土木四 b03501032 羅弘

1. 此次在training的部分，首先將資料用pandas讀進後，針對非01項做hot-encode且對用robustscaler()將資料normalize，使得預測比較準確，之後便是先測試過Naive\_Bayes、DecisionTreeClassifier、linear regression但是測試後發現準確率和要求差異過大，所以最後使用catboostclassifier作為training model，在使用後，就達到兩個line的需求了
2. 基本上給的參數中比較重要的一點是，需要將一些原本並非01表示的數據變成01表示，像是education、婚姻狀態，都可以藉由將1~4分別對應的類別，改變成為多項的yes or no 的是非題，而基本上將這些參數改變成這樣的表達方式後，基本上參數都是被我所使用的。
3. 此次feature中，個人認為最重要的便是pay bill 和limit\_bal需要做前處裡，我們可以知道一個人的pay and bill 是和他的limit有絕對關係，所以我認為在這邊需要消除limitbal和pay、bill之間的關係，所以便決定pay&bill都要除以limit\_bal(舉例來說15萬上限的人不會借到20萬)所以其實一個人pay and bill是會被Limit\_bal做限制。在直觀上來說，原本沒有對這邊做處理時，準確率只有0.77(未達simple line)但加入後便上升到0.79(兩條線都過了)
4. 此次使用python 3.6和catboost的原因，第一python2.7在win10上面安裝時會有ascii的error所以上網查了以後發現，3.6板就沒這個問題，所以就使用了。  
   第二catboost是一個頗新的opensource 的machine learning method裡面，是用gradinet 提升的方式來做學習，那雖然他和sklearn 的幾種學習方式做比較後，花的時間多蠻多的，但是準確率(accuracy)上升很多，那選用的原因，也是因為在前面五次的嘗試中，發現諸如Naive\_Bayes、DecisionTreeClassifier、linear regression表現真的不好，所以在多方嘗試下最後才選擇使用catboost