

機器學習在股票型態辨識之應用

指導教授：中山財管所 王昭文 教授

簡報製作：中山財管所 蘇彥庭 研究助理

演講摘要

- 在證券投資之技術分析領域中，大致可分為指標、K棒及型態3類分析方法，如何從中找出有用的技術分析方法，提升交易策略的報酬率，一直是投資人常在思考的問題，機器學習提供另一種解決方法。本演講以上海及深圳A股市場為例，介紹如何運用機器學習模型，強化技術分析交易策略的報酬率與勝率。透過深度學習模型中的多層感知器 (Multi Layer Perceptron, 簡稱MLP)，將股價向上突破20日均價進場，股價小於關鍵K棒則出場之交易策略，以進場日的技術指標及型態做為特徵，每筆樣本報酬率的漲跌進行模型訓練，藉以預測未來的股價漲跌。實證結果發現，透過機器學習模型，確實可以有效地提升原有交易策略的績效。
-

技術分析之關鍵K棒交易策略

- 回測標的：中國上證及深證A股2015年平均市值前300名之股票
 - 回測期間：2005年01月至2017年12月
 - 進場條件：當日收盤價向上突破20日移動平均線
 - 出場條件：跌破關鍵K棒出場價格
 - 交易成本：買入千分之3、賣出千分之4
-

移動平均線介紹

- 移動平均線用以反映目前投資人的股價持有成本
- n日簡單移動平均線(Simple Moving Average, SMA)計算方法：

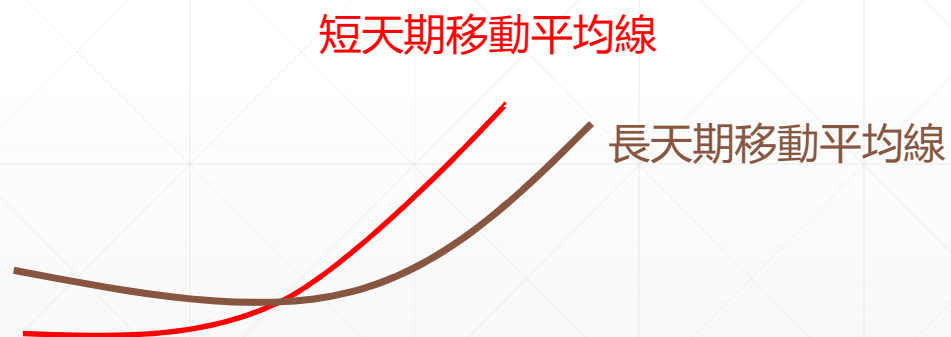
$$SMA_n = \frac{P_t + P_{t-1} + \dots + P_{t-n+1}}{n}, \text{ t 為現在時間點}$$

第t天	T-6	T-5	T-4	T-3	T-2	T-1	T
收盤價格	10	11	12	15	14	12	9
5日移動平均線	-	-	-	-	(10+11+12+15+14)/5 =12.4	(11+12+15+14+12)/5 =12.8	(12+15+14+12+9)/5=12.4

