

以類神經網路方法建構房價估價模型-以高雄市實價登錄資料為例

Development of Reality Pricing Models by Applying Artificial Neural Network Based on Actual Price Registration Data in Kaohsiung City

姜林杰祐¹

國立高雄應用科技大學 金融資訊研究所 教授
clcy@cc.kuas.edu.tw

黃惠芬²

國立高雄應用科技大學 金融資訊研究所 研究生
1104342111@gm.kuas.edu.tw

摘要

本研究目的為使用內政部所提供的公開資訊建構房價估價量化模型。分析資料取自內政部實價登錄資料服務平台及內政部不動產資訊平台，分析期間自 2013 年 1 月到 2016 年 12 月，研究範圍包含高雄市高交易量地區的集合型住宅及非商用透天宅，運用公開資訊的房屋特徵值，並以類神經網路分析方法，建構出高雄市的住宅用房屋價格估計模型。

研究中，影響房價的因素除了參考實價登錄資料的項目外，慮及高雄捷運對於住宅的交通便捷性影響，與總體經濟變數亦可能為影響房產價格的重要資訊；因此，研究將建物地理位置轉成 GPS 座標，並估算建物到公共運輸點的距離，以及同期台股股價指數、五大行庫平均房貸利率、消費者物價指數與 M1b 貨幣供給額等變數，評估其對於房價的影響。

透過本研究所建構的房價估計模型，可提供房屋市場的買賣方或仲介端估算出合理的開價基礎。本研究雖僅以高雄市高交易量地區的房價估計做為實證範圍，但模型亦可運用於其他地區之房價預估。

關鍵詞：實價登錄、房價、類神經網路

第一章 緒論

一般民間或交易行為所謂的不動產，則稱之為房地產(Real Estate)，其具體內容包含了土地及建物兩大類型；最大特徵是，土地不僅沒有折舊，在地狹人稠的台灣地區，尤其是都會地區，因需求大於供給或是投資(機)因素推動的影響，還可能隨時間增值，甚至造成短期內之大幅上漲。

而建物的耐用年限，以最常見的鋼筋混凝土造來說，其物理耐用年限達六十年之久，近年更由於重視制震以及災防需求而採用的鋼骨結構更，大幅增加安全性及物理耐用年限。而隨著土地成本的遞增以及政府公共建設的投入，亦使得建物的經濟使用年限普遍的超越了物理使用年限，而增加的效益也就反映在房價的推升上。

由於不動產之產品以及市場具有：(1)產品差異性大，(2)兼具消費財與投資財(或資本財)特性，(3)短期供給缺乏彈性，以及(4)不完全競爭市場等特徵，故一般民眾對於不動產市場價格相對也欠缺可資提供買賣、投資的決策工具。

購置住宅為多數民眾生涯規劃的重大里程碑，但在置產價金高而且貸款付款期間長的壓力下，消費者購屋置產之前，確實需要能夠進行大量相關資訊蒐集或長期觀察房價變化，提供擬定購地(屋)決策依據。

高雄捷運系統於 2008 年開始運行，計有紅橘兩線(不考慮性質與捷運略有差異的輕軌捷運)，縱橫貫穿高雄市人口稠密區域，目前僅有十字交叉，尚未構成捷運網路，沿線地標整理如表 1.1 所示，捷運沿線已經包含高雄近半的交通運輸據點、百貨公司商圈、公園休閒綠地、文化藝術活動園區、學校、就業點等熱區，囊括計劃購屋者的期待於一身，捷運沿線之房地交易價格之變化亦為本研究所關注之重點之一。

表 1.1 高捷沿線地標整理(資料來源：本研究整理)

高捷沿線地標整理	紅 線	橘 線
交通運輸系統據點	高雄國際機場、高雄車站、高鐵左營站	國道一號中正路交流道
百貨公司商圈	大魯閣草衙道、統一夢時代百貨、好市多賣場、三多商圈、漢神巨蛋商圈、高鐵新光三越百貨	鳳山市區商圈、大統百貨、文化中心商圈
文化、公園綠地、休閒	勞工公園、中央公園、生態園區、楠梓都會公園、世運主館場、橋頭糖廠	大東文化園區、衛武營藝術文化中心、高雄文化中心、愛河新灣區、駁二園區、哈瑪星、西子灣
學 校	高雄醫學院、高雄海洋科大、高雄第一科大	高雄師範大學、中山大學
就業點	高雄軟體園區、楠梓加工區	中正路辦公商業區

