Introduction to Artificial Intelligence and Machine Learning
Instructor: Tian-Li Yu **B03901133 Gary Shih**
Final Project Documentation **December 11, 2017**

Problem 1

簡述此次作業主要使用的演算法，並說明為何使用此演算法

Basically, the algorithm I use in this homework is **gradient boosting on decision trees**. The reason why I use this algorithm is that it shows state-of-the-art results on classification problems.

Problem 2

呈上題，簡述此演算法比較重要的參數，並說明如何選取適合的參數值

There are several important parameters in this kind of algorithm such as learning rate, loss function, and all parameters related to decision trees. Basically, all of them can be tuned based on a held-out set seperated from the training data.

Problem 3

簡述對資料的輸入特徵 (features) **的處理方式**

- LIMIT_BAL: scalar
- SEX: scalar
- EDUCATION: one-hot vector $\in \mathbb{R}^7$
- MARRIAGE: one-hot vector $\in \mathbb{R}^4$
- AGE: scalar
- PAY_[$1 \sim 6$]: one-hot vector $\in \mathbb{R}^{11}$
- BILL_AMT[$1 \sim 6$]: scalar
- PAY_AMY[$1 \sim 6$]: scalar
- Y: scalar

Problem 4

若程式中有使用較特殊的套件，請敘述其名稱及版本，並簡述為何使用此套件

catboost 0.3.1

The reason why I choose to use this framework is that it may be the best implementation or variation of gradient boosting decision tree algorithm now. On its website, it shows better results on several benchmarks compared to another popular framework, XGBoost.