常見的移動平均策略

黃金交叉

短天期移動平均線向上突破長天期移動平均線
股價走多頭的機率高



死亡交叉

短天期移動平均線向下跌破長天期移動平均線
股價走空頭的機率高



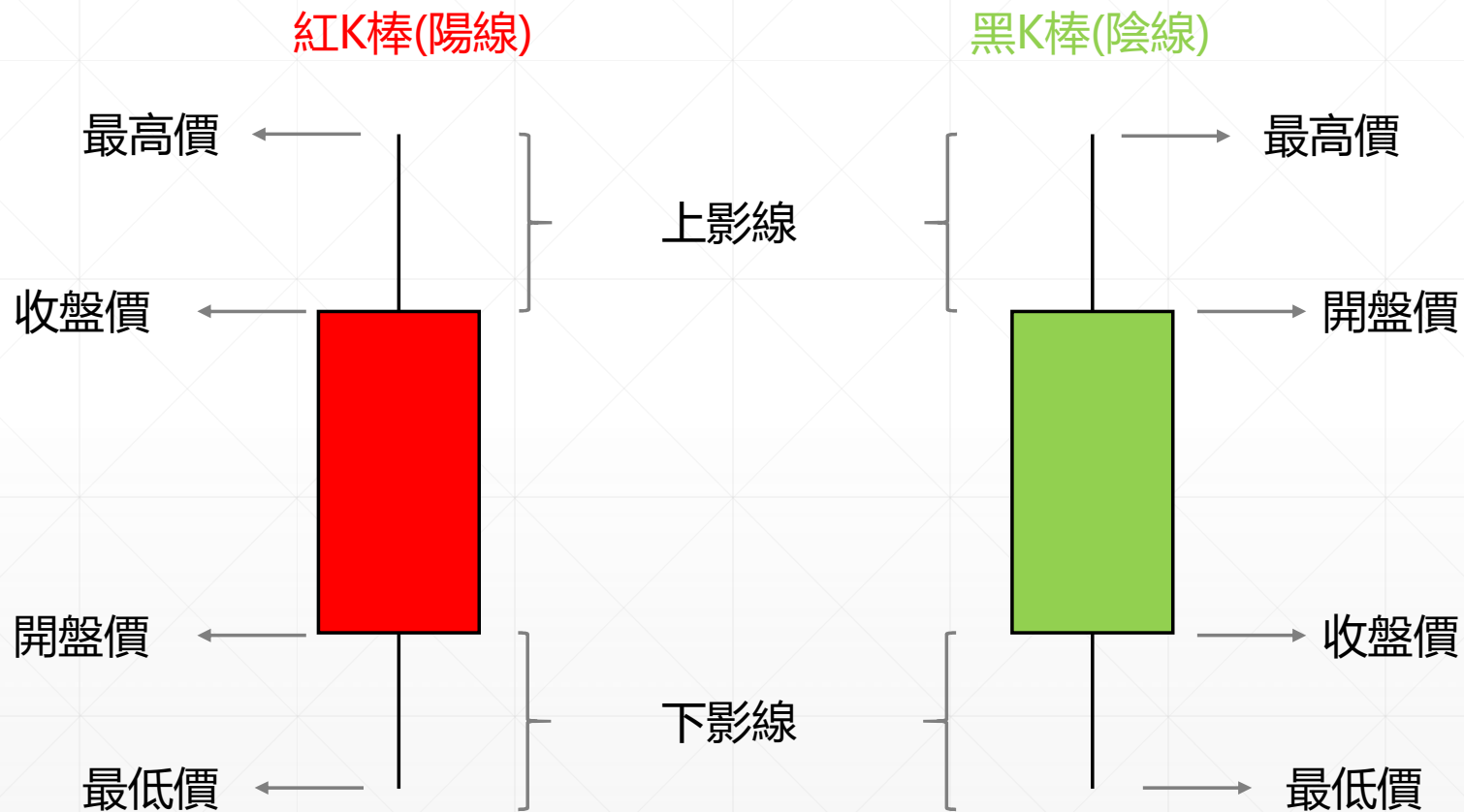
黃金/死亡交叉範例

上證指數



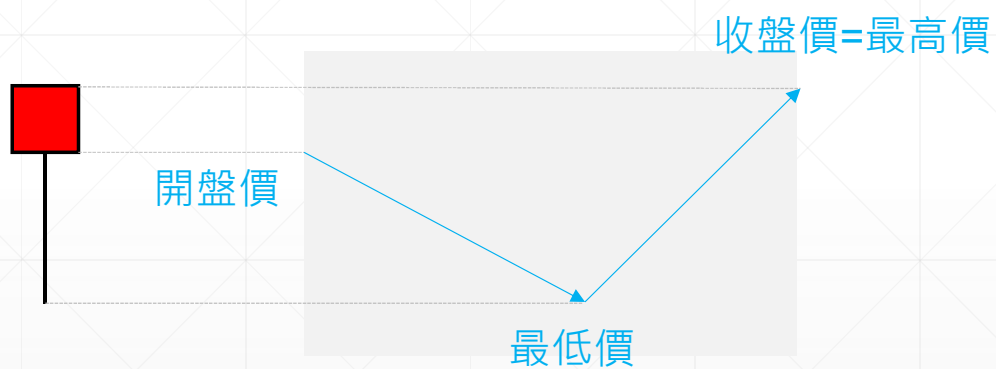
K棒

- K棒是由開盤價、最高價、最低價及收盤價所組成，用以描繪買方和賣方交戰的過程

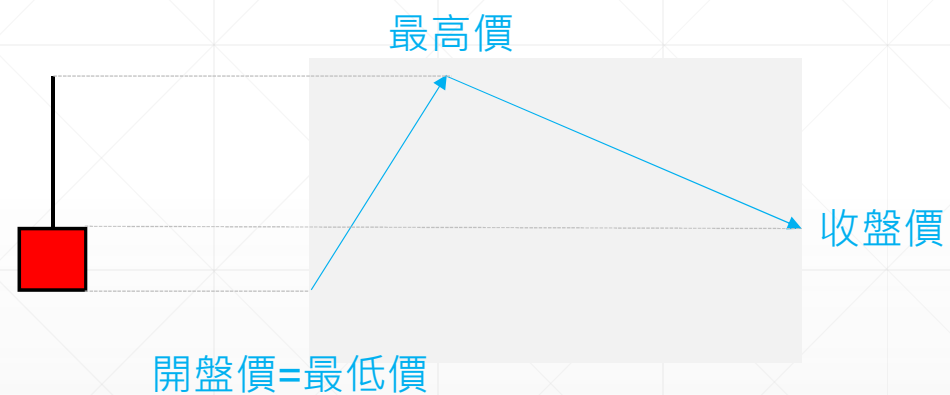


K棒意涵

- 紅K棒，長下影線

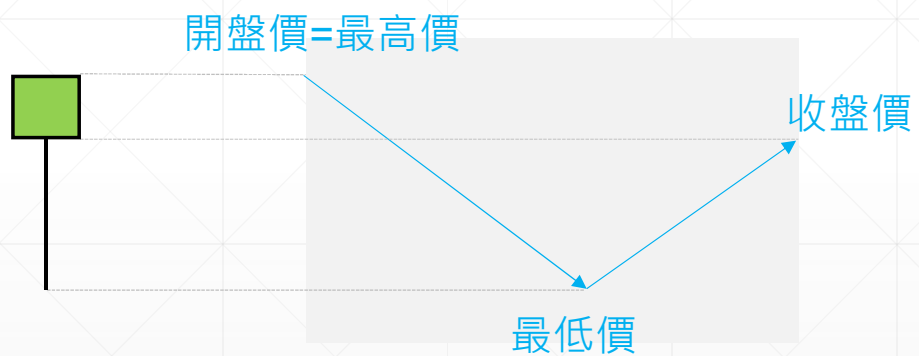


- 紅K棒，長上影線

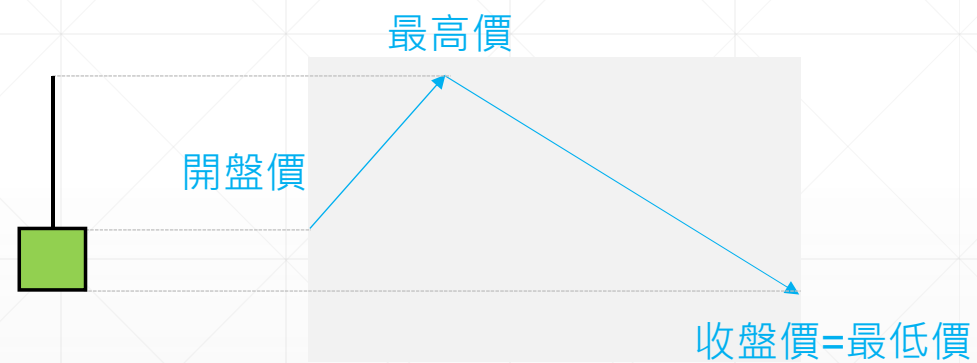


K棒意涵

- 黑K棒，長下影線

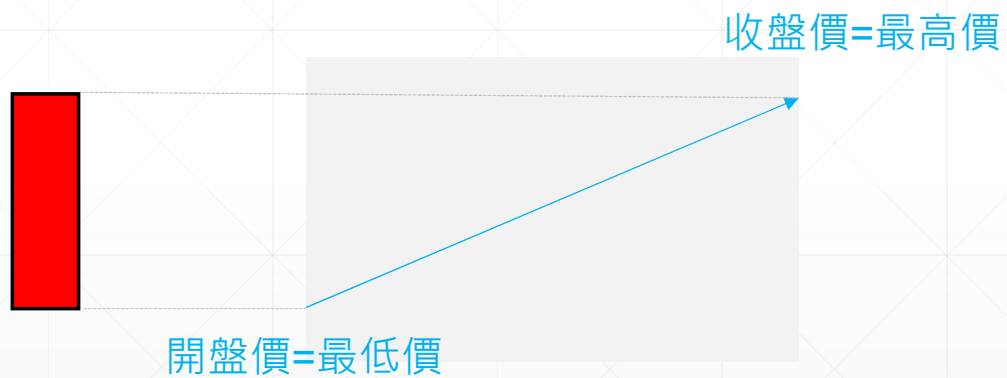


- 黑K棒，長上影線

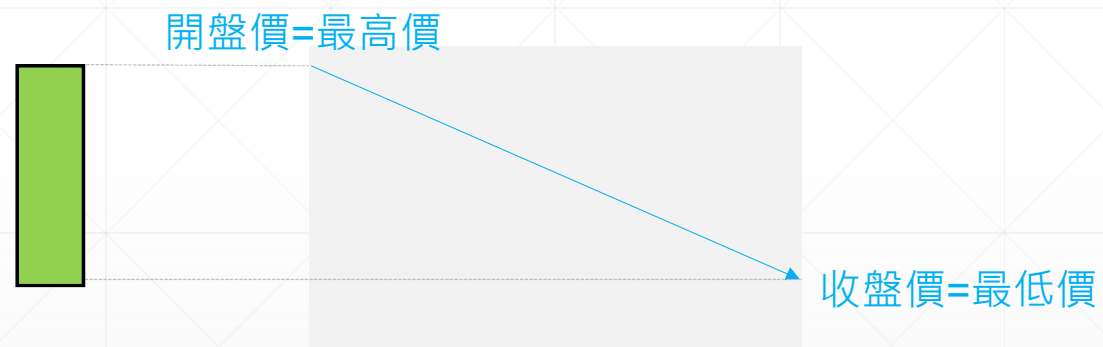


K棒意涵

- 長紅K棒



- 長黑K棒



關鍵K棒出場

- 關鍵K棒：長紅K棒和長黑K棒
 - 以進場價格跌5%的價位做為目標出場價格
 - 進場後逐日判斷當日是否為關鍵K棒，若為關鍵為關鍵K棒，則計算出場價格
 - 若為紅K棒，出場價格為當日開盤價向下5%
 - 若為黑K棒，出場價格為當日收盤價向下5%
 - 若關鍵K棒的首場價格比目標出場價格來得高，則做為新的目標出場價格
 - 關鍵K棒出場即為移動停利(損)的概念
-