由於不動產交易資料不透明，房屋的成交價格欠缺具公信力的平台可提供民眾查詢，只能透過仲介業者提供所謂的行情資料作為參考依據，但這些資料因為缺乏公信力，以致民眾沒有辦法取得房屋的真實交易價格資訊。消費者為資訊弱勢的一方，買賣糾紛時有所聞，反映出房地產交易透明化、公開化是不動產交易公平之重要基礎；交易資訊愈透明公開，買賣雙方之權益愈受到保障。有鑑於此，內政部推動實價登錄制度，修正平均地權條例、地政士法及不動產經紀業管理條例等條文，強制規定土地及房屋的買賣案件、房仲經手之租賃案件及代銷預售屋案件之不動產交易資訊登錄，並於區段化、去識別化後提供查詢，讓民眾在買賣土地或房屋時可以有具公信力之資訊可以參考。

本研究整理不動產交易實價查詢系統於 2013 ~ 2016 年高雄市各行政區之不動產交易，表 1.2 所示為 2013 ~ 2016 年間高雄集合住宅成交案件前五區之案件數量及成交總金額摘要統計。

表 1.2 高雄集合住宅成交案件前五區案件數及總金額統計(資料來源：本研究整理)

成交案件數量前五區		成交總金額前五區	
行政區	佔全市交易案件量百分比 %	行政區	佔全市成交總金額百分比%
三民區	16.71	鼓山區	21.42
鳳山區	15.3	左營區	15.42
左營區	12.67	三民區	15.18
鼓山區	12.18	鳳山區	12.26
楠梓區	11.63	楠梓區	8.5
小計	68.49	小計	72.78

依據前述不動產交易實價查詢系統資料顯示，目前高雄市交易頻繁之前五個行政區域，依成交量及成交金額所占比率，依序分別為：三民區、鳳山區、左營區、鼓山區、楠梓區。

其中，除楠梓區之區域位置係屬郊區外，其餘四區則緊鄰高雄都會核心(苓雅區、新興區)之周邊。此五個行政區域均因有大量之重劃或區段徵收土地釋出，加上捷運路線貫穿以及大型公共設施之投入與改良等因素，造成較其他行區域更多的新屋供給，並吸引區域內外人口進駐，進而呈現交易熱絡之景象。

自 2013 至 2016 年間，統計此五個行政區域，其市場成交案件量約占全高雄市的 68.49%，而成交總金額占全高雄市之比率也高達 72.78%，所以本研究即以此五個行區域內主要的市場交易建築類型—中高樓層集合住宅，做為研究對象，就其座落區段(門牌)、總樓層數、建物型態、主要用途、建築完成年月、環境影響因素等特徵予以分析，研究資料源自內政部不動產資訊的公開資訊，希望藉此建構量化分析模型，提出可供消費者執行買賣屋決策之參考依據。近年來，因為內政部實價登錄資訊的揭露，使得重啟不動產估價的研究更具意義，此即本研究的動機；有了正確的資料，是否能支持更準確的房產估價模型的開發，是本論文的研究目標。

第二章 演算法理論基礎

(一) 運用現代科技輔助執行不動產大量估價之方法

在「不動產估價技術規則」第 19 條第 7 項中提及：

「計量模型分析法：蒐集相當數量具代表性之比較標的，透過計量模型分析，求出各主要影響價格因素與比較標的價格二者之關係式，以推算各主要影響價格因素之調整率及調整額之方法。」以及第 20 條：

「應用前條計量模型分析法應符合下列條件：

- 一、須蒐集應用計量模型分析關係式自變數個數五倍以上之比較標的。
- 二、計量模型分析採迴歸分析者，其調整後判定係數不得低於零點七。
- 三、截距項以外其他各主要影響價格因素之係數估計值同時為零之顯著機率不得大於百分之五。」。

