

計算思維與問題解決

Final Project

學生姓名:莊上緣

學生系級:資訊系112

學生學號:F74086250

說明

- 這份description會包含以下幾個部分:

- 1.第一題~第七題的:

- 解題時的一些文字說明及程式思路
 - 需印在cmd之執行結果(文字)
 - 需印出之圖片
 - cmd的執行結果截圖(會和需印出之圖片以四張為單位放在同一頁)

- 2.關於這門課的學期修課心得與收穫

- 註:

- 1.由於這次教授給的作業說明之PDF中的“ What you have to do” 包含了7點，因此我在這份description中就會以1~7題來稱呼他們。

- 2.由於在切分trainSet與testSet時皆會用到random，因此以下的執行結果中僅為這份程式某次之結果，有些地方可能和教授的PDF中執行出來的數值與圖形有些微的不同，是隨機抽樣所導致的。

- 3.由於在logistic regression中，也就是第三題~第五題，會有總共9000次的trial，所以執行程式會需要等好幾分鐘才會全部執行完(且每次跳出圖片後，需要關閉圖片，程式才會繼續往下執行，因為我是每題都有一個plt.show())。

- 4.在第三題~第五題中，由於會進行許多次1000trial，為了在等待時確保程式有在運行，我有做了進度條的顯示，每100個trial就會印出一條目前進度訊息。

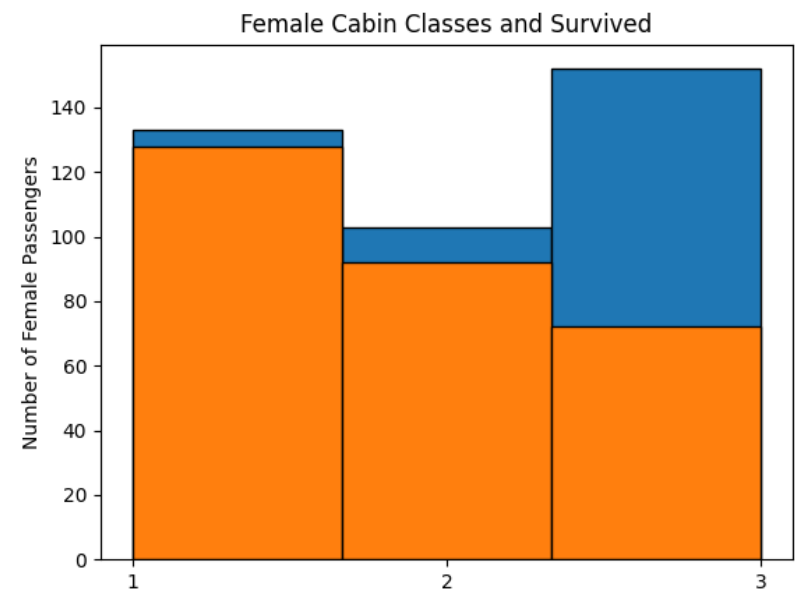
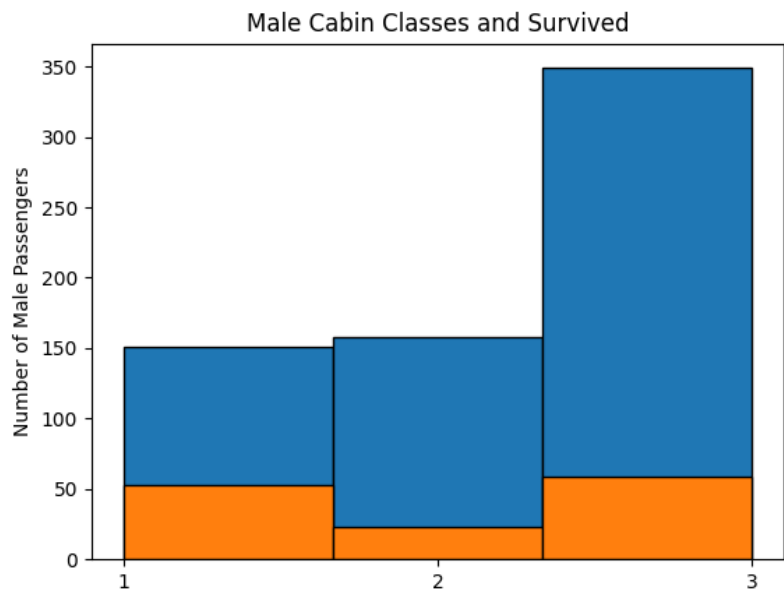
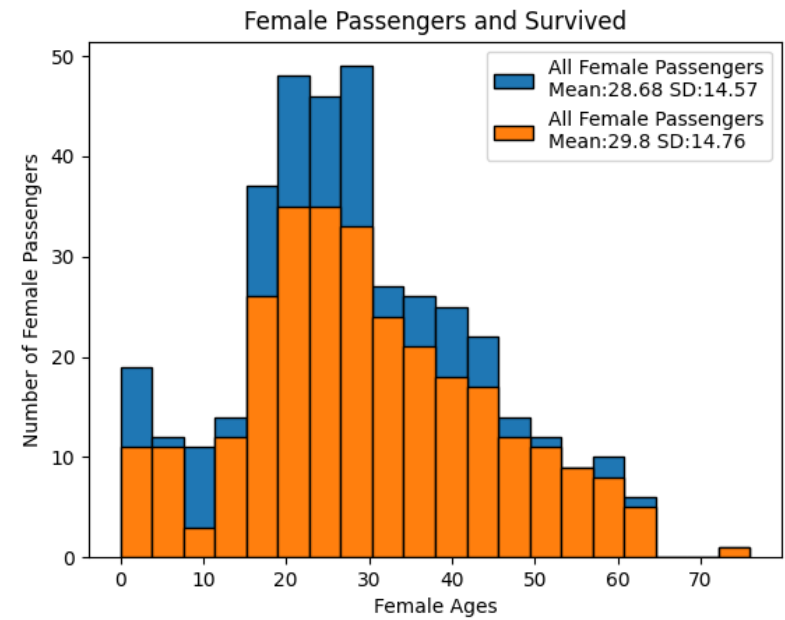
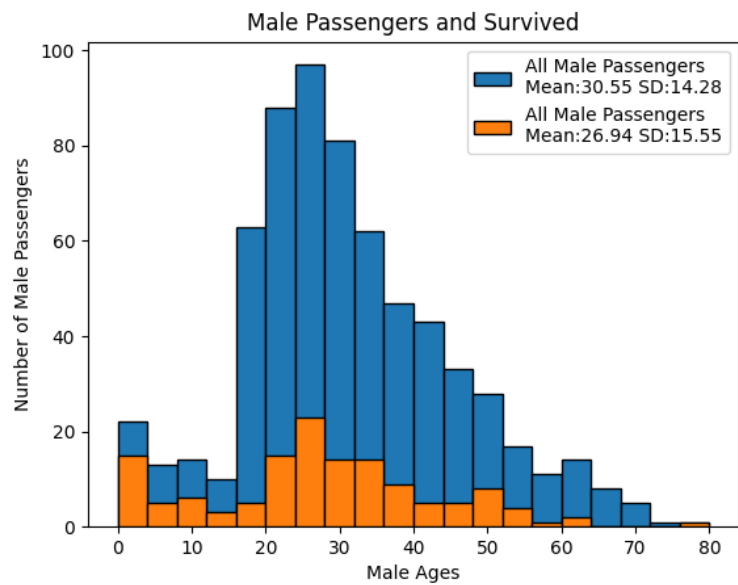
第一題

- 讀檔且建立object，這裡我是使用老師在講解馬拉松選手那個範例時所使用的程式碼去改的。

```
605 #####
606 #####
607 #####
608 ##### Question 1 #####
609 examples=buildPassengerExamples('TitanicPassengers.txt')
610 #####
611 #####
```

第二題

- 1.將不同性別分開並繪製總人數與生還人數的直方圖
- 2.同上，分開性別，且分成不同的CabinClass去畫總人數與生還人數之直方圖



第三題

- 對所有乘客以logistic regression進行處理，畫圖並得出各種統計量如下：

Logistic Regression:

