

FinTech Final Project Report

學號：R07922142 系級：資工所碩一 姓名：歐政鷹

在這次的 Final Project 中，我利用了課堂中學到的兩種指標：均線(Moving Average, MA)及 RSI(Relative strength indicator) 並利用 Dynamic Programming 及 logistic regression 模型，來作為我的預測模型中 Baseline 建立的方法。

我利用 Dynamic Programming 對 Daily OHLCV 中的 Close price 進行分析，找出現有資料中每天的最佳 action，然後再利用每天當天過去 14 天的資料，透過 logistic regression 來預測出當天的 action。由於 Dynamic Programming 是會用到未來的資料來作預測，因此以 logistic regression 所做出來的結果當然與 Dynamic Programming 的結果有著一定的差距，圖 1 顯示出 logistic regression 預測結果與 baseline 作比較的情況。所以我以以上的結果作為 baseline，希望能透過改善及優化 logistic regression model 來得到更好的預測結果。

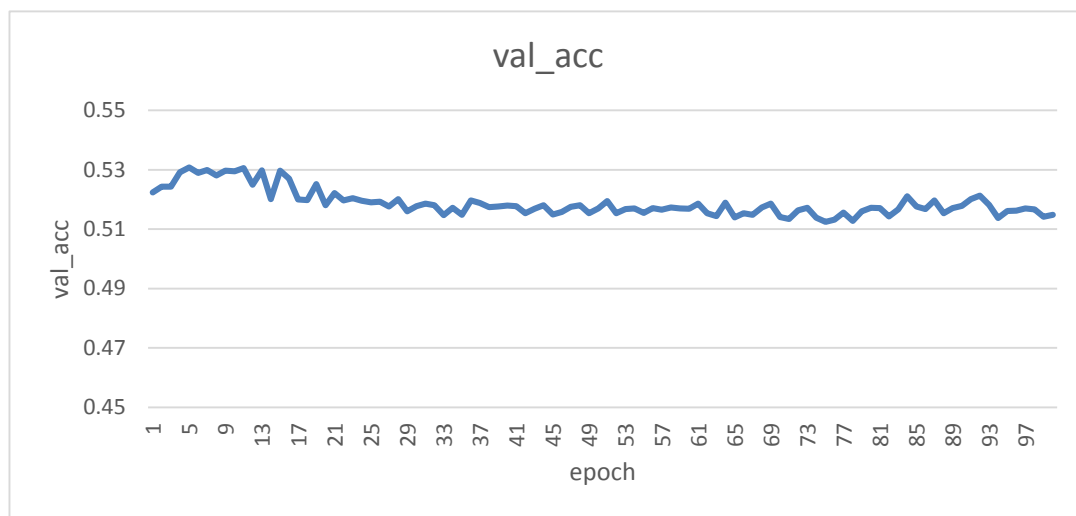


圖 1. logistic regression 訓練曲線

在 logistic regression 的優化方面，我先以加入更多 Feature 作考慮，一方面可以觀察加入不同的 feature 是否會有更好的結果，另一方面可以讓之後在作 feature selection 時有更多的選擇。

我希望可以加入一些具判別性的 feature，因此我以 MA 及 RSI 作為我的選擇。我先以所有資料為 MA 及 RSI 找出合適的長短期 Window size，alpha 及 beta 值，在尋找的過程中，為了避免出現 overfitting 的情況，所以回報率的多寡並不是尋找的目標，而是志在尋找出具有回報的參數組合。

在尋找出合適的參數組合後，我把其 MA 及 RSI 結果加入到 feature 中，再利用 logistic regression 作訓練。在得到初步的訓練結果後，再作 features selection 以找出最佳的 features 組合為目標。而在得到最佳的 features 組合後，我再透過 ensemble，一方面可以提升 model 預測準確度，一方面可以減少單一 model 出現 overfitting 的情況。在 ensemble 的方法裡，不同的 model 是透過 MA 及 RSI 不同的參數組合訓練而成，不同的參數組合都是以可以得到正回報的前提下找出，因此我認為透過 ensemble 的方式可以讓 model 有更好的結果，圖 2 顯示出在經過 features selection 及 ensemble 後，logistic regression 預測結果與 baseline 作比較的情況，可明顯看出結果有顯著的改善。

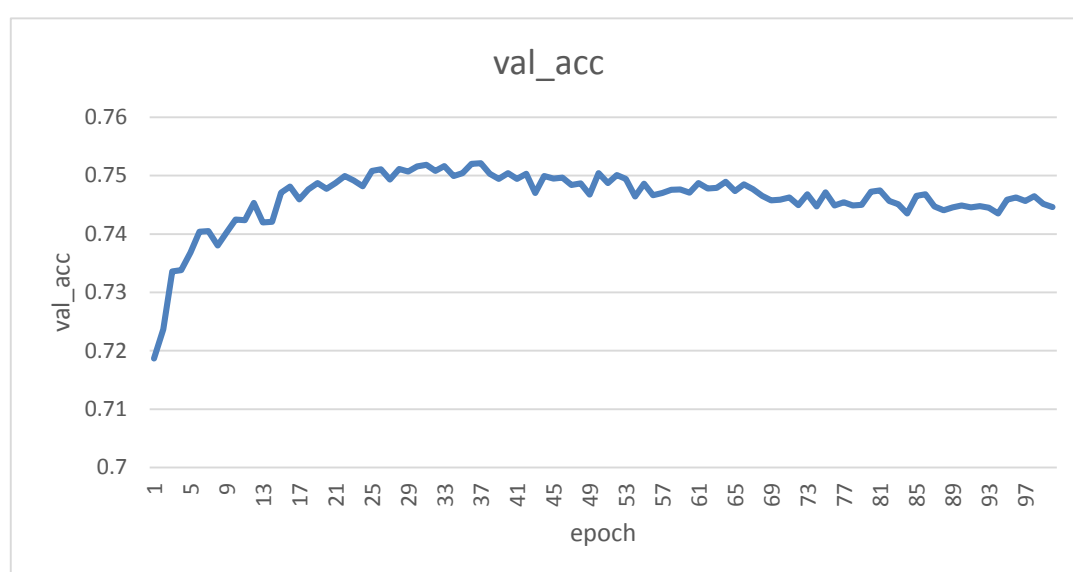


圖 2. logistic regression 訓練曲線

最後，我利用上述 mode 來作為我 final project 的結果，希望能取得好的結果。

在我的 model 中，我是透過每天的 action 作為 model 好壞的比較依據，而沒有考慮到我的 action 最終得到的回報率的好壞。雖然不確定以考慮回報率的方式是否會有更好的結果，但我認為這是將來 model 要繼續改善時一個可考慮的方向，說不定可以讓 model 得到更好的結果。