技術分析之關鍵K棒交易策略示例



技術分析之關鍵K棒交易策略示例



技術分析之關鍵K棒交易策略示例

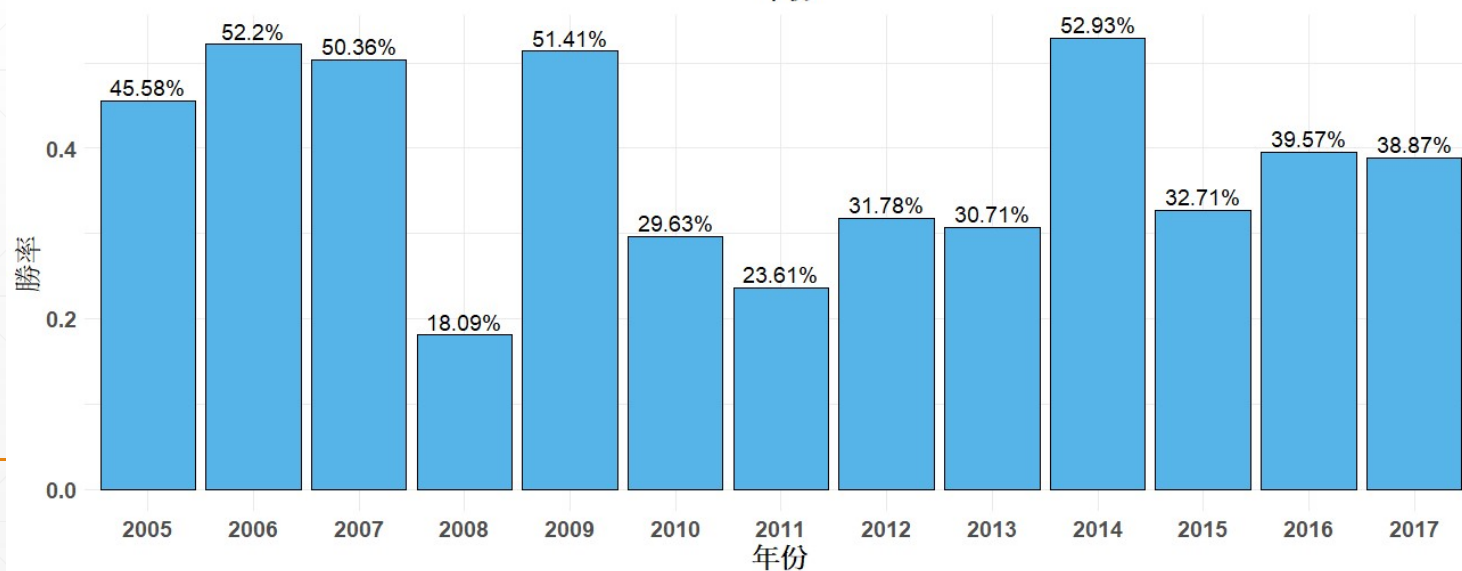
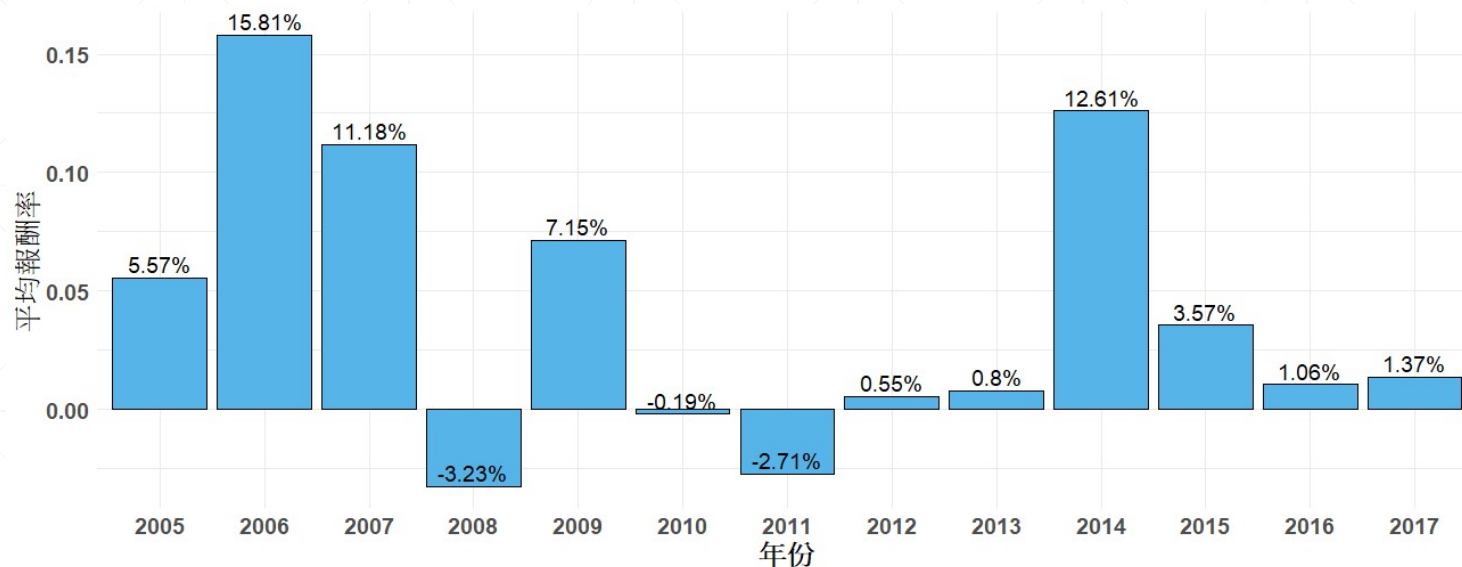


技術分析之關鍵K棒交易策略

- 回測標的：中國上證及深證A股2015年平均市值前300名之股票
- 回測期間：2005年01月至2017年12月
- 進場條件：當日收盤價向上突破20日移動平均線
- 出場條件：跌破關鍵K棒出場價格
- 交易成本：買入千分之3、賣出千分之4
- 策略績效：

平均報酬率	勝率	交易筆數	平均持有日數	年化報酬率	報酬率標準差	最大報酬率	最小報酬率
3.55%	37.52%	40,027筆	41.14日	31.52%	23.68%	1106.46%	-25.36%

技術分析之關鍵K棒交易策略績效分析



以機器學習強化策略

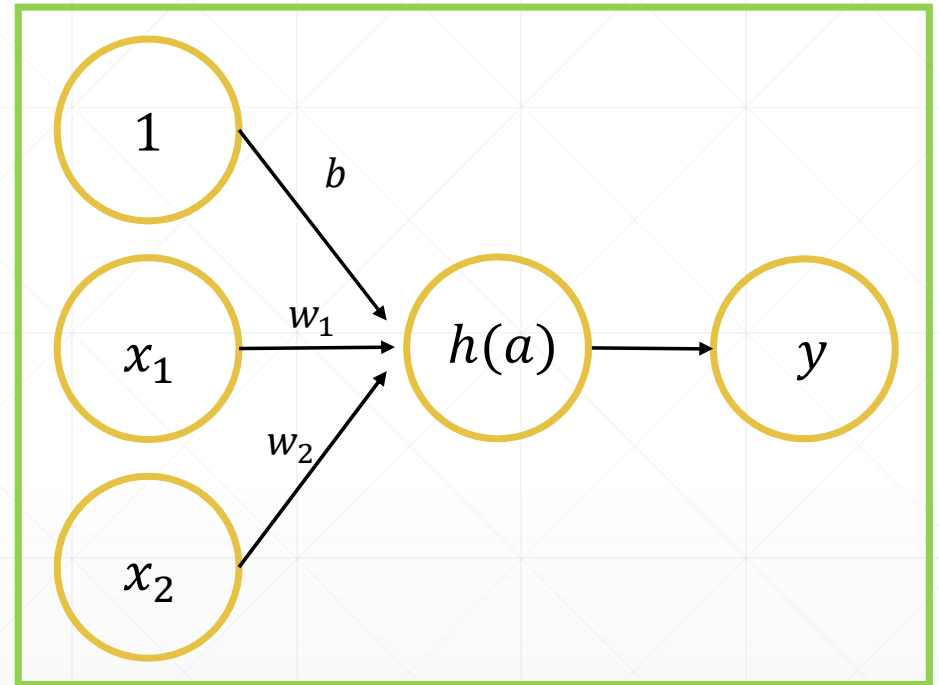
- 利用股票型態特徵及日周月的技術分析指標，以深度學習模型預測上證及深證A股指數的進場時機點，提升原有策略的績效指標。
- 訓練期(Train)：2005年01月至2015年12月，共21,444筆交易。
- 預測期(Test)：2016年01月至2017年12月，共7,394筆交易。
- 預測目標：各筆交易的報酬率漲跌兩分類。 $Y_i = \begin{cases} 0, & \text{Return}_i \leq 0 \\ 1, & \text{Return}_i > 0 \end{cases}$