Averages for all examples 1000 trials with $k=0.5$

Mean weight of C1 = 1.143, 95% confidence interval = 0.12

Mean weight of C2 = -0.084, 95% confidence interval = 0.101

Mean weight of C3 = -1.059, 95% confidence interval = 0.113

Mean weight of age = -0.033, 95% confidence interval = 0.006

Mean weight of male gender = -2.407, 95% confidence interval = 0.156

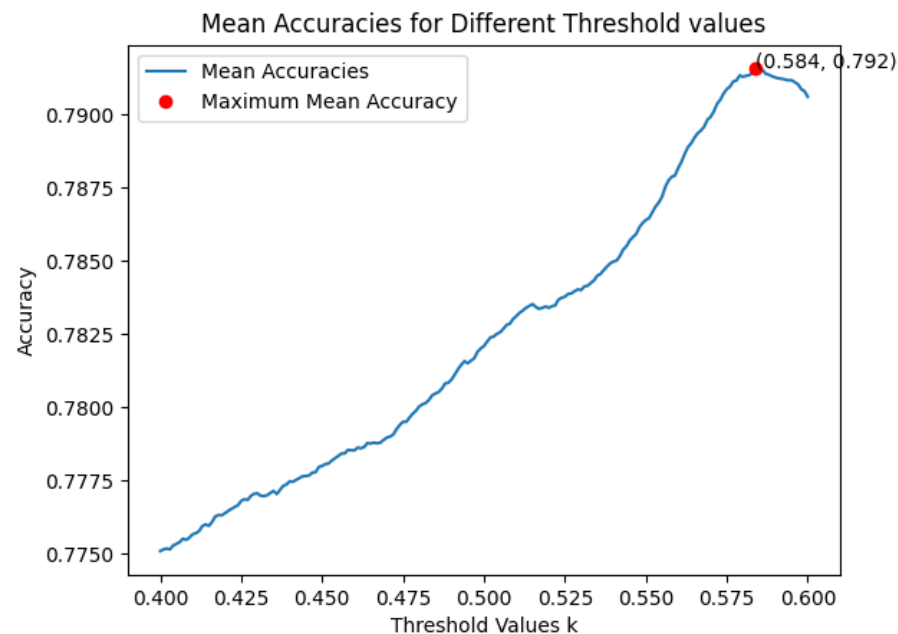
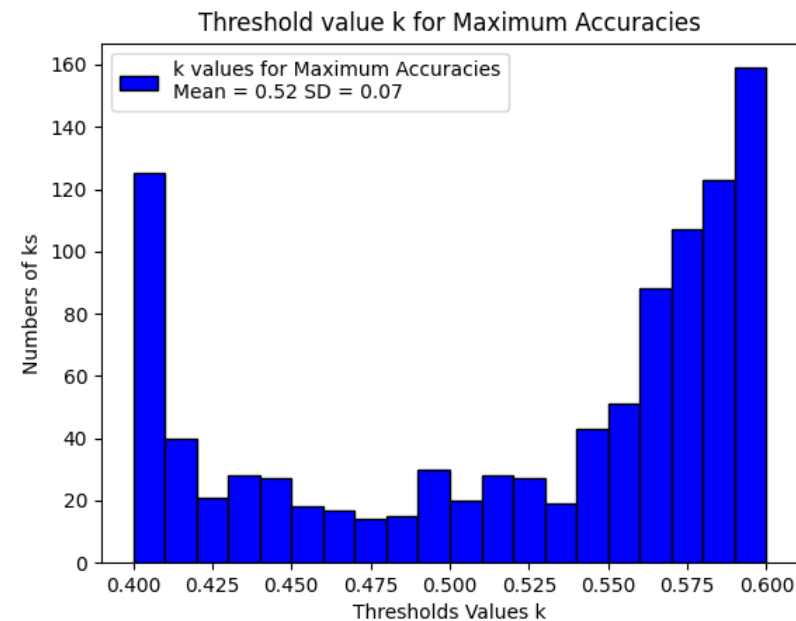
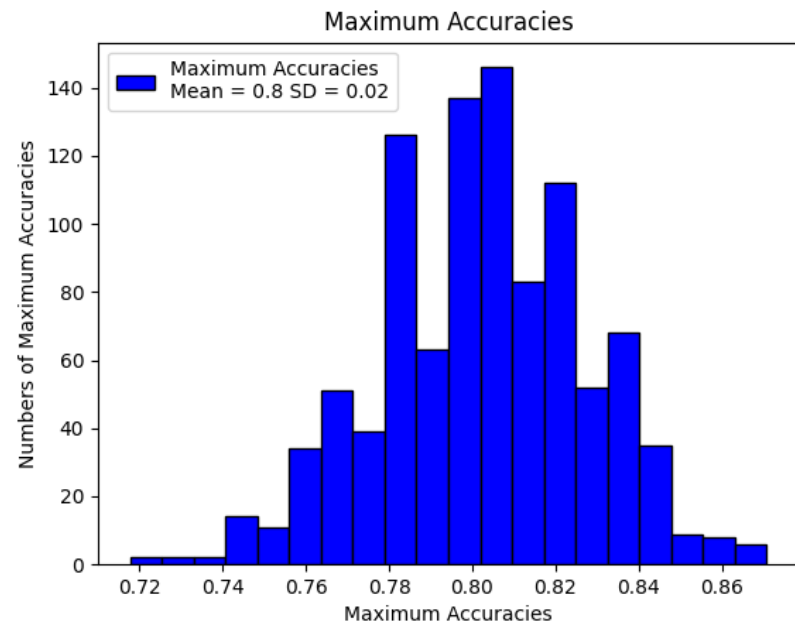
Mean accuracy = 0.782, 95% confidence interval = 0.051

Mean sensitivity = 0.703, 95% confidence interval = 0.093

Mean specificity = 0.837, 95% confidence interval = 0.065

Mean pos. pred. val. = 0.748, 95% confidence interval = 0.094

Mean AUROC = 0.839, 95% confidence interval = 0.053



```
D:\NCKUCSIE\grade3\grade3-2\Computational_Think_and_Problem_Solving\Fi
t>python final_project.py
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)

Logistic Regression:
Averages for all examples 1000 trials with k=0.5
Mean weight of C1 = 1.143, 95% confidence interval = 0.12
Mean weight of C2 = -0.084, 95% confidence interval = 0.101
Mean weight of C3 = -1.059, 95% confidence interval = 0.113
Mean weight of age = -0.033, 95% confidence interval = 0.006
Mean weight of male gender = -2.407, 95% confidence interval = 0.156
Mean accuracy = 0.782, 95% confidence interval = 0.051
Mean sensitivity = 0.703, 95% confidence interval = 0.093
Mean specificity = 0.837, 95% confidence interval = 0.065
Mean pos. pred. val. = 0.748, 95% confidence interval = 0.094
Mean AUROC = 0.839, 95% confidence interval = 0.053
```

第四題(zScaling)

- 對所有乘客的age進行z-scaling之後，以logistic regression進行處理，畫圖並得出各種統計量如下：

Logistic Regression with zScaling:

Averages for all examples 1000 trials with k=0.5

Mean weight of C1 = 1.138, 95% confidence interval = 0.119

Mean weight of C2 = -0.081, 95% confidence interval = 0.098

Mean weight of C3 = -1.056, 95% confidence interval = 0.115

Mean weight of age = -0.477, 95% confidence interval = 0.086

Mean weight of male gender = -2.406, 95% confidence interval = 0.154

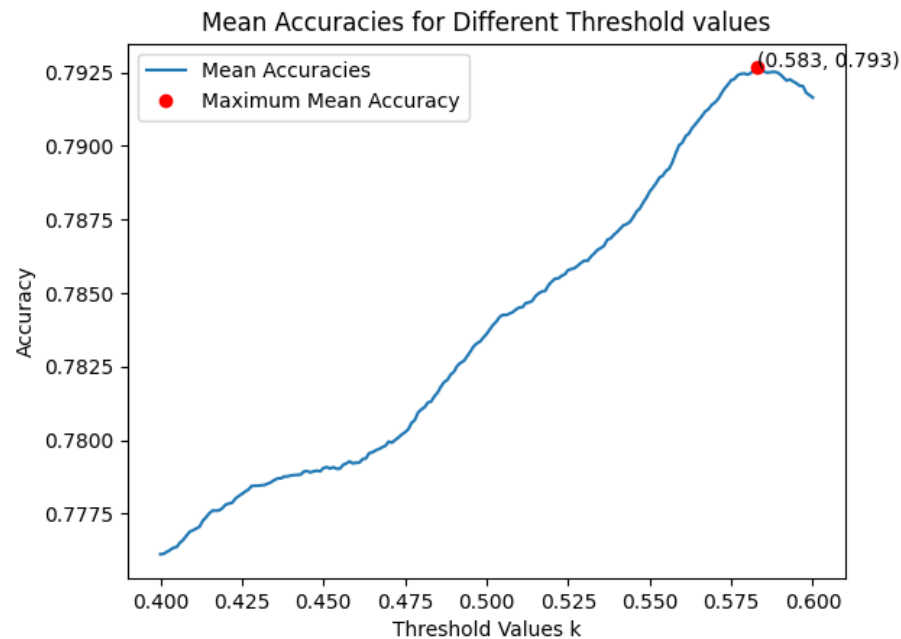
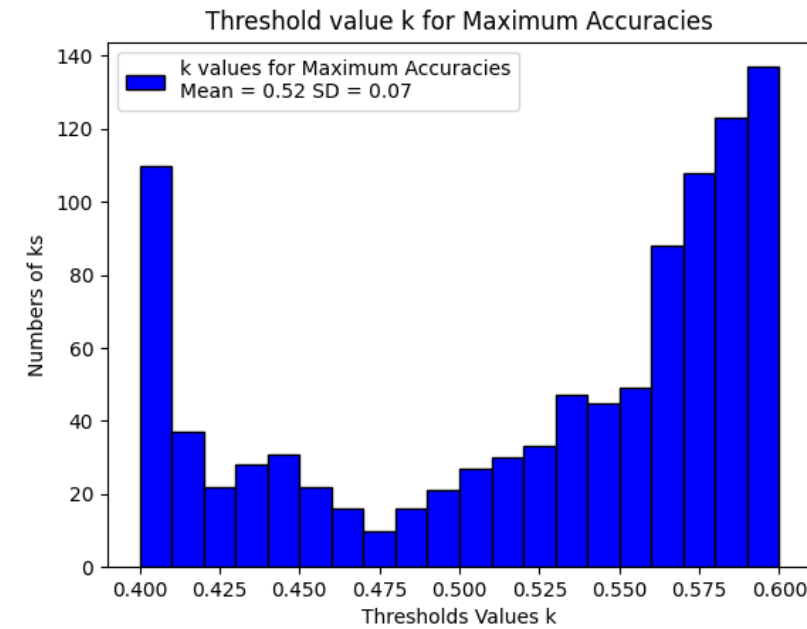
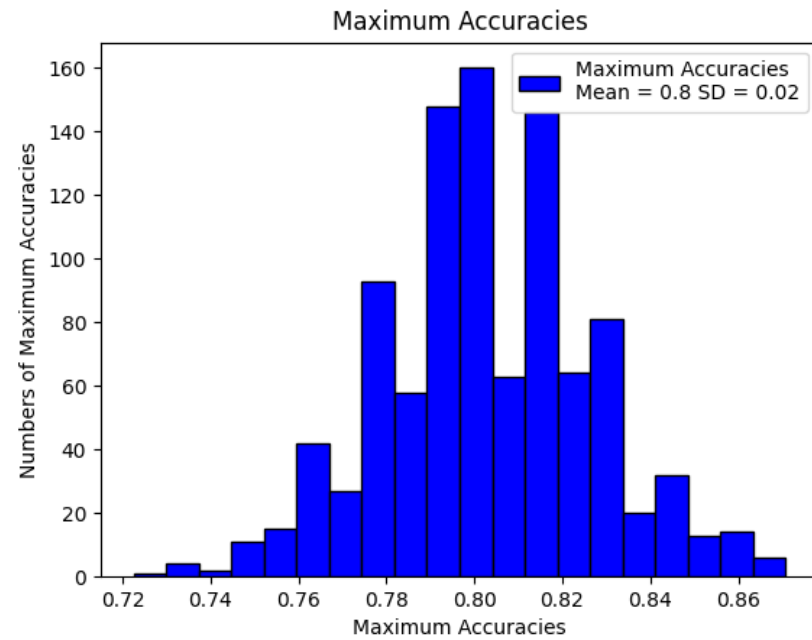
Mean accuracy = 0.784, 95% confidence interval = 0.05