此即大量估價(Mass Appraisal)之應用基本規定，本研究將依循之。

台灣目前的實價登錄資料內容符合大量估價資料的運用，經由原始資料標準化程序和統計預測模型，應可得到系統性的預測房價結果。

美國估價基金會擬定的標準估價作業準則(USPAP, Uniform Standards of Professional Appraisal Practice)，對大量估價的步驟與內涵定義為：(1)鑑定勘估財產，(2)界定與勘估財產相關的市場區域範圍，(3)界定該市場區域內影響價格的特徵，(4)建立可反映市場內影響價格的特徵模型結構，(5)發展模型分級(Calibration)用以判斷個別特徵影響價格的程度，(6)將模型所導出的推論運用於勘估財產的特徵，(7)檢視大量估計的結果。

依據美國估價基金會編撰之專業估價準則(USPAP)第六章(Standard 6)所規範之大量估價工具模型有：

1. 複迴歸(Multiple Regression Analysis)
2. 類神經網路(Artificial Neural Network)
3. 時間序列分析(Time Series Analysis)
4. 稅賦評價模型(Tax Assessment Value Model)

台灣於實價登錄制度實施後，不動產資料量與資料品質相較過去已有大幅改善，大量估價之實務運用可行性與準確度將可同步提升。

(二) Z-Score 標準差法

一組數據原始數值減去其平均數，再除以標準差後所得到的新數值，經過 Z 公式轉換後，具有平均數為 0，標準差為 1 的特性；當 Z 分數小於 0 時，表示該觀察值落在平均數以下，當 Z 分數大於 0，表示該觀察值落在平均數以上；數值越大，表示距離平均數越遠，若觀察值恰等於平均數，則 Z 分數為 0，此法又稱為標準差法。

(三) 複迴歸模型

複迴歸(多元迴歸)是用來探討一個應變數 Y 和多個自變數(X_1, X_2, \dots, X_n)的關係，整理複迴歸的表示式如(式 2.1)：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (\text{式 2.1})$$

β_0 為常數， $\beta_1 \dots \beta_n$ 為迴歸係數， ε 為誤差

迴歸分析之應變數與自變數皆需為數值，由取得的樣本，計算出迴歸方程式，透過迴歸方程式自變數 X_i 的係數，瞭解該自變數對應變數的影響力，也可以找出最大的影響變數，進行統計上的解釋，也可以用來預測應變數，以自變數的變動，估算應變數的改變量。

迴歸模型的顯著性檢定，一般都使用 F 檢定，F 檢定將所有自變數計算進來，觀察應變數 Y 和所有自變數 X_n 間是否有統計的顯著性。

判斷係數(Coefficient of Determination) R^2 用來解釋線性迴歸模式的適配度(Goodness of Fit)，說明整個模型的解釋力， $R^2=0$ 時，表示應變數(Y)與自變數(X_n)沒有線性關係， $R^2 \neq 0$ 時，表示應變數(Y)被自變數(X_n)所解釋的比率

(四) 類神經網路模型

類神經網路是模仿生物神經網路的結構和功能的運算模型，圖 2.1 以生物神經元和人工神經元來類比，生物神

經元經由多個樹突接受外界訊息，傳導給細胞核，若來源訊號足以觸發反應，則細胞核傳遞加工訊息至突觸，突觸將接受的加工訊息，輸出至多個神經元，各個神經元持續接受訊息、反應及輸出步驟，一直到最初的外來訊息獲得解決(或無解)時，才會結束；類神經元(人工神經元)是簡單的模擬生物神經元，它從外界環境(輸入設備)或者其它人工神經元取得資訊，並加以運算，之後輸出結果到外界環境(輸出設備)或者其它人工神經元。