	平均報酬率	勝率	交易筆數	平均持有日數	年化報酬率	報酬率標準差	最大報酬率	最小報酬率
全部樣本	3.55%	37.52%	40,027筆	41.14日	31.52%	23.68%	1106.46%	-25.36%
訓練期樣本	3.29%	35.91%	21,444筆	37.21日	32.25%	21.19%	478.53%	-15.08%
預測期樣本	1.40%	39.90%	7,394筆	57.39日	8.88%	12.37%	116.44%	-25.36%

註：此處訓練期樣本數與預測期樣本數相加未等於全部樣本，是因為刪除特徵缺值影響

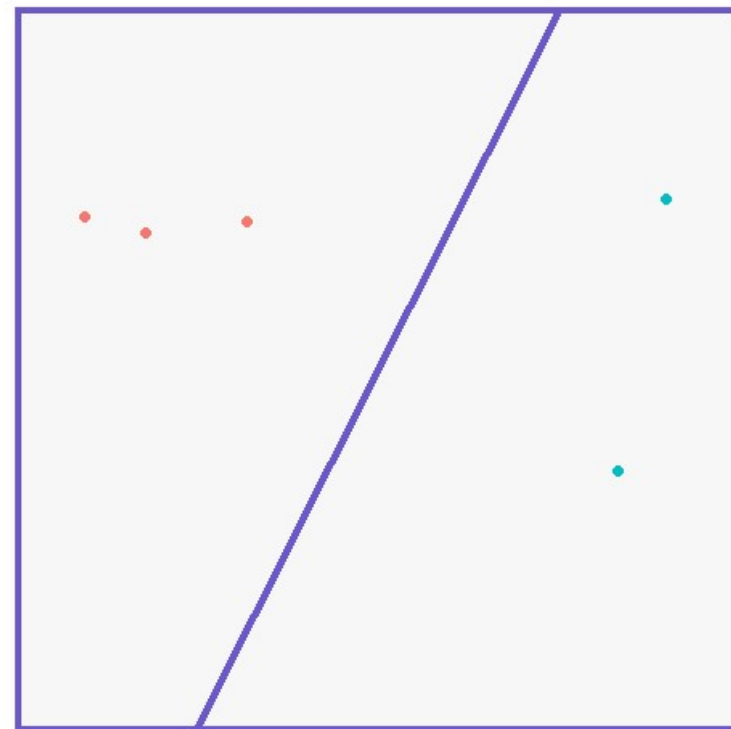
感知器介紹

- 感知器模型：
- $a = b + w_1X_1 + w_2X_2$
- $Y = h(a)$
- $h(x)$ 為激勵函數(Active Function)
- X_1 和 X_2 為特徵(Features)
- w_1 和 w_2 為特徵權重
- Y 為預測目標
- 透過權重的調整，讓模型可以學到某個函數，並且區分資料分類



感知器訓練過程






- 兩分類預測問題
- 感知器模型： $Y = \text{sign}[w_1X_1 + w_2X_2 - w_0]$
- 感知器模型訓練步驟：
 1. 初始化：權重 w_i
 2. 激勵：輸入第一筆樣本 (x_1, x_2) 得到輸出值 Y
 3. 權重修改：依據實際的 Y_d 值與感知器輸出的 Y 值之間的差異進行調整
 - 3-1 計算誤差： $e = Y_d - Y$
 - 3-2 計算調整量： $\Delta w_i = \alpha \cdot x_i \cdot e$
 - 3-3 調整權重： $w_i = w_i + \Delta w_i$
 4. 重複2-3步驟，依序輸入樣本進行權重修改，直到模型能夠完全正確分類或者達到迭代次數上限



圖片來源：<https://github.com/micahprice/perceptron-GIF>

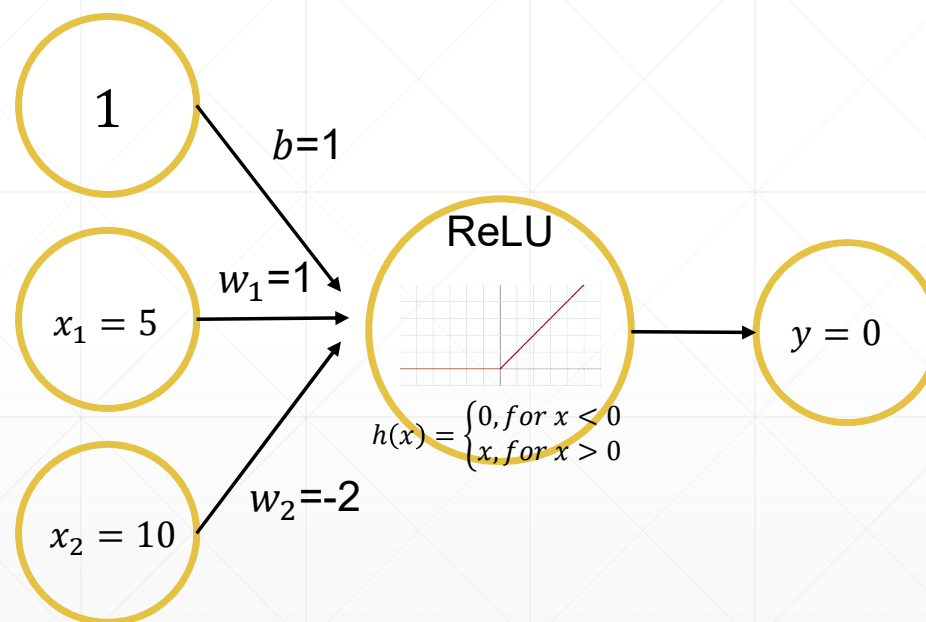
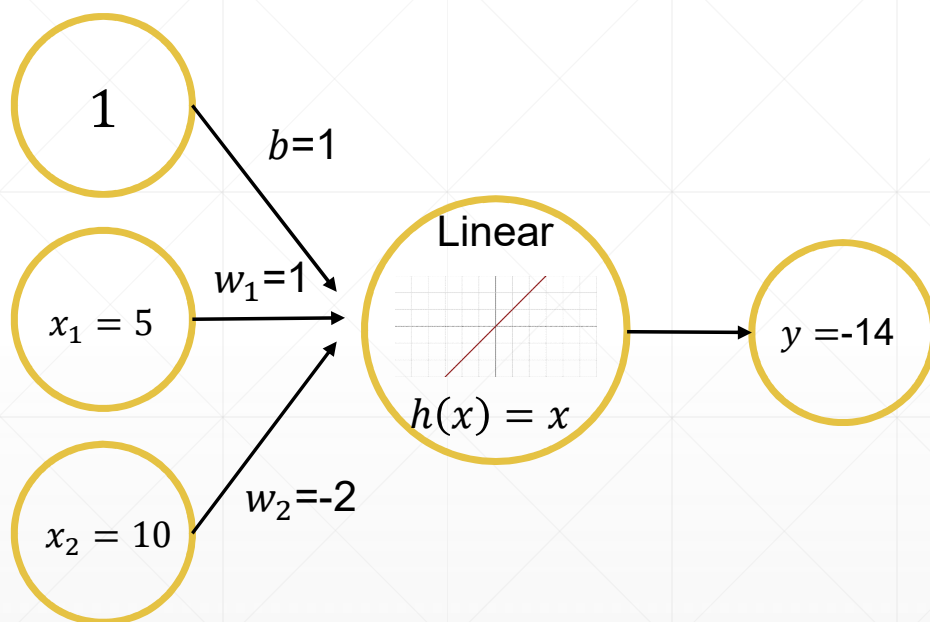
激勵函數介紹