Mean sensitivity = 0.705, 95% confidence interval = 0.095

Mean specificity = 0.838, 95% confidence interval = 0.067

Mean pos. pred. val. = 0.749, 95% confidence interval = 0.095

Mean AUROC = 0.839, 95% confidence interval = 0.054



```

1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)

Logistic Regression with zScaling:
Averages for all examples 1000 trials with k=0.5
Mean weight of C1 = 1.138, 95% confidence interval = 0.119
Mean weight of C2 = -0.081, 95% confidence interval = 0.098
Mean weight of C3 = -1.056, 95% confidence interval = 0.115
Mean weight of age = -0.477, 95% confidence interval = 0.086
Mean weight of male gender = -2.406, 95% confidence interval = 0.154
Mean accuracy = 0.784, 95% confidence interval = 0.05
Mean sensitivity = 0.705, 95% confidence interval = 0.095
Mean specificity = 0.838, 95% confidence interval = 0.067
Mean pos. pred. val. = 0.749, 95% confidence interval = 0.095
Mean AUROC = 0.839, 95% confidence interval = 0.054

```

第四題(iScaling)

- 對所有乘客的age進行i-scaling之後，以logistic regression進行處理，畫圖並得出各種統計量如下：

Logistic Regression with iScaling:

Averages for all examples 1000 trials with k=0.5

Mean weight of C1 = 1.071, 95% confidence interval = 0.11

Mean weight of C2 = -0.068, 95% confidence interval = 0.098

Mean weight of C3 = -1.002, 95% confidence interval = 0.114

Mean weight of age = -2.048, 95% confidence interval = 0.368

Mean weight of male gender = -2.404, 95% confidence interval = 0.156

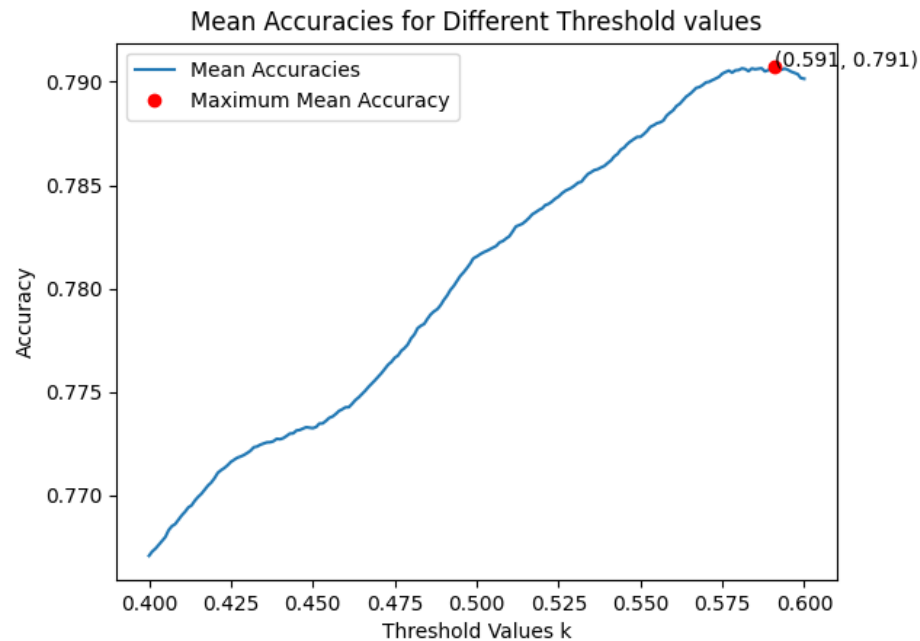
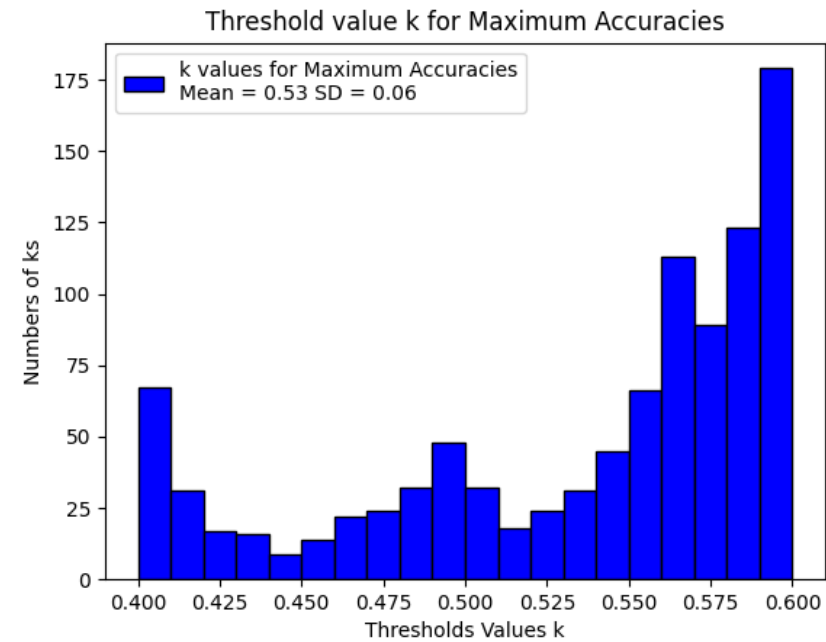
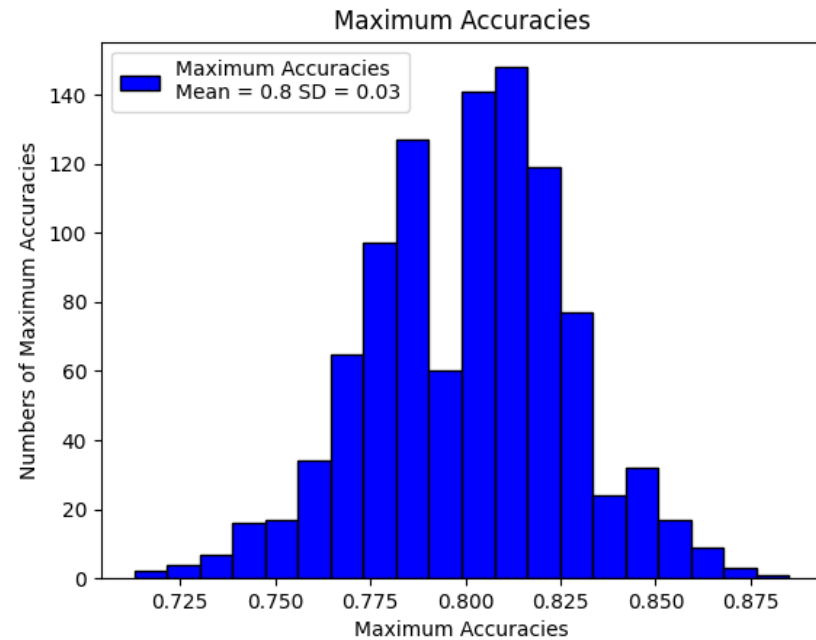
Mean accuracy = 0.782, 95% confidence interval = 0.054

