薪於情於理也都說得過去。但另一方面則可以觀察到一些不太能夠直覺聯想到的關聯性，比如說投資損失最多的人其實反而很有可能年薪超過 50 萬美元。甚至還有非常匪夷所思的現象，比如說教育程度為幼稚園的人相對更有可能年薪大於 50 萬，而只有國中或高中畢業的人卻很有可能年薪少於 50 萬。