- 激勵函數：將輸入訊號的總和轉換成輸出訊號的函數。
- 不同的激勵函數有各自的輸出，進而對整個模型造成影響。

激勵函數	圖形	函式	輸出值域
Binary Step		$h(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1, & \text{for } x > 0 \end{cases}$	$\{0, 1\}$
Linear		$h(x) = x$	$(-\infty, \infty)$
Sigmoid		$h(x) = \frac{1}{1 + e^{-x}}$	$(0, 1)$
Tanh		$h(x) = \frac{2}{1 + e^{-2x}} - 1$	$(-1, 1)$
ReLU		$h(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } x > 0 \end{cases}$	$[0, \infty)$

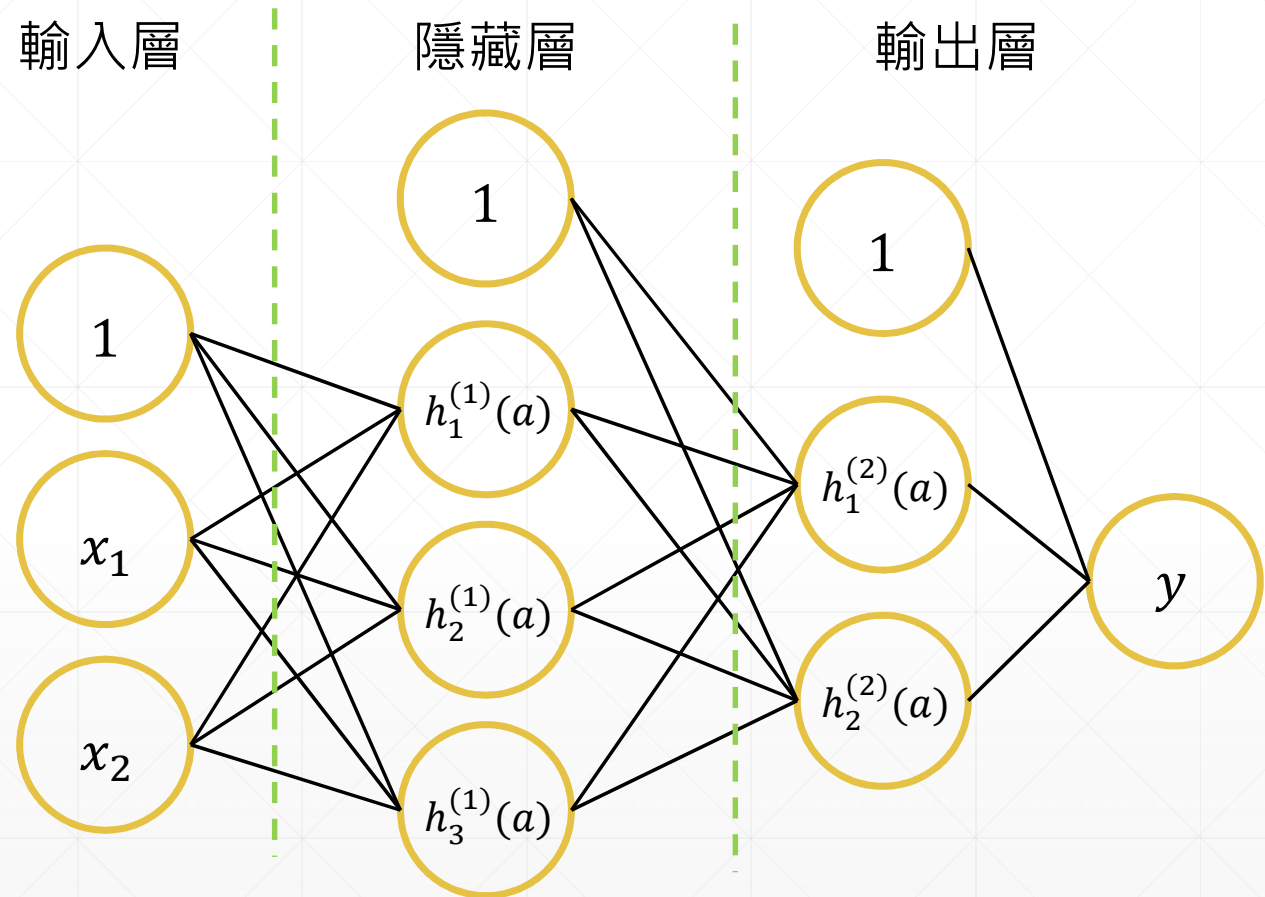
激勵函數圖形引用至維基百科：https://en.wikipedia.org/wiki/Activation_function