Mean sensitivity = 0.699, 95% confidence interval = 0.098

Mean specificity = 0.839, 95% confidence interval = 0.066

Mean pos. pred. val. = 0.75, 95% confidence interval = 0.098

Mean AUROC = 0.836, 95% confidence interval = 0.054



```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

Logistic Regression with iScaling:

Averages for all examples 1000 trials with k=0.5

Mean weight of C1 = 1.071, 95% confidence interval = 0.11

Mean weight of C2 = -0.068, 95% confidence interval = 0.098

Mean weight of C3 = -1.002, 95% confidence interval = 0.114

Mean weight of age = -2.048, 95% confidence interval = 0.368

Mean weight of male gender = -2.404, 95% confidence interval = 0.156

Mean accuracy = 0.782, 95% confidence interval = 0.054

Mean sensitivity = 0.699, 95% confidence interval = 0.098

Mean specificity = 0.839, 95% confidence interval = 0.066

Mean pos. pred. val. = 0.75, 95% confidence interval = 0.098

Mean AUROC = 0.836, 95% confidence interval = 0.054

第五題(男性，無scaling)

- 將所有乘客分開成男性與女性個別處理，以logistic regression進行處理，畫圖並得出各種統計量如下，需特別注意的是pos.pred.val那行，因為男性乘客原本就只有少數人生還，因此model在進行預測的時候，很可能會在某幾次trial出現預測該testSet皆死亡的狀況，此時truePos與falsePos皆為0，因此會造成計算統計量時return nan:

Logistic Regression with Male and Female Separated:

Averages for Male Examples 1000 trials with k=0.5

Mean weight of C1 = 1.098, 95% confidence interval = 0.16

Mean weight of C2 = -0.53, 95% confidence interval = 0.153

Mean weight of C3 = -0.557, 95% confidence interval = 0.14

Mean weight of age = -0.047, 95% confidence interval = 0.009

Mean weight of male gender = 0.011, 95% confidence interval = 0.057

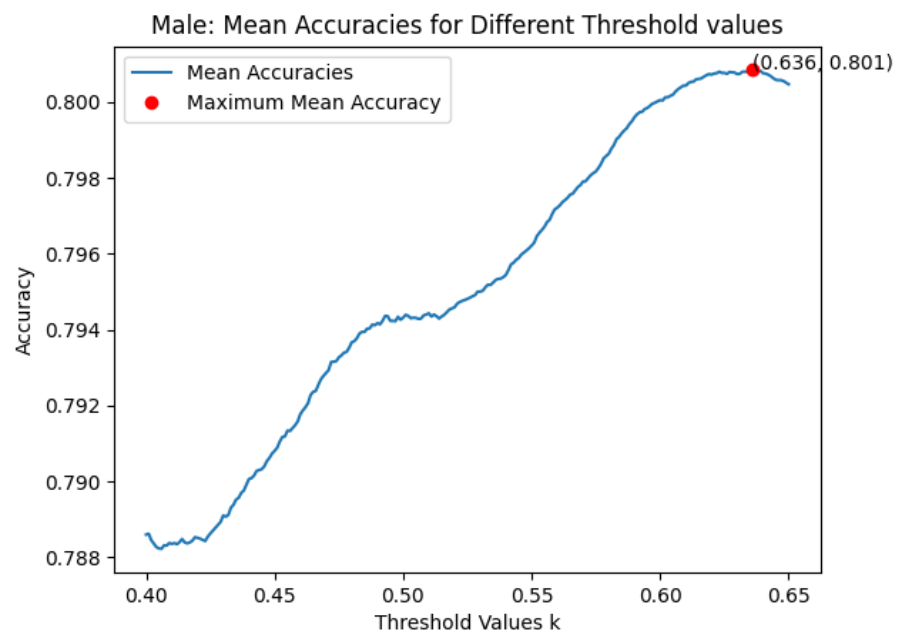
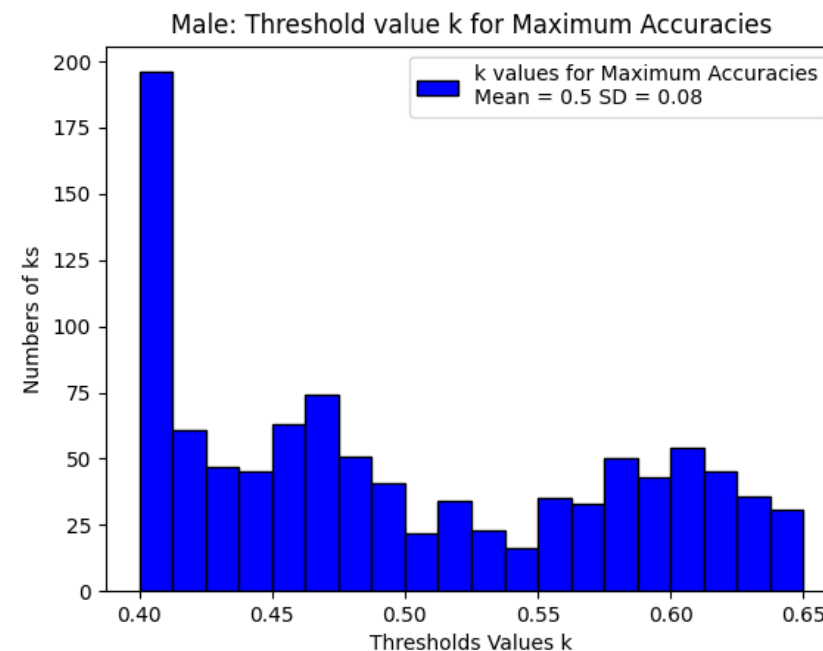
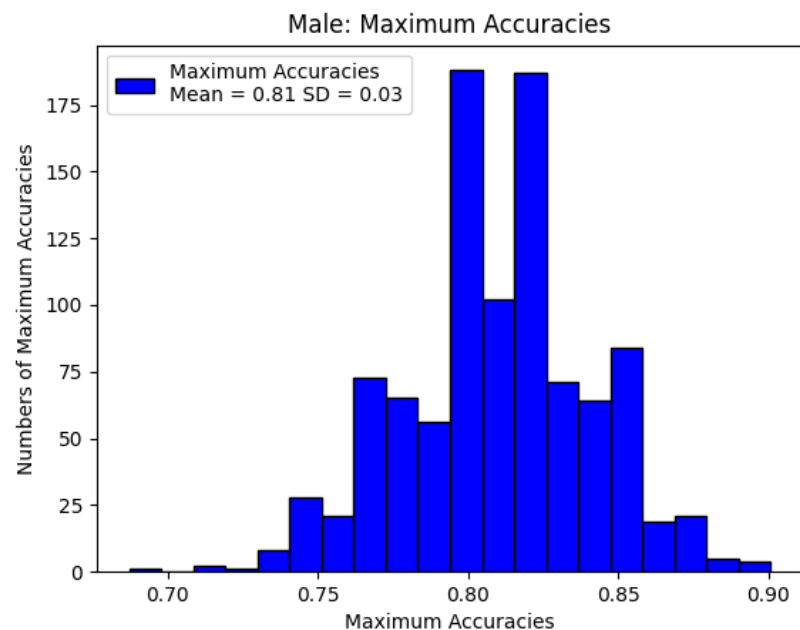
Mean accuracy = 0.794, 95% confidence interval = 0.062

Mean sensitivity = 0.082, 95% confidence interval = 0.097

Mean specificity = 0.979, 95% confidence interval = 0.032

Mean pos. pred. val. = nan, 95% confidence interval = nan

Mean AUROC = 0.688, 95% confidence interval = 0.105



```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

Logistic Regression with Male and Female Separated:
Averages for Male Examples 1000 trials with k=0.5
Mean weight of C1 = 1.098, 95% confidence interval = 0.16
Mean weight of C2 = -0.53, 95% confidence interval = 0.153
Mean weight of C3 = -0.557, 95% confidence interval = 0.14
Mean weight of age = -0.047, 95% confidence interval = 0.009
Mean weight of male gender = 0.011, 95% confidence interval = 0.057
Mean accuracy = 0.794, 95% confidence interval = 0.062
Mean sensitivity = 0.082, 95% confidence interval = 0.097
Mean specificity = 0.979, 95% confidence interval = 0.032
Mean pos. pred. val. = nan, 95% confidence interval = nan
Mean AUROC = 0.688, 95% confidence interval = 0.105

第五題(女性，無scaling)

- 將所有乘客分開成男性與女性個別處理，以logistic regression進行處理，畫圖並得出各種統計量如下，需特別注意的是male gender那行，由於這次處理的乘客全為女性，因此性別為男性所佔的權重本來就會是0了。

Logistic Regression with Male and Female Separated:

Averages for Female Examples 1000 trials with k=0.5

Mean weight of C1 = 1.414, 95% confidence interval = 0.246

Mean weight of C2 = 0.405, 95% confidence interval = 0.212

Mean weight of C3 = -1.819, 95% confidence interval = 0.193

Mean weight of age = -0.015, 95% confidence interval = 0.011

Mean weight of male gender = 0.0, 95% confidence interval = 0.0

Mean accuracy = 0.764, 95% confidence interval = 0.085

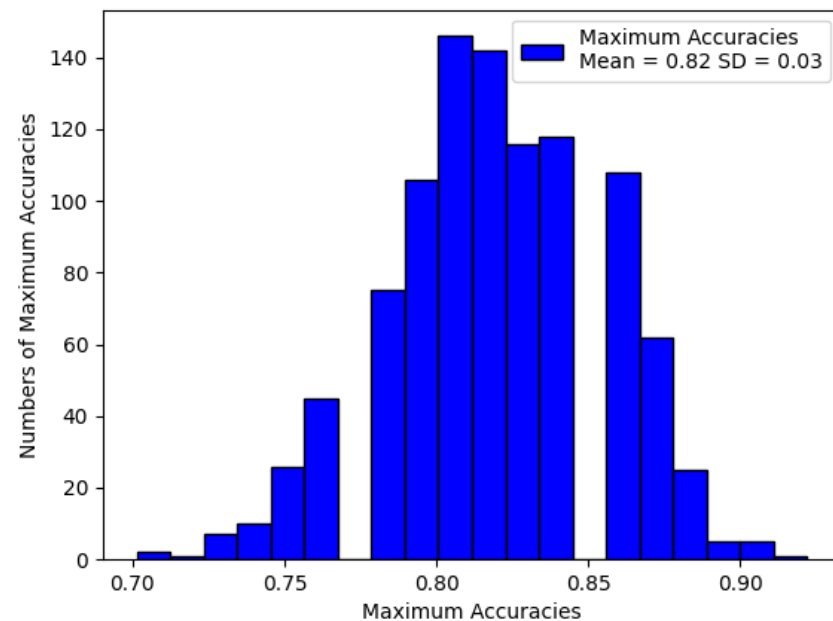
Mean sensitivity = 0.849, 95% confidence interval = 0.152

Mean specificity = 0.526, 95% confidence interval = 0.342

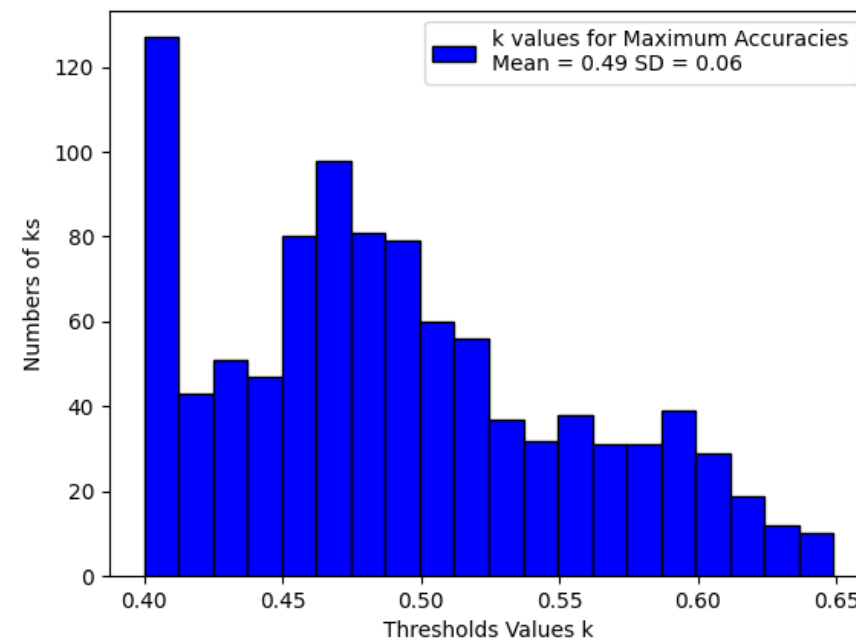
Mean pos. pred. val. = 0.847, 95% confidence interval = 0.121