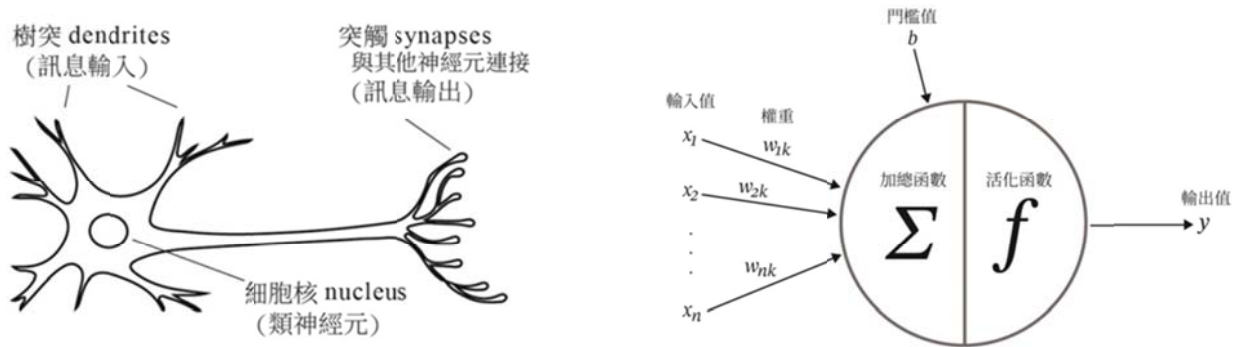


圖 2.1 生物神經元示意圖和人工神經元示意圖對比 (圖片來源：本研究整理)

x_1, x_2, \dots, x_n ：為神經元的輸入變數值(Input)

w_{ik} ：輸入變數的權重(Weights)， x_i 乘上 w_{ik} 就等於外部輸入的神經脈衝，對神經元 k 來說，所有的輸入訊號用加總函數(Σ)來表示，如(式 1)：

$$Net_k = \sum_{i=1}^n x_i w_{ik} \quad (1)$$

b ：為門檻值(Threshold)或偏誤值(Bias)， Net_k 必須大於門檻值，才能夠傳遞至神經元，否則輸入的神經脈衝就受到抑制。

$f(\cdot)$ ：為活化函數(Activation Function)，其目的是將加總函數 Net_k 的值做映射得到所需要的輸出向外傳遞，通常是非線性函數，常用的非線性活化函數。

Y ：為輸出(Output)，亦即所需要的結果。

類神經網路的學習過程是透過修正權重的方式，使得輸出值能夠接近實際值，網路的訓練重心在於如何自動的、有效率的調整權重大小以及學習速率，目前最常使用的方式是最陡坡降法(The Gradient Steepest Descent Method)，就是將誤差以可微分函數表示，透過微分斜率的梯度來產生權重的修正量，誤差為正向時(輸出值大於實際值)，降低神經權重，反之則增加神經權重，如此不斷反覆學習，讓誤差降低。

本論文採最廣為使用的監督式學習網路—倒傳遞類神經網路(Back-propagation Neural Network)。倒傳遞網路是由多層的神經元結構所構成，其中最外層接收輸入變數的稱之為輸入層(Input Layer)，而最後產生預測結果的神經元則稱之為輸出層(Output Layer)，而介於中間的神經元稱之為隱藏層(Hidden Layer)。

倒傳遞神經網路演算法的網路學習過程，包括了正向傳遞與反向傳遞兩個方向傳遞之演算。正向傳遞，就是資料由輸入層經過相關權重處理後傳遞至隱藏層，透過轉換函數可以計算每一神經元所對應的輸出值及其誤差函數的過程，當正向傳遞輸出層不能得到期望的輸出值時，則向後傳遞誤差訊息，修正各層神經元的權重和門檻值，此一過程稱之為反向傳遞，透過正向及反向來回運算，不斷的重複正向與反向傳遞運算藉以產生一組最佳之權重值，而網路的測試過程僅透過正向傳遞利用最佳權重值產生輸出值，圖 2.2 為倒傳遞神經網路學習步驟說明。