激勵函數範例



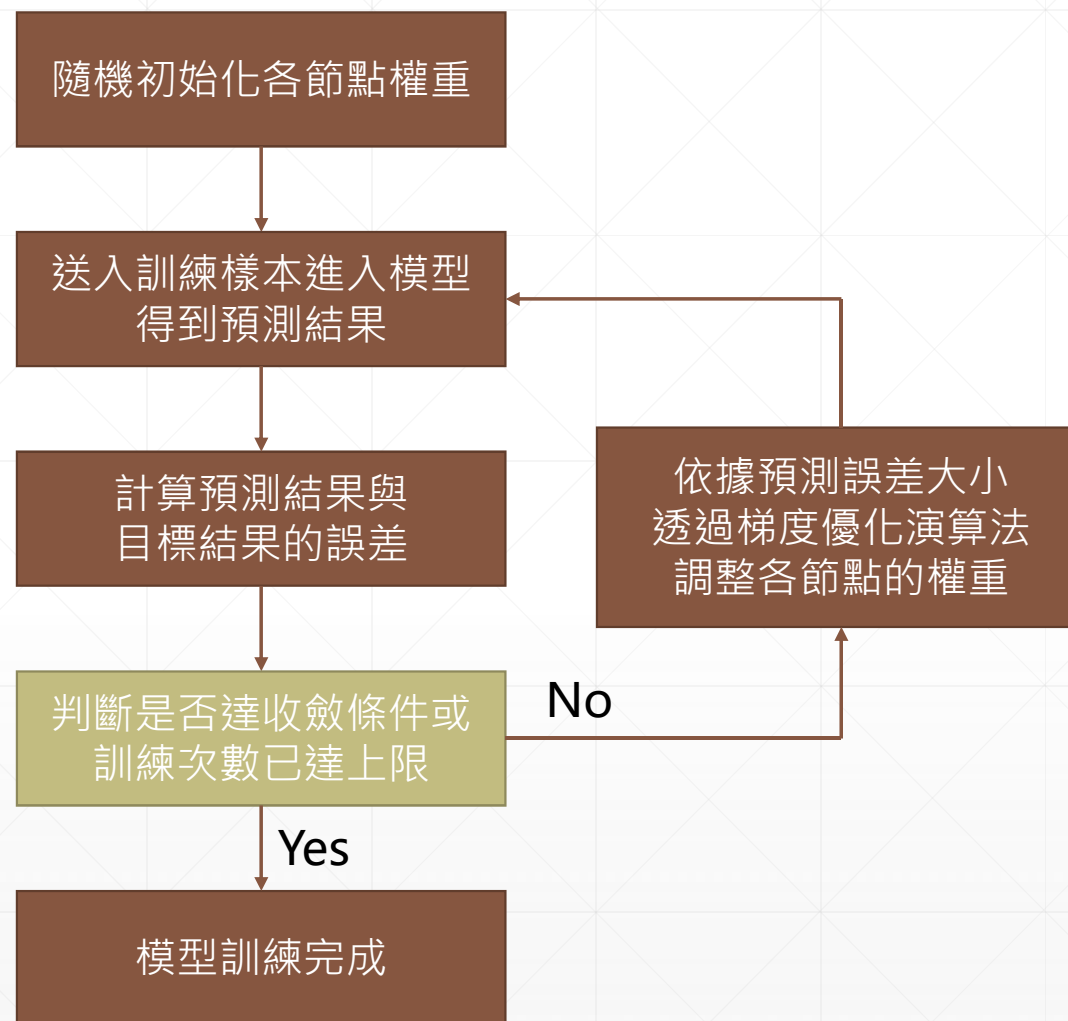
多層感知器

- 多層感知器(Multi Layer Perceptron)包含輸入層、輸出層以及至少1個隱藏層。
- 透過多層的設定，使模型能夠學習非線性函數。
- 互動網站：
<http://playground.tensorflow.org/>

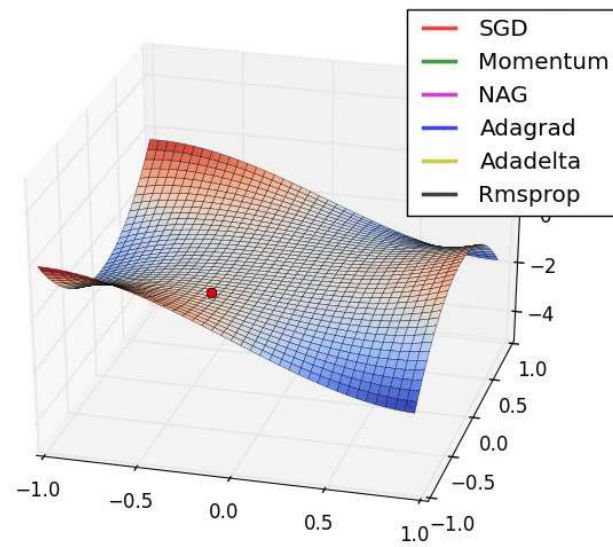
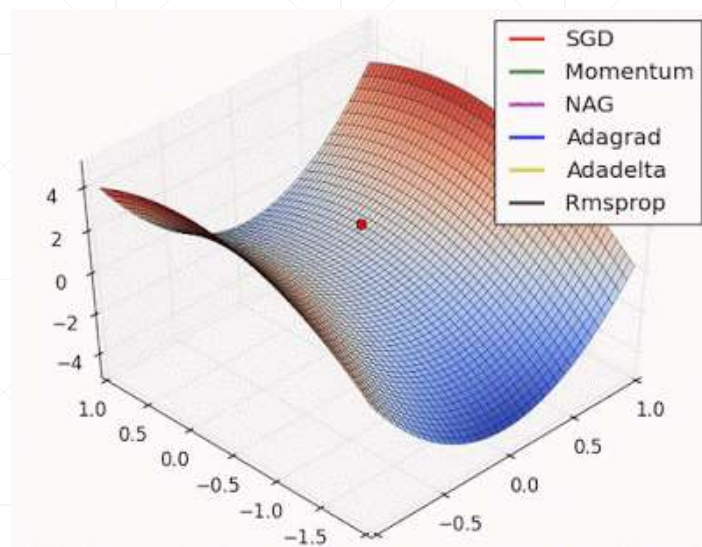
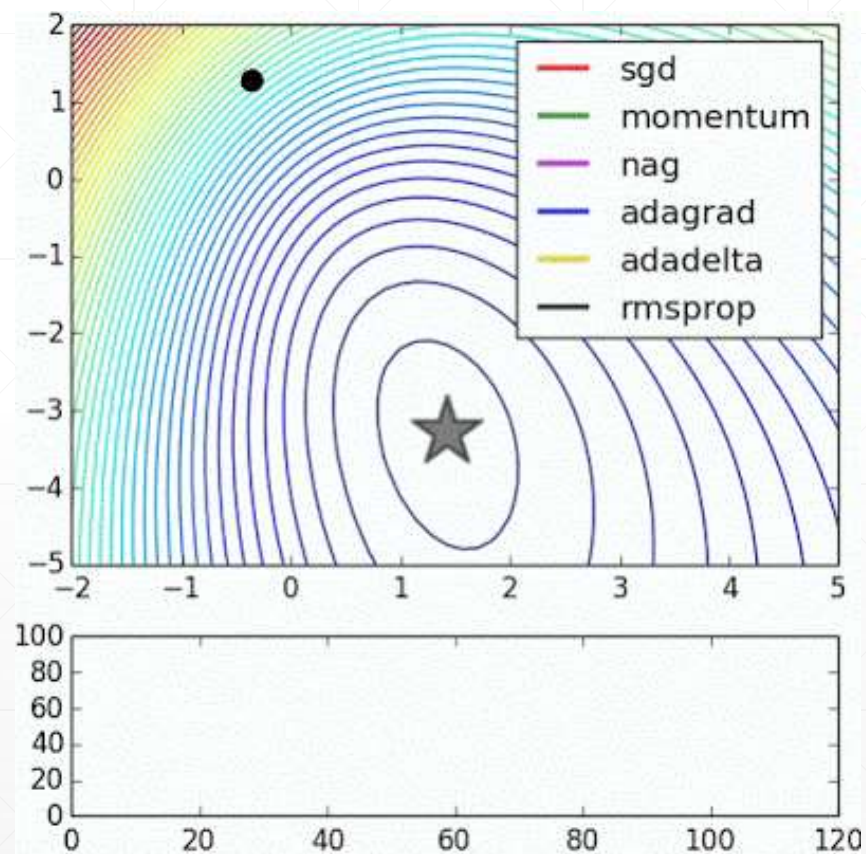


模型訓練方式

- 反向傳播法
(Backward-Propagation)
- 讓訓練樣本通過模型得到初步預測結果，並計算預測與實際結果的誤差。再將誤差往回傳遞到模型中，調整模型內各節點的權重。不斷重複上步驟，持續調整模型，直到預測誤差達到收斂條件或訓練次數達到使用者設定上限。