Mean AUROC = 0.827, 95% confidence interval = 0.095

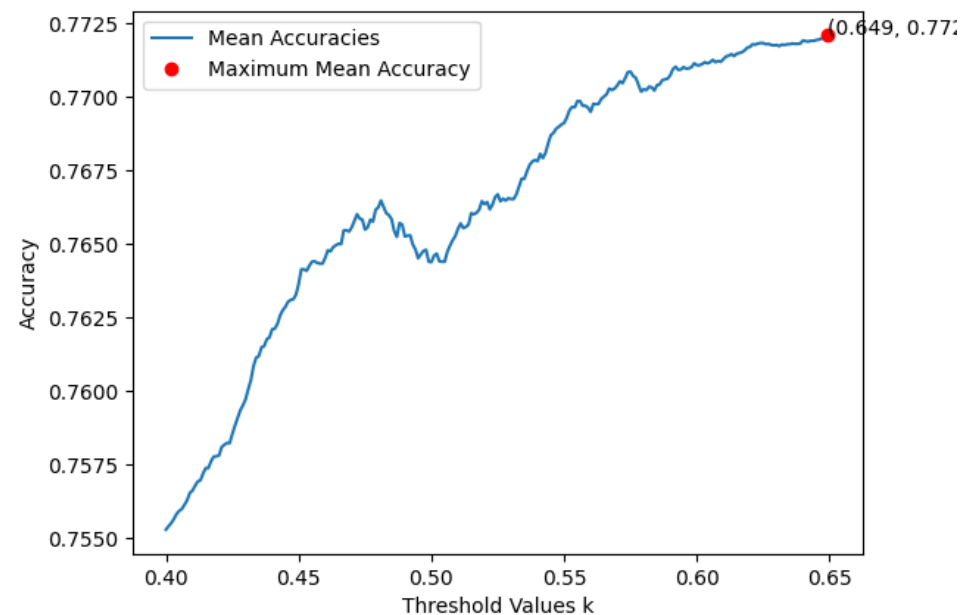
Female: Maximum Accuracies



Female: Threshold value k for Maximum Accuracies



Female: Mean Accuracies for Different Threshold values



```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

```
Logistic Regression with Male and Female Separated:
Averages for Female Examples 1000 trials with k=0.5
Mean weight of C1 = 1.414, 95% confidence interval = 0.246
Mean weight of C2 = 0.405, 95% confidence interval = 0.212
Mean weight of C3 = -1.819, 95% confidence interval = 0.193
Mean weight of age = -0.015, 95% confidence interval = 0.011
Mean weight of male gender = 0.0, 95% confidence interval = 0.0
Mean accuracy = 0.764, 95% confidence interval = 0.085
Mean sensitivity = 0.849, 95% confidence interval = 0.152
Mean specificity = 0.526, 95% confidence interval = 0.342
Mean pos. pred. val. = 0.847, 95% confidence interval = 0.121
Mean AUROC = 0.827, 95% confidence interval = 0.095
```

第五題(男性，z-scaling)

- 將所有乘客分開成男性與女性個別處理，先對age進行z-scaling，再以logistic regression進行處理，畫圖並得出各種統計量如下：

Logistic Regression with Male and Female Separated with zScaling:

Averages for Male Examples 1000 trials with k=0.5

Mean weight of C1 = 1.107, 95% confidence interval = 0.155

Mean weight of C2 = -0.537, 95% confidence interval = 0.149

Mean weight of C3 = -0.558, 95% confidence interval = 0.139

Mean weight of age = -0.047, 95% confidence interval = 0.009

Mean weight of male gender = 0.012, 95% confidence interval = 0.056

Mean accuracy = 0.726, 95% confidence interval = 0.074

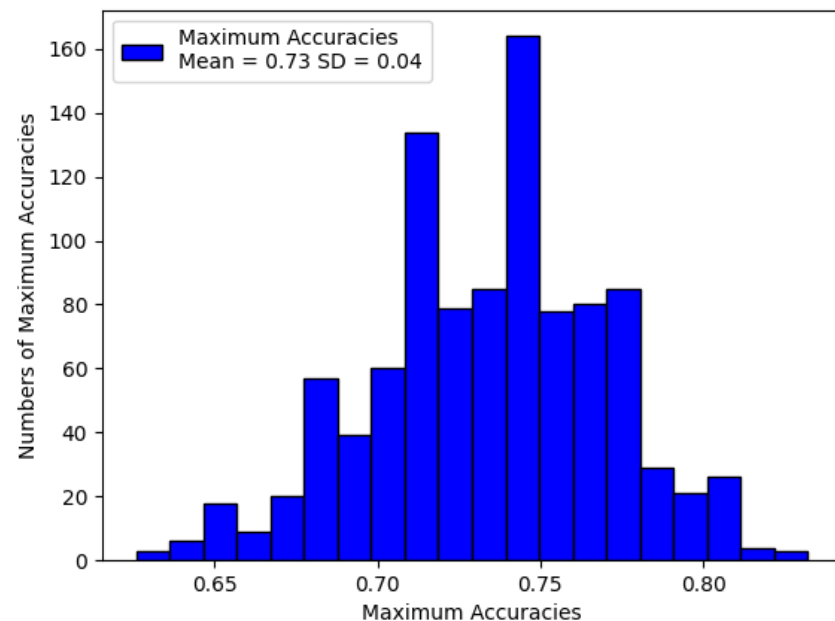
Mean sensitivity = 0.391, 95% confidence interval = 0.171

Mean specificity = 0.811, 95% confidence interval = 0.081

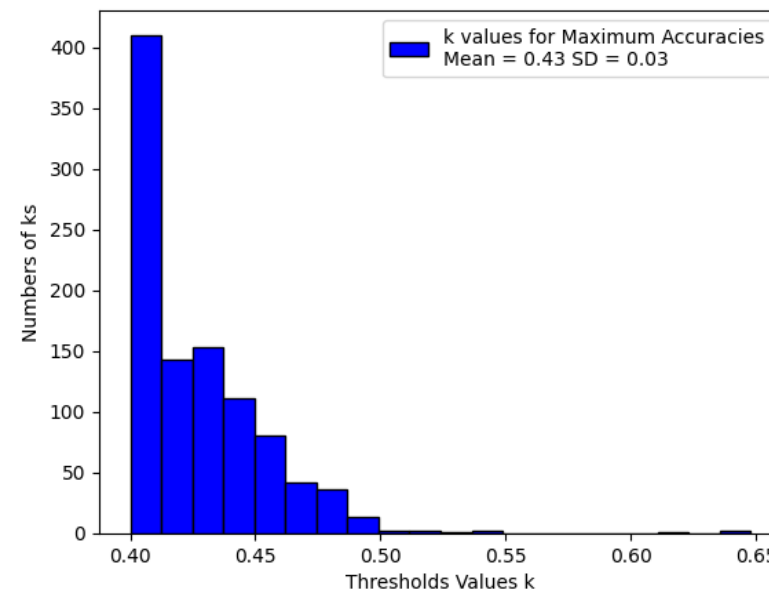
Mean pos. pred. val. = 0.348, 95% confidence interval = 0.158

Mean AUROC = 0.639, 95% confidence interval = 0.117

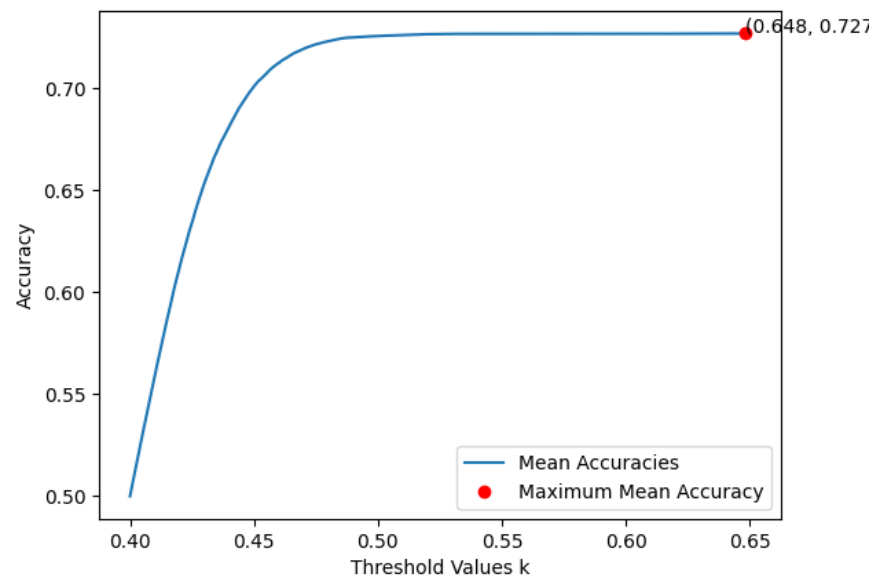
Male: Maximum Accuracies



Male: Threshold value k for Maximum Accuracies



Male: Mean Accuracies for Different Threshold values



