

運用財務報表特徵面向與技術指標預測報酬率

計量財務期末報告
財精四 C 07155328 白祐瑄

一、摘要

使用 0050.TW 成分股財務報表不同特徵面向與個股技術指標做為因子。並將所有因子進行有效性因子篩選法後，剩餘因子做為輸入機器學習之特徵。藉由機器學習預測個股未來一季之報酬。此份報告作法參考台灣管理學刊中文章。

二、資料來源與數據因子說明

財務報表因子：通過爬蟲方式抓取 Goodinfo 網站上個股資料。不同特徵面向因子使用如下表（一）。

表（一）

收益特徵	營業利益率、營業毛利率、股東權益報酬率、資產報酬率、稅前淨利率。
成長特徵	營業利益季成長率、毛利率季成長率、稅前淨利季成長率、稅後淨利季成長率、營收季成長率、資產總額季成長率
活動特徵	總資產週轉率、固定資產周轉率、存貨周轉率
穩定特徵	流動比率、速動比率、負債比率、長期資金適合率、股東權益報酬率
價值特徵	稅後淨利、本期淨利

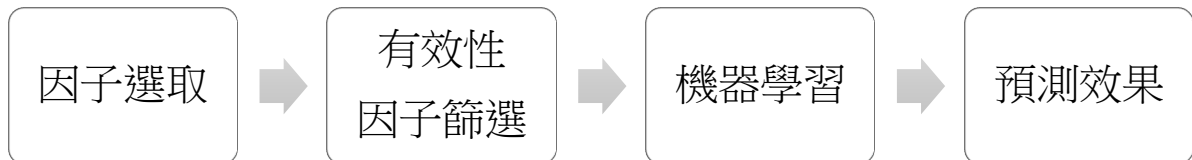
技術指標：運用 python 技術指標套件 TA-Lib 配合 yahoo finance 計算。採用的技術指標以及其指標所含參數如下表（二）。

表（二）

Simple Moving Average 簡單移動平均	SMA
Weighted Moving Average 加權移動平均	WMA
Exponential Moving Average 指數移動平均	EMA
Moving Average Convergence/Divergence 指數平滑異同移動平均	MACD、MACDsignal、MACDhist
Relative Strength Index 相對強弱指數	RSI
Bollinger Bands 布林帶	UPPER、MIDDLE、LOWER
Momentum 動量	MOM
Beta	BETA
Average Directional Movement Index 平均動向指數	ADX
Stochastic 隨機指標	slowK、slowD

三、研究流程

此報告以 0050 成分股的 2010 年第一季至 2020 年第 3 季資料做為訓練資料，將 2020 年第四季做為測試資料。由前 43 季資料進行有效性因子篩選之結果做為機器學習的輸入特徵，並進行訓練。最後由 2020 年第四季的財務報表資料與技術指標，進行未來一季的報酬率預測。



四、有效性因子篩選法

深度學習模型要能準確預測，輸入變量的挑選有其重要性。因此，在進行模型訓練之前，會採用特徵初篩的方式來對輸入特徵作調整。特徵篩選主要是把上述提到的因子與技術指標，在模型訓練前，透過設定的方式來初步篩選，得出較為有效的因子當作輸入變量後，再納入深度學習模型來進行訓練。

要判斷因子是否能顯著區隔報酬，作者以最直觀的想法依據該因子來執行投資組合，因此，使用分組報酬的方式來檢驗。依據每個因子的計算值來對所有股票樣本做排序，每季把由大到小的數值分成設定組數，將股票部位來進行等權重投資，最後依據前 43 季最高分數和最低分數的組合股價差距是否有顯著差異，來做為有效性因子的判斷依據，故會 43 個組合股價差距樣本。本研究是以 t 檢定的 p 值小於 0.10 為判斷標準，以此方法來找出有效性因子。

本次報告以每季每因子下依大小排序分為 5 組，因每季樣本為 0050 成分股，故每季每個因子應為十股一組進行等權重投資。而在做完 t 檢定後，將 43 季中 p 值小於 0.1 的數量找出，並取前 10 個數量較多的因子做為有效性因子篩選法之結果放入機器學習特徵。有效性因子篩選法之結果，包含以下特徵：毛利季成長率、營業毛利率、平均動向指數、布林帶（上線）、稅前淨利季成長率、布林帶（中線）、指數平滑異同移動平均（柱狀圖）、指數平滑異同移動平均（信號線）、營業利益率、稅後淨利季成長率。

五、模型設定

使用 MLP 多層感知機。多層感知機是一種前向傳遞類神經網路，至少包含三層結構（輸入層、隱藏層和輸出層），並且利用到「倒傳遞」的技術達到學習(model learning)的監督式學習。作者使用 MLP 的方式是以股票所擁有的特徵因子為輸入變量，輸入變量在經過 MLP 模型運算後，得出相應的報酬變數作為輸出變量(output)，與實際的報酬變數比較誤差後，再反向傳播回模型中調整權重進行最適化，以此得到最終的 MLP 預測模型。在深度學習模型設定方面，以 ReLU 當作激活函數，可以解決梯度消失問題、使深度學習架構具有稀疏性與避免過大的運算量，並採用適應性動差估計(adaptive moment estimation, Adam)來當作模型的優化函數。