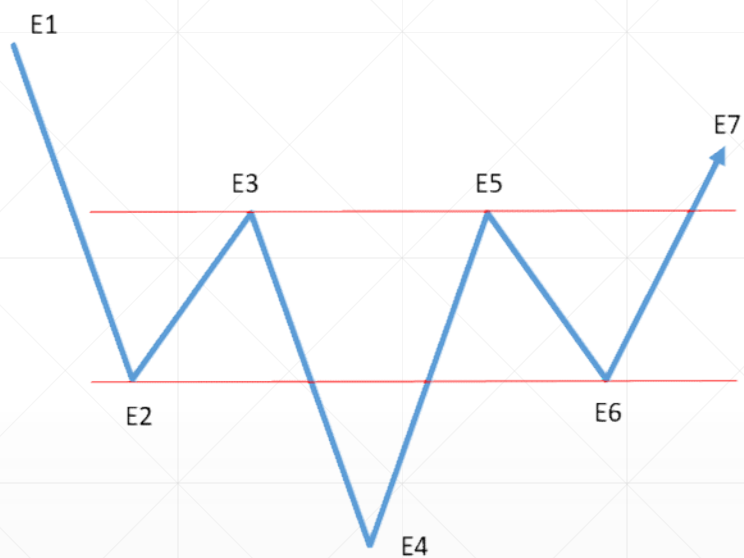


梯度優化演算法

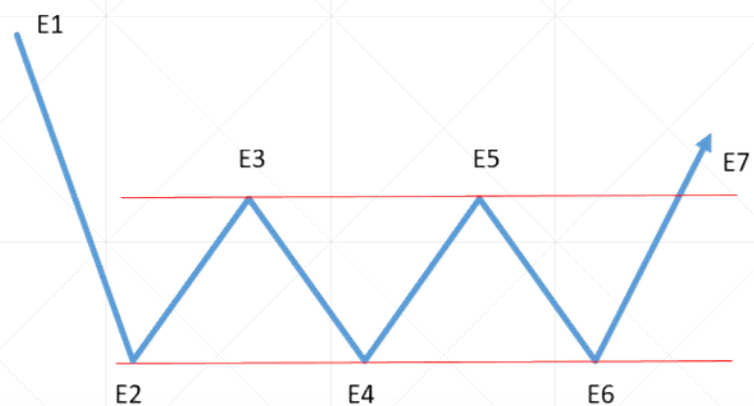


圖片來源：<http://www.denizyuret.com/2015/03/alec-radfords-animations-for.html>

股票型態特徵

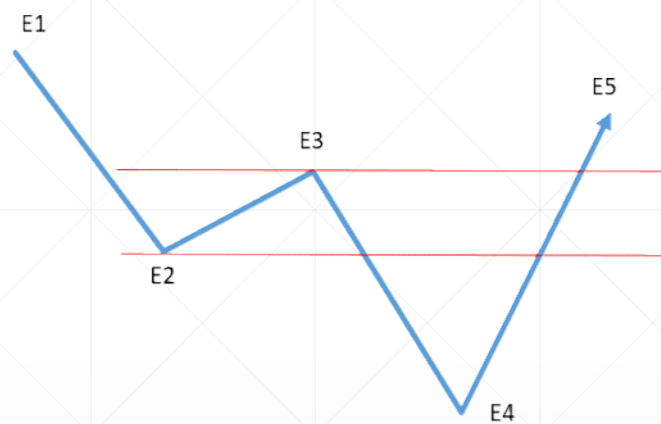


頭肩底

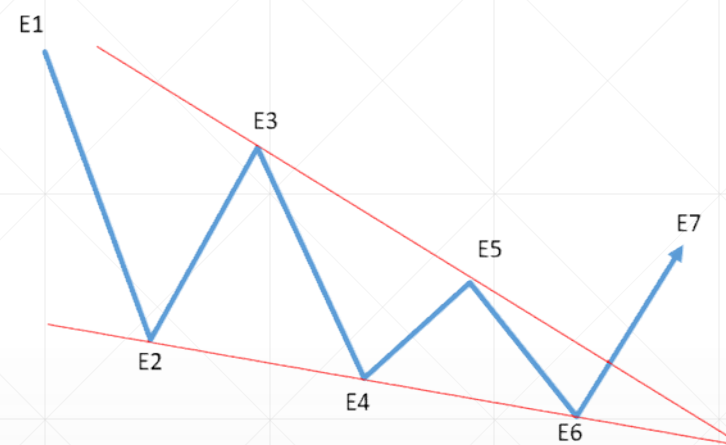


三重底

股票型態特徵



V底



下降楔形

機器學習特徵

- 學習特徵共有型態及技術分析指標兩個面向，共計有63個特徵
 - 型態特徵
 - 進場點價格與頂底之間的關係
 - 頂底之間的斜率關係
 - 頂底之間的時間長度關係
 - 技術分析指標特徵
 - 日、周及月頻布林通道指標
 - 日、周及月頻KD指標
 - 日、周及月頻MACD指標
 - 日、周及月頻均線與進場點價格之間的關係
 - 日、周及月頻均量與進場點成交量之間的關係
-

多層感知器模型設定

- 程式執行環境：R語言中使用Keras套件 <https://keras.rstudio.com/>
- 使用多層感知器模型(Multi Layer Perceptron)
- 對訓練期和預測期特徵值進行標準化，將值域縮放至[0,1]區間
- 誤差函數：分類交叉熵損失(Categorical Cross-Entropy)
- 梯度優化演算法：adam(Adaptive Moment Estimation)
- 訓練次數：2000次
- Batch-size：全部訓練期樣本
- 驗證期(validation)切割比率設定為10%

多層感知器模型架構



預測分類結果

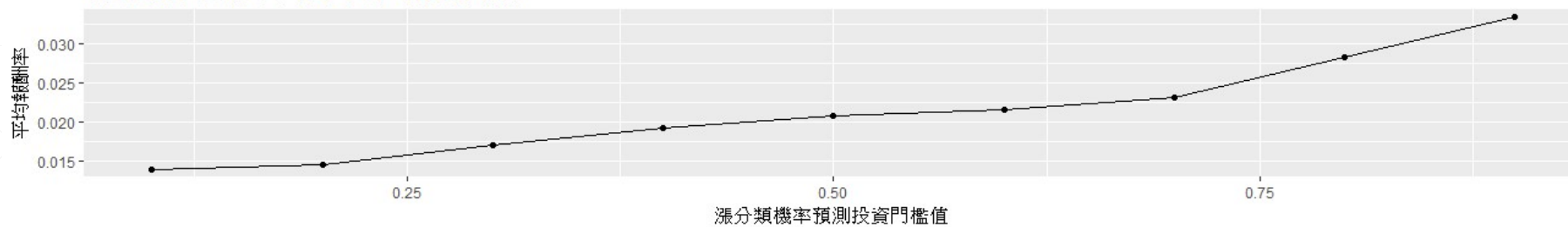
- 預測整理準確率：51.74%
- 預測漲實際漲準確率：41.21%

混淆矩陣(Confusion Matrix)			
		實際	
		跌分類	漲分類
預測	跌分類	2,378	1,502
	漲分類	2,066	1,448

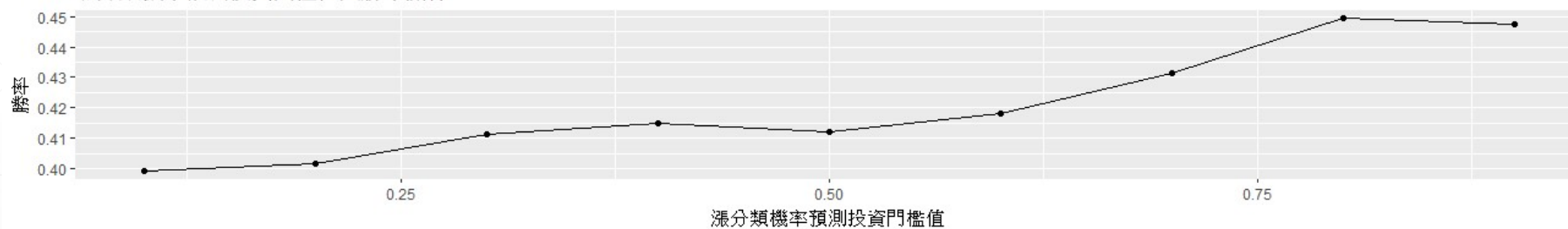
- 在財務投資領域中，重視預測漲分類準確率，因為預測漲才會實際拿錢去投資
 - 以各樣本預測漲的機率，透過門檻值設定來決定是否實際投資
-