```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

Logistic Regression with Male and Female Separated with zScaling:
Averages for Male Examples 1000 trials with k=0.5
Mean weight of C1 = 1.107, 95% confidence interval = 0.155
Mean weight of C2 = -0.537, 95% confidence interval = 0.149
Mean weight of C3 = -0.558, 95% confidence interval = 0.139
Mean weight of age = -0.047, 95% confidence interval = 0.009
Mean weight of male gender = 0.012, 95% confidence interval = 0.056
Mean accuracy = 0.726, 95% confidence interval = 0.074
Mean sensitivity = 0.391, 95% confidence interval = 0.171
Mean specificity = 0.811, 95% confidence interval = 0.081
Mean pos. pred. val. = 0.348, 95% confidence interval = 0.158
Mean AUROC = 0.639, 95% confidence interval = 0.117

第五題(女性，z-scaling)

- 將所有乘客分開成男性與女性個別處理，先對age進行z-scaling，再以logistic regression進行處理，畫圖並得出各種統計量如下：

Logistic Regression with Male and Female Separated with zScaling:

Averages for Female Examples 1000 trials with k=0.5

Mean weight of C1 = 1.415, 95% confidence interval = 0.257

Mean weight of C2 = 0.408, 95% confidence interval = 0.221

Mean weight of C3 = -1.823, 95% confidence interval = 0.197

Mean weight of age = -0.016, 95% confidence interval = 0.011

Mean weight of male gender = 0.0, 95% confidence interval = 0.0

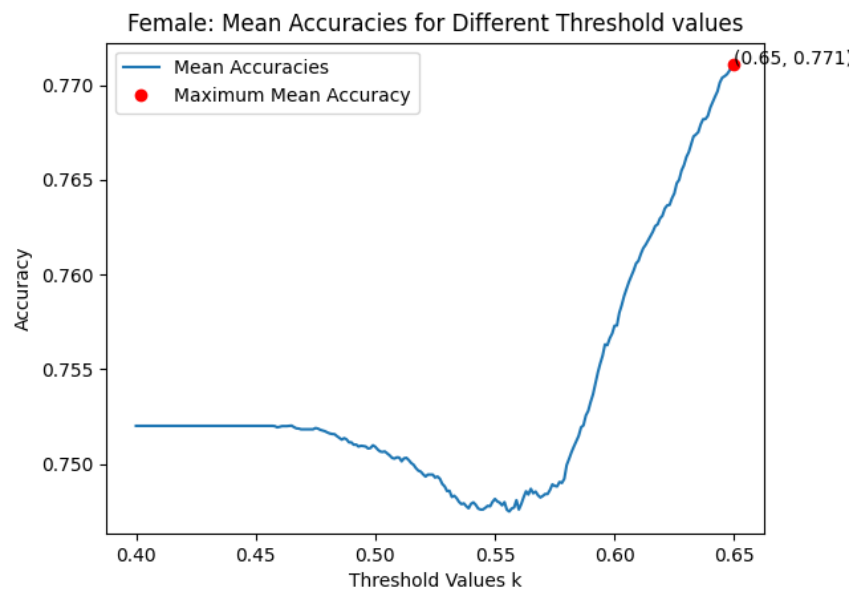
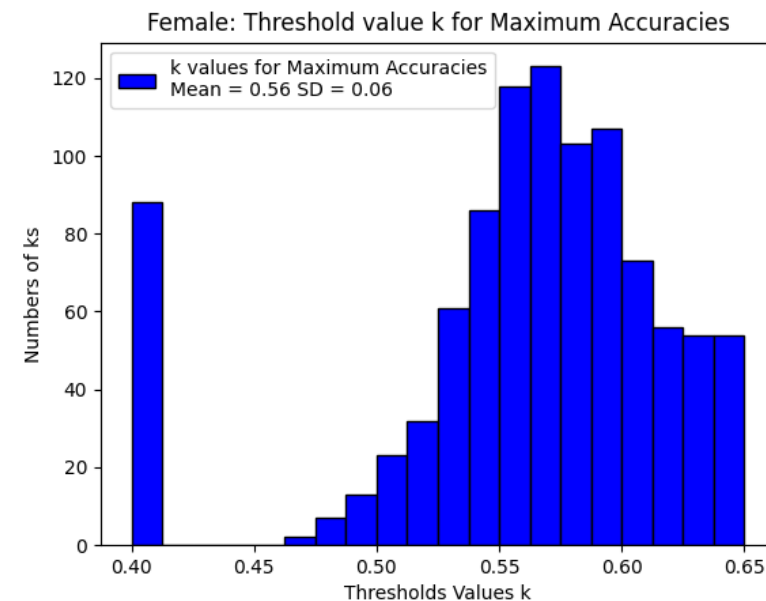
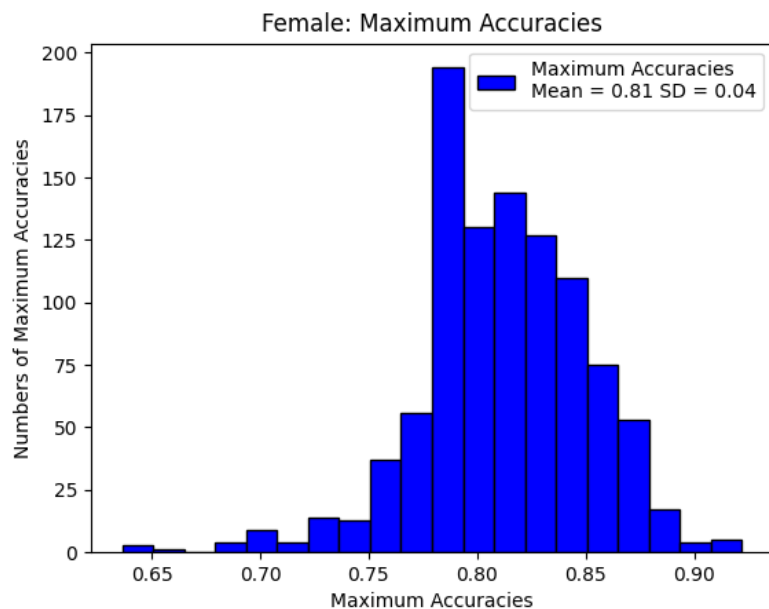
Mean accuracy = 0.751, 95% confidence interval = 0.088

Mean sensitivity = 0.994, 95% confidence interval = 0.082

Mean specificity = 0.017, 95% confidence interval = 0.213

Mean pos. pred. val. = 0.755, 95% confidence interval = 0.098

Mean AUROC = 0.816, 95% confidence interval = 0.093



```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

Logistic Regression with Male and Female Separated with zScaling:
Averages for Female Examples 1000 trials with k=0.5
Mean weight of C1 = 1.415, 95% confidence interval = 0.257
Mean weight of C2 = 0.408, 95% confidence interval = 0.221
Mean weight of C3 = -1.823, 95% confidence interval = 0.197
Mean weight of age = -0.016, 95% confidence interval = 0.011
Mean weight of male gender = 0.0, 95% confidence interval = 0.0
Mean accuracy = 0.751, 95% confidence interval = 0.088
Mean sensitivity = 0.994, 95% confidence interval = 0.082
Mean specificity = 0.017, 95% confidence interval = 0.213
Mean pos. pred. val. = 0.755, 95% confidence interval = 0.098
Mean AUROC = 0.816, 95% confidence interval = 0.093

第五題(男性，i-scaling)

- 將所有乘客分開成男性與女性個別處理，先對age進行i-scaling，再以logistic regression進行處理，畫圖並得出各種統計量如下:

Logistic Regression with Male and Female Separated with iScaling:

Averages for Male Examples 1000 trials with k=0.5

Mean weight of C1 = 1.101, 95% confidence interval = 0.161

Mean weight of C2 = -0.534, 95% confidence interval = 0.152

Mean weight of C3 = -0.554, 95% confidence interval = 0.14

Mean weight of age = -0.047, 95% confidence interval = 0.008

Mean weight of male gender = 0.012, 95% confidence interval = 0.056

Mean accuracy = 0.726, 95% confidence interval = 0.07

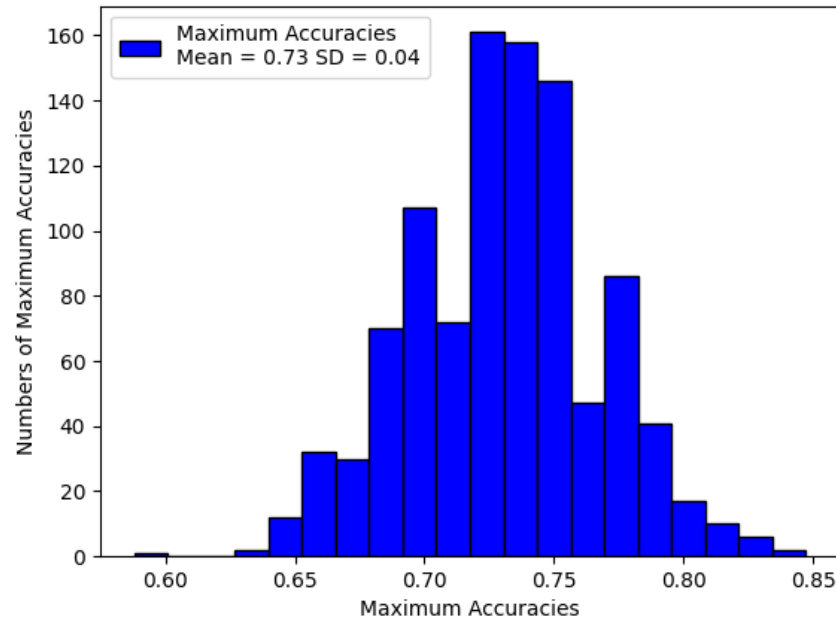
Mean sensitivity = 0.395, 95% confidence interval = 0.173

Mean specificity = 0.811, 95% confidence interval = 0.068

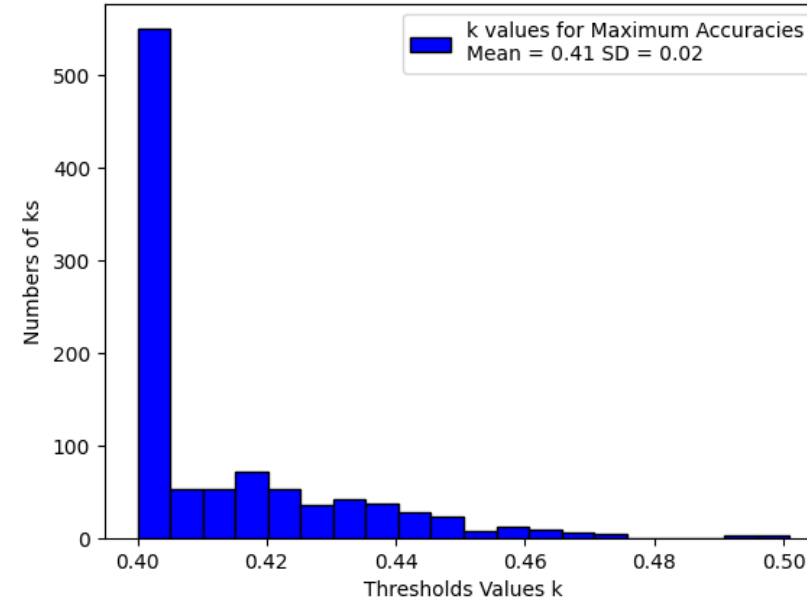
Mean pos. pred. val. = 0.35, 95% confidence interval = 0.158

Mean AUROC = 0.598, 95% confidence interval = 0.109

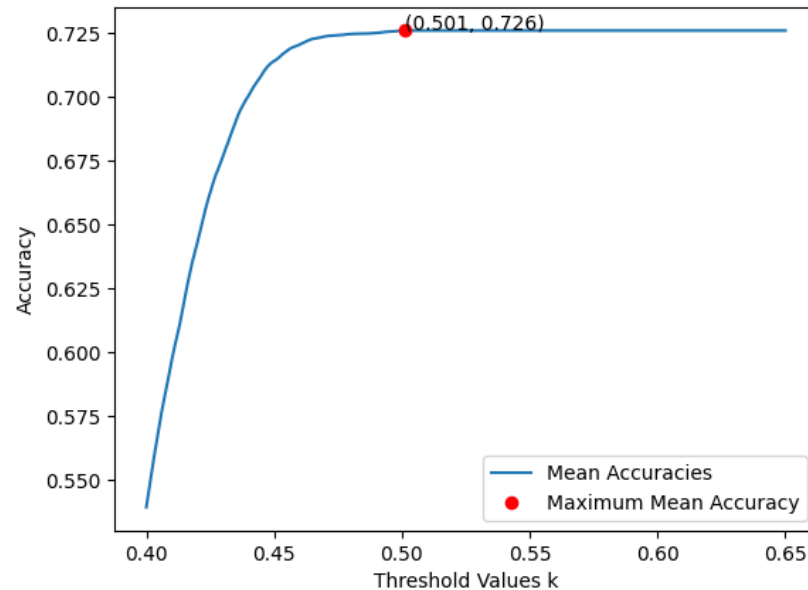
Male: Maximum Accuracies



Male: Threshold value k for Maximum Accuracies



Male: Mean Accuracies for Different Threshold values