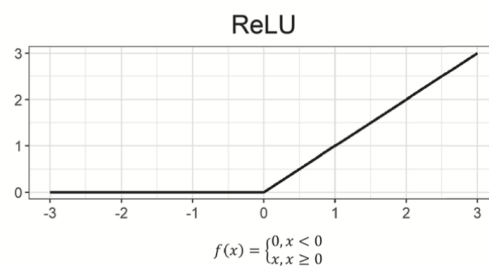


圖 1，激活函數 ReLU 示意圖

而為避免只依賴一次模型訓練而導致預測偏誤，作者採用重複五次的模型訓練所得之五次預測值進行加總平均做為最終預測報酬。因此為進行此作法，在模型參數設定上，不對隨機數種子進行設定。

六、結果呈現

下圖 2 為由 2020 第四季資料預測未來一季的個股報酬，也就是 2021-01-01 至 2021-03-31 之報酬預測與真實報酬值。因橫軸為個股代碼，因此由左而右的高低趨勢在此不具意義。由圖可見，預測效果並不好。

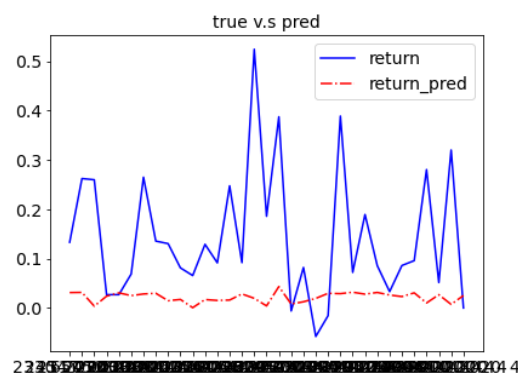


圖 2

圖 3 橫軸為真實值，縱軸為預測值。預測效果越好，其點分佈應接近 45 度斜值線。

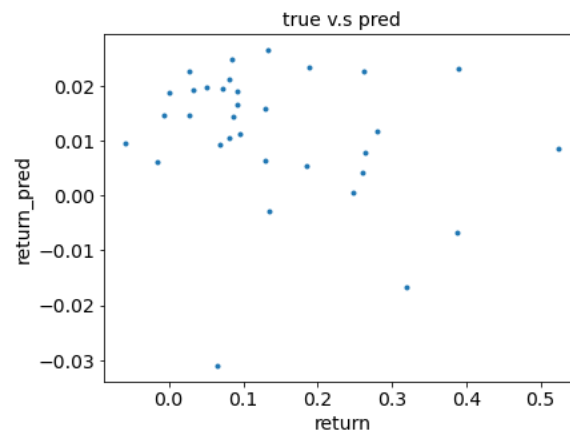


圖 3

因預測結果不好，所以做進一步觀察，檢查是否為過擬合的情況。下圖 4、圖 5 為訓練樣本的真實值與預測值，可以發現此模型在訓練時就已經是訓練不好的情形了。其原因可能是沒有進行交叉驗證及在因子與特徵的選取不夠妥當。

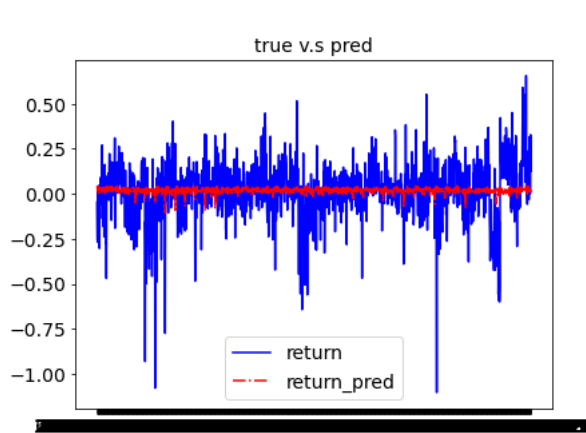


圖 4

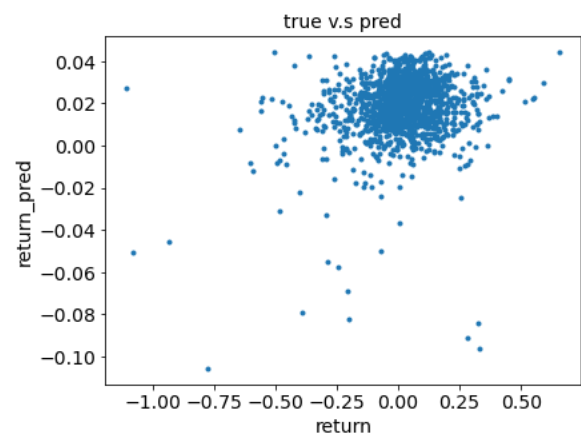


圖 5

七、參考資料

吳錦文; 王昭文; 謝育展. 深度學習在 Smart Beta 交易策略之應用. 台灣管理學刊, 2020, 20.2: 77-110.