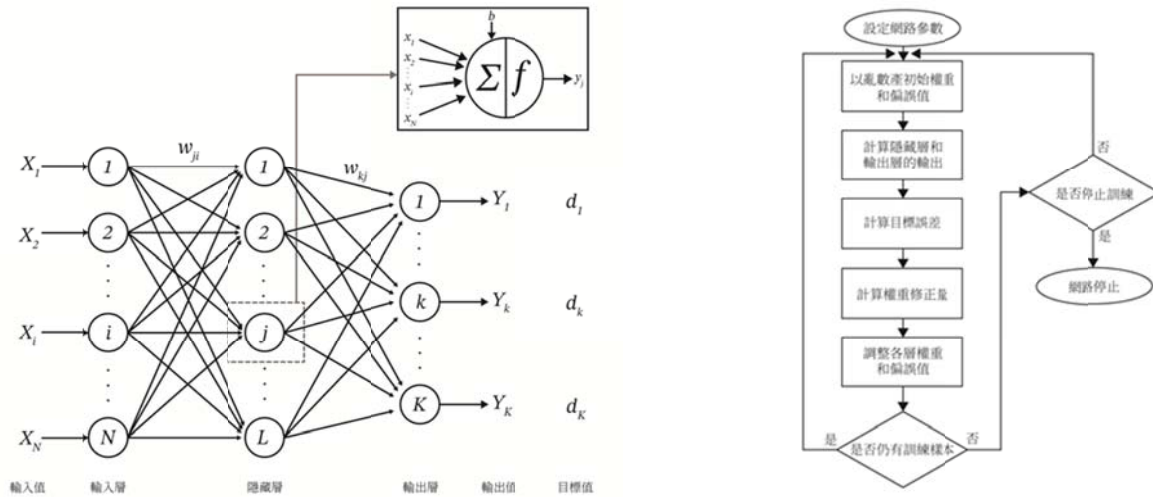


圖 2.2 倒傳遞神經網路演算法的網路架構及學習步驟說明 (圖片來源：本研究整理)

(五) 相關分析

以皮爾遜積差相關(Pearson Product Moment Correlation)方法分析兩者的相關程度，相關係數可作為連續變數 X、Y 之間線性相關的指標，稱皮爾森相關係數(Pearson's r)。

對樣本資料而言，皮爾森相關係數(Pearson Correlation Coefficient)的定義如(式 2)：

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} \quad (2)$$

r_{XY} = 樣本相關係數， S_{XY} = XY 共變異數， S_X = X 的樣本標準差， S_Y = Y 的樣本標準差

相關係數(R)的平方(R^2)可做為判斷係數或解釋變異量的比例。

第三章 資料來源和研究範圍

(一) 房價資料來源—內政部不動產交易實價查詢服務網

該網站(<http://lvr.land.moi.gov.tw/homePage.action>)提供的不動產買賣公開資料，下載年期為 2013 年第一季至 2016 年第四季，共四年、十六季的高雄市不動產成交案件實際資訊，下載的實價登錄資料檔案共有 28 個欄位，如表 3.3 實價登錄原始資料欄位及內容範例(資料來源：本研究整理)所示。

(二) 整體經濟指標資料來源

消費者物價指數、五大行庫平均房貸利率與 M1b 貨幣供給額數據來源為：

<http://pip.moi.gov.tw/V2/E/SCRE0201.aspx> 網站之「內政部不動產資訊平台 > 住宅統計 > 價格指標」，選取其中 M1b 貨幣供給額包括通貨淨額、支票存款以及活期存款與活期儲蓄存款等資料項。台股股價指數來源則為台灣經濟新報(TEJ)系統。

(三) 資料整合及分類

總計下載的原始資料筆數為 168,672 筆，篩選交易標的為「(土地+建物)+車位」或「(土地+建物)」的住宅用途資料筆數為 112,811 筆，漏失及補值資料處理之後，計有 103,738 筆資料供本研究使用。增加欄位『接近捷運站』，筆者以辦理購屋貸款為由，詢問玉山銀行、中國信託商銀、台北富邦銀行和國泰人壽保戶房貸等放款單位核貸人員，大致上對於認定住宅「接近捷運站」的概念距離約 1.5 公里左右，本研究採取較為保守的 1.2 公里為準，以建物的 GPS 座標計算出建物至高捷各站點的距離，房屋距離最近的捷運站小於 1.2 公里之案件，賦予數值 1，其他則為數值 0。增加欄位『建材類別』，欄位『主要建材』字串內含有「鋼筋混凝土」的交易案，賦予數值 1，沒有則為數值 0。完成上述計算補值，依欄位『建物型態』拆解出「集合性住宅」獨立檔案。