```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

Logistic Regression with Male and Female Separated with iScaling:
Averages for Male Examples 1000 trials with k=0.5
Mean weight of C1 = 1.101, 95% confidence interval = 0.161
Mean weight of C2 = -0.534, 95% confidence interval = 0.152
Mean weight of C3 = -0.554, 95% confidence interval = 0.14
Mean weight of age = -0.047, 95% confidence interval = 0.008
Mean weight of male gender = 0.012, 95% confidence interval = 0.056
Mean accuracy = 0.726, 95% confidence interval = 0.07
Mean sensitivity = 0.395, 95% confidence interval = 0.173
Mean specificity = 0.811, 95% confidence interval = 0.068
Mean pos. pred. val. = 0.35, 95% confidence interval = 0.158
Mean AUROC = 0.598, 95% confidence interval = 0.109

第五題(女性，i-scaling)

- 將所有乘客分開成男性與女性個別處理，先對age進行i-scaling，再以logistic regression進行處理，畫圖並得出各種統計量如下:

Logistic Regression with Male and Female Separated with iScaling:

Averages for Female Examples 1000 trials with k=0.5

Mean weight of C1 = 1.414, 95% confidence interval = 0.264

Mean weight of C2 = 0.406, 95% confidence interval = 0.222

Mean weight of C3 = -1.819, 95% confidence interval = 0.192

Mean weight of age = -0.015, 95% confidence interval = 0.011

Mean weight of male gender = 0.0, 95% confidence interval = 0.0

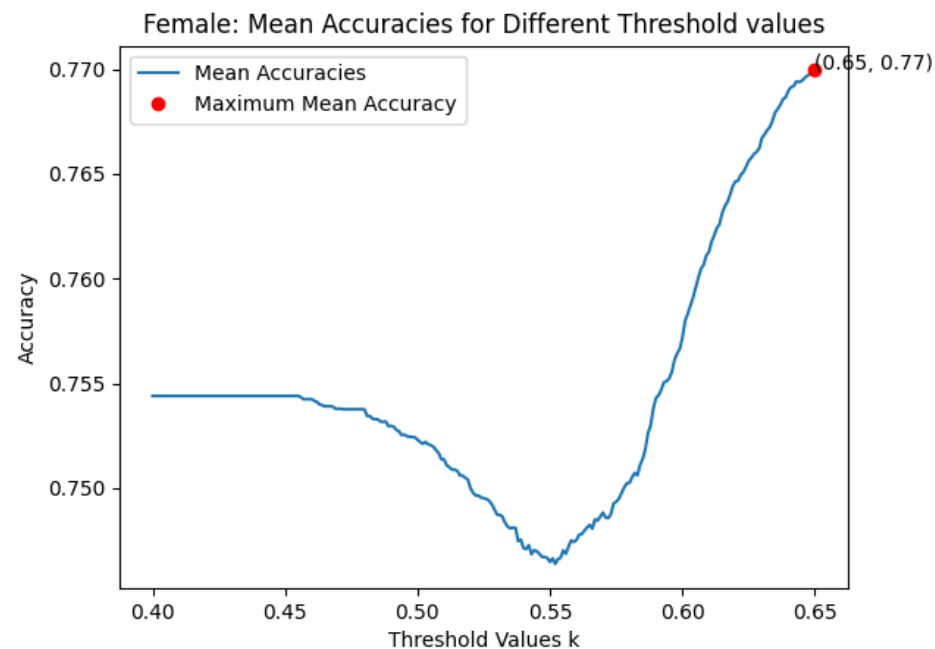
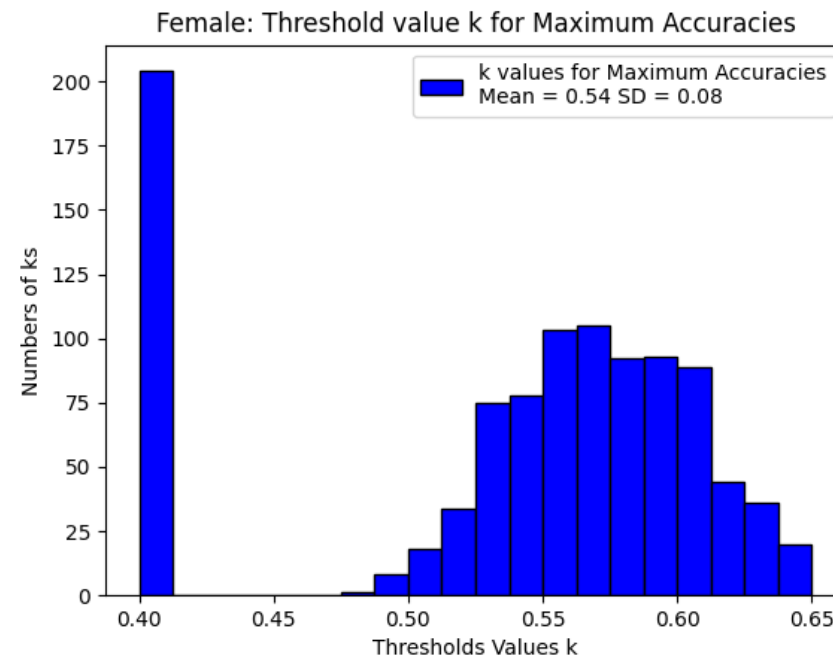
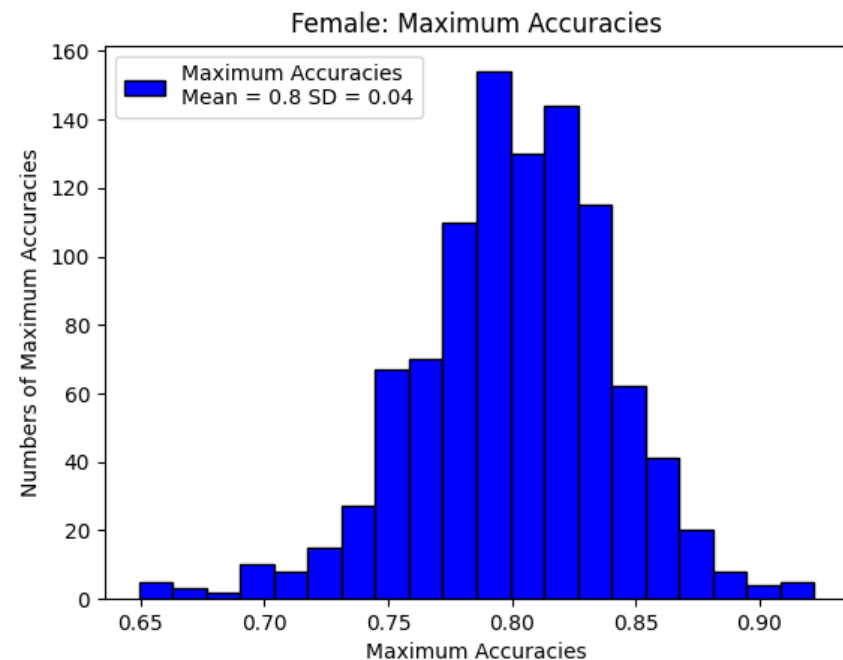
Mean accuracy = 0.752, 95% confidence interval = 0.086