預測品質

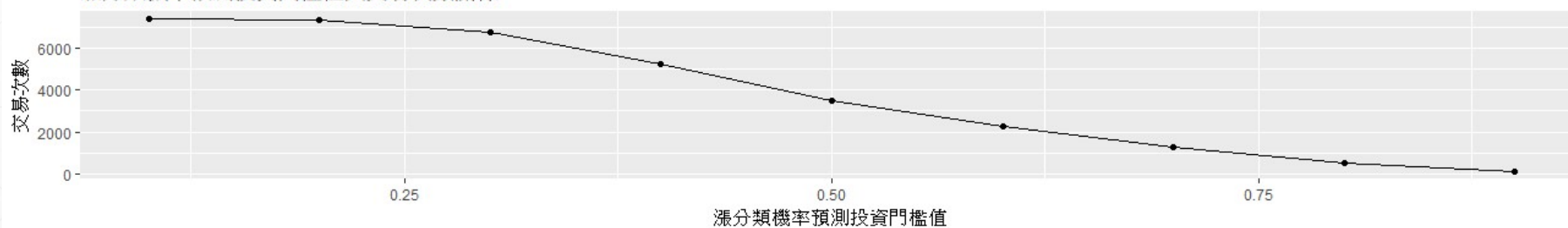
漲分類機率預測投資門檻值與平均報酬率關係



漲分類機率預測投資門檻值與勝率關係



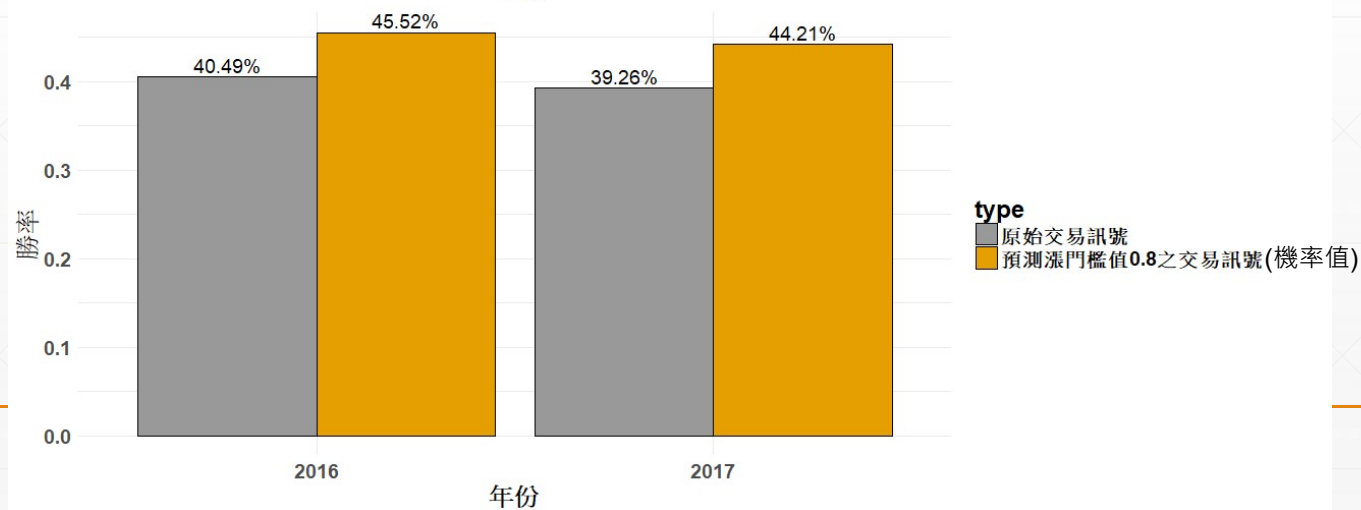
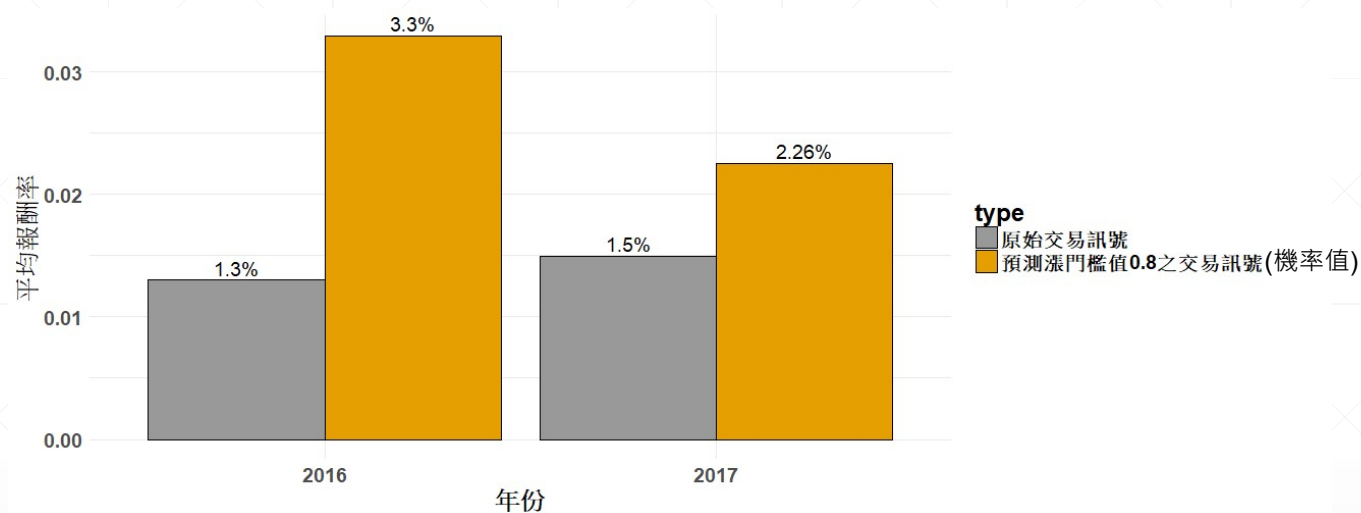
漲分類機率預測投資門檻值與交易次數關係



預測績效分析

	平均報酬率	勝率	交易筆數	平均持有日數	年化報酬率	報酬率標準差	最大報酬率	最小報酬率
預測期 全部樣本	1.40%	39.90%	7,394筆	57.39日	8.88%	12.37%	116.44%	-25.36%
模型預測樣本 門檻值=0.5	2.08%	41.21%	3,514筆	62.22日	12.40%	12.77%	116.44%	-14.38%
模型預測樣本 門檻值=0.6	2.16%	41.81%	2,260筆	62.44日	12.62%	12.65%	116.44%	-14.38%
模型預測樣本 門檻值=0.7	2.31%	43.15%	1,284筆	63.48日	13.28%	12.71%	116.44%	-14.38%
模型預測樣本 門檻值=0.8	2.83%	44.93%	523筆	63.99日	16.16%	13.87%	116.44%	-10.83%
模型預測樣本 門檻值=0.9	3.35%	44.76%	105筆	62.78日	19.45%	14.03%	53.92%	-10.83%

預測績效分析



結語

- 在整個AI投資策略的研發中，時常遇到的困難是我們應該要如何從手中的資訊，找到獲利的關鍵因素。在財務領域之AI程式交易開發過程中，由於價格走勢並非如圖形辨識或手寫辨識有固定或相似的型態重覆發生，故目前仍需要仰賴人類過去的交易經驗和主觀想法，先行開發一個雛型策略，再應用機器學習方法強化原來的交易策略。
 - 在本次演講中，我們首先呈現一個基本的交易策略，並透過機器學習模型的協助，來提升交易策略的績效。在2016年至2017年之間，平均報酬率由1.40%提升至2.83%，勝率由39.90%提升至44.93%。此結果證明機器學習領域確實能應用於財務投資交易策略領域中。
-