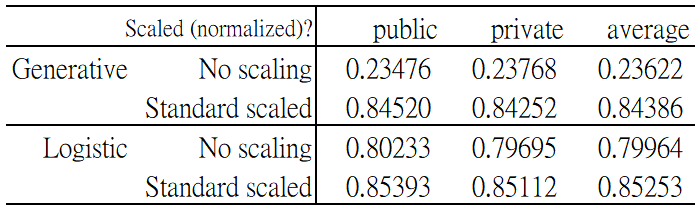
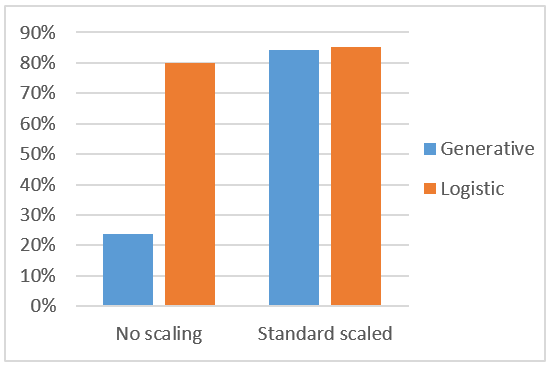
學號：B04901060 系級： 電機三 姓名：黃文璁

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

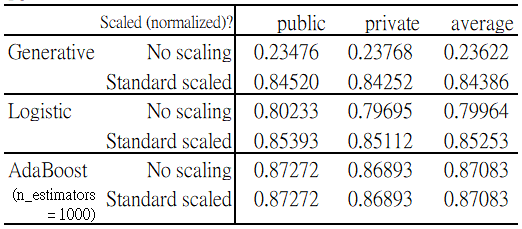
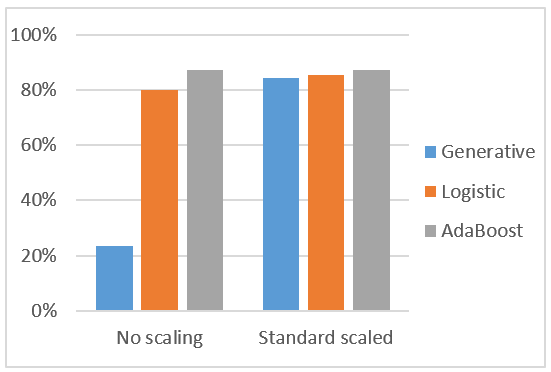
　　在沒有Standard scale (normalize) 的情況下，logistic的準確度遠高於generative model。

　　在有Standard scale (normalize) 的情況下，logistic和generative model的準確度較接近，但logistic還是稍微好一些。故**整體來說，logistic regression的準確度較好**

**2.請說明你實作的best model，其訓練方式和準確率為何？**

答：

**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。**

　　Generative model在沒有normalize的情況下準確度非常差，但normalize後準確度就上升到和logistic接近的水準。整體來說**generative model和logistic regression在normalize後準確度都有提升。**

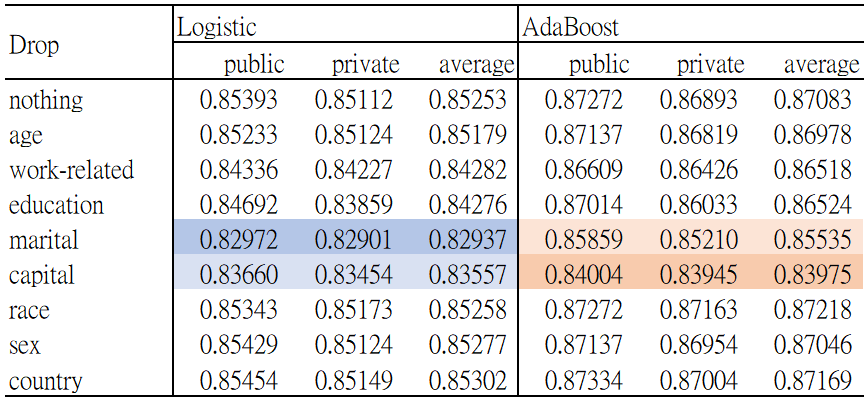
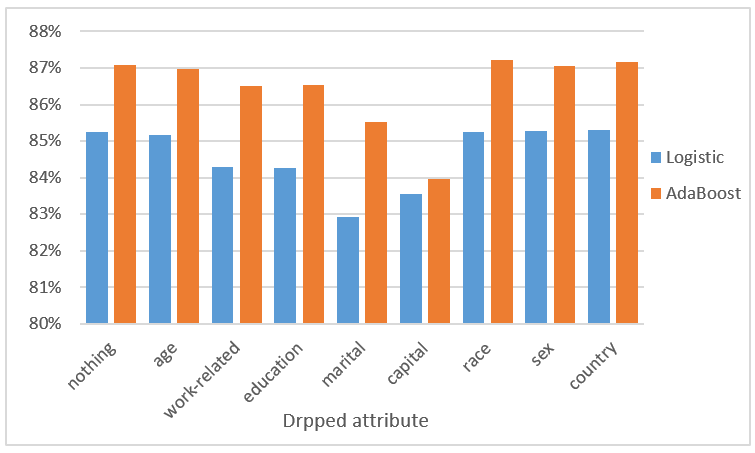
　　而本次在best model中使用的**AdaBoost則不受normalize影響**，這是由於scikit-learn中的AdaBoostClassifier預設使用的estimator是DecisionTreeClassifier，而decision tree較不受normalize影響結果。

**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：

**5.請討論你認為哪個attribute對結果影響最大？**

分別利用logistic和AdaBoost兩個模型，每次drop一個attribute，測試Kaggle分數：

　　兩種模型中attribute drop之後影響最大的都為marital和capital gain/loss，故可以歸納出：

**marital**和**capital gain/loss**影響預測結果最大。