Mean sensitivity = 0.99, 95% confidence interval = 0.106

Mean specificity = 0.025, 95% confidence interval = 0.272

Mean pos. pred. val. = 0.759, 95% confidence interval = 0.102

Mean AUROC = 0.811, 95% confidence interval = 0.089



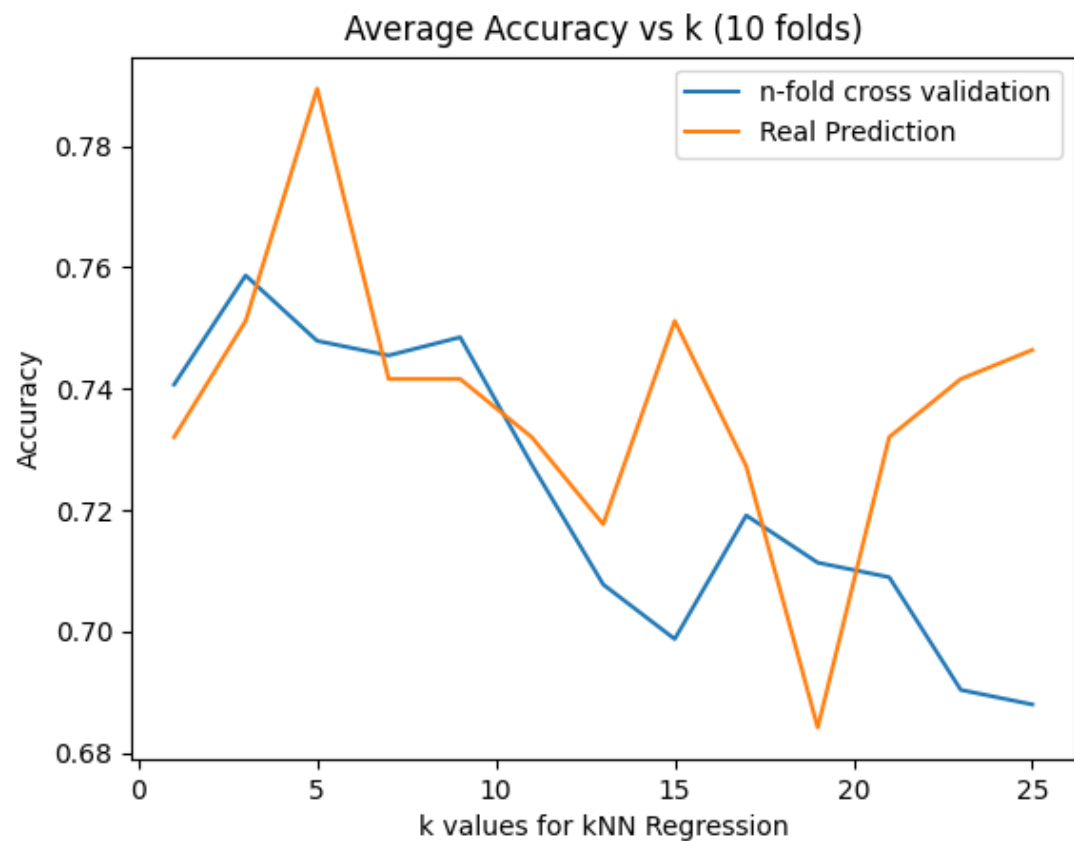
```
1000 trials... (100/1000)
1000 trials... (200/1000)
1000 trials... (300/1000)
1000 trials... (400/1000)
1000 trials... (500/1000)
1000 trials... (600/1000)
1000 trials... (700/1000)
1000 trials... (800/1000)
1000 trials... (900/1000)
1000 trials... (1000/1000)
```

Logistic Regression with Male and Female Separated with iScaling:
Averages for Female Examples 1000 trials with k=0.5

Mean weight of C1 = 1.414, 95% confidence interval = 0.264
Mean weight of C2 = 0.406, 95% confidence interval = 0.222
Mean weight of C3 = -1.819, 95% confidence interval = 0.192
Mean weight of age = -0.015, 95% confidence interval = 0.011
Mean weight of male gender = 0.0, 95% confidence interval = 0.0
Mean accuracy = 0.752, 95% confidence interval = 0.086
Mean sensitivity = 0.99, 95% confidence interval = 0.106
Mean specificity = 0.025, 95% confidence interval = 0.272
Mean pos. pred. val. = 0.759, 95% confidence interval = 0.102
Mean AUROC = 0.811, 95% confidence interval = 0.089

第六題

- 先以 $k=3$ 進行KNN，接著再使用n-fold cross validation找出能使accuracy最佳的 k ，印出各種統計量並且畫圖
- 我在寫這題的時候，用n-fold cross validation找最佳的 k 時，出現 $k=5$ 的機率最高，也有出現 $k=3,7,9$ 的，這裡就以某次執行時出現的最佳的 $k=3$ 為例。
- 在繪製real prediction與n-fold cross validation之不同 k 對應到的accuracy時，我發現n-fold cross validation在 $k=9$ 以上之後幾乎都會呈現accuracy隨之下降的狀況，但real prediction卻有時候會這樣，但有時候反而會在 $k=23,25$ 左右時，accuracy又回升的狀況。



k-NN Prediction for Survive with k=3:

TP,FP,TN,FN = 56 21 108 24

| | | |
|----------------------|-----|----|
| | TP | FP |
| Confusion Matrix is: | 56 | 21 |
| | 108 | 24 |
| | TN | FN |

Accuracy = 0.785

Sensitivity = 0.7

Specificity = 0.837

Pos. Pred. Val. = 0.727

Using n-fold cross validation to find proper k for k-NN Prediction

K for Maximum Accuracy is: 3

TP,FP,TN,FN = 46 29 106 28

| | | |
|----------------------|-----|----|
| | TP | FP |
| Confusion Matrix is: | 46 | 29 |
| | 106 | 28 |
| | TN | FN |

Accuracy = 0.727

Sensitivity = 0.622

Specificity = 0.785

Pos. Pred. Val. = 0.613

Predictions with maximum accuracy k: 3

Cross Validation Accuracies is: 0.7586826347305389

Predicted Accuracies is: 0.7272727272727273

第七題

- 將乘客分成男性與女性個別去以 $K=3$ 做KNN，得出各種統計量，並將TP,FP,TN,FN合併之後去跑confusion matrix

Try to predict male and female separately and combined with k=3:

For Male:

TP,FP,TN,FN = 7 18 91 15

| | | |
|----------------------|----|----|
| | TP | FP |
| Confusion Matrix is: | 7 | 18 |
| | 91 | 15 |
| | TN | FN |

Accuracy = 0.748

Sensitivity = 0.318

Specificity = 0.835

Pos. Pred. Val. = 0.28

For Female:

TP,FP,TN,FN = 50 14 9 4

| | | |
|----------------------|----|----|
| | TP | FP |
| Confusion Matrix is: | 50 | 14 |
| | 9 | 4 |
| | TN | FN |

Accuracy = 0.766

Sensitivity = 0.926

Specificity = 0.391

Pos. Pred. Val. = 0.781

Combined Predictions Statistics:

TP,FP,TN,FN = 57 32 100 19

| | | |
|----------------------|-----|----|
| | TP | FP |
| Confusion Matrix is: | 57 | 32 |
| | 100 | 19 |
| | TN | FN |

Accuracy = 0.755

Sensitivity = 0.75

Specificity = 0.758

Pos. Pred. Val. = 0.64

學期修課的心得與收穫

在修教授開的這門計算思維與問題解決之前，我曾經修過的相關的課程也有資料結構、演算法、機率統計、資料科學導論、機器學習等，而這些所學的東西都在這學期的這門課被統整在一起並且以實務帶動理論的方式讓我對他們更加熟悉。這門課涵蓋了演算法中的最佳化問題求解、圖論；統計學中的隨機、中央極限定理、蒙地卡羅、信賴區間、抽樣等重要概念；以及在機器學習中的分群與迴歸這兩大重要的議題。我覺得老師在每堂課中所附的範例程式碼以由簡單至深入的方式呈現，令我對這些理論概念是如何以程式碼實作的，有了更清晰的認識。

我認為除了理論知識與實作技巧外，我在老師這堂課上學到的兩個最重要的東西，就是「**模組化的程式寫法**」以及「**資料視覺化之分析**」。雖然我在其他課也學過C++等OOP相關的課，但其實也並不常在我的程式中加入class,object等元素，直到我修了這門課後，我發現老師的範例程式碼有很大一部分都是基於class與自定義function等方式來完成的，我在修課初期有些不適應這種程式碼的架構，時常會花非常多時間來研究老師的程式碼才能看懂，但是在做最後三四份作業(線性迴歸、kmeans、knn、final_project)時，我因為熟悉了老師的程式碼，就深切地體會到了這種模組化的程式寫法的強大之處，有許多功能，比如findK,getLabel,getStats,yscaleFeatures 等，只要將class,function定義好，便可很輕鬆地對程式進行改動與擴充，該份程式就能更廣泛地應用在不同的dataset上。而資料視覺化則是因為老師在作業中時常會讓我們以pylab,matplotlib繪圖，在後來幾份作業中，我深刻地體會到了這種方式的厲害之處，有時只要看圖畫出來的異常點，很可能就可以大概推測程式中哪邊的寫法有問題需要改善了，且將資料視覺化後，也能以更加直觀的方式來讓我們對於程式執行結果進行分析。

最後要說的是，很感謝教授這學期開了這門課，雖然我並不是工科系的學生，但我很慶幸這學期有加選到這門課，我在這門課收穫了非常多的東西，很開心這學期能在教授的課堂上學習與成長！