集合性住宅檔案增加欄位『建物型態編號』，欄位『建物型態』字串內含有「住宅大樓」賦予數值 1，「公寓」賦予數值 2，「華廈」賦予數值 3，「套房」賦予數值 4。

表 3.1 為 2013~2016 年間，高雄市集合性住宅成交案件基本統計量，表 3.2 為案件是否在捷運系統沿線範圍內的數量統計，統計數據顯示捷運沿線的案件量高於其他區段，而且平均單價也較高，而且有購買車位的案件，平均

單價高於未購買車位的案件。表 3.4 依各行政區的成交單價由高至低排序之基本統計。表 3.5 依各行政區的屋齡排序，屋齡數值為負值表示預售屋，數值為 0 表示建物落成當年成交，數據顯示舊有都會核心區三民、鳳山、苓雅、新興、前鎮、鹽埕等區屋齡相對較高，而近年重劃區域左營、鼓山、楠梓、仁武等原衛星城鎮則屋齡較低，說明高雄市住宅正由舊都會核心區向外圍擴張。

表 3.1 集合性住宅成交案件基本統計量

變數	平均值	標準差	最小值	最大值	眾數	案量	中位數
總價	6599169.86	6334219.70	17345	275880000	5000000	74580	5000000
單價	44862.18	18155.25	7814	114173.00	27780	74580	43062

表 3.2 集合性住宅捷運系統沿線成交案件單價比較

分析變數：單價					
靠近捷運站	案量	平均值	標準差	最小值	最大值
否	36527	42787.46	17425.28	7814.00	113764.00
是	38053	46853.71	18613.29	7818.00	114173.00

表 3.3 實價登錄原始資料欄位及內容範例(資料來源：本研究整理)

1	鄉鎮市區	三民區	鳳山區	鼓山區
2	交易標的	房地(土地+建物)	房地(土地+建物)	房地(土地+建物+車位)
3	土地區段位置或建物區門牌	高雄市三民區德翠路31-60號	高雄市鳳山區凱旋路332巷1-30號	高雄市鼓山區馬卡道路391~420號
4	土地移轉總面積平方公尺	6263	26.9	26.09
5	都市土地使用分區	住	住	住
6	非都市土地使用分區			
7	非都市土地使用編定			
8	交易年月日	1010706	1010719	1010725
9	交易筆棟數	土地1建物1車位0	土地1建物1車位0	土地1建物1車位1
10	移轉層次	全	二層	十三層
11	總樓層數	五層	五層	二十八層
12	建物型態	透天厝	公寓(5樓含以下無電梯)	住宅大樓(11層含以上有電梯)
13	主要用途		國民住宅	住家用
14	主要建材	鋼筋混凝土造	鋼筋混凝土造	鋼筋混凝土造
15	建築完成年月	680612	710513	1010111
16	建物移轉總面積平方公尺	277	7365	29191
17	建物現況格局-房	8	3	4
18	建物現況格局-廳	6	2	2
19	建物現況格局-衛	5	2	2
20	建物現況格局-隔間	有	有	有
21	有無管理組織	無	無	有
22	總價元	8150000	2000000	20300000
23	單價每平方公尺	29422	27155	69536
24	車位類別			坡道平面
25	車位移轉總面積平方公尺	0	0	25.5
26	車位總價元	0	0	0
27	備註			
28	編號	RPRNMLPKQHFFDE18CA	RPTNMLRLOQHFFBS3DA	RPSNMLTJQIHFFAE28CA

集合性住宅房價預估模型使用變數：車位、土地移轉總面積平方公尺、移轉層次、總樓層數、建物型態、主要建材、屋齡、建物移轉總面積平方公尺、格局-房、格局-廳、格局-衛、總價元、靠近捷運系統、消費者物價指數、五大行庫平均房貸利率、台股股價指數和 M1b 貨幣供給額,計 17 個變數,其中房屋總價為目標變數。

表 3.4 行政區集合性住宅成交案件單價比較

分析變數：單價					
行政區	案件量	平均值	標準差	最小值	最大值
鼓山區	9081	62134.48	17648.93	8216	113764
前金區	1535	52524.12	18963.88	10629	112963
左營區	9450	52223.81	15919.69	8247	113224

表 3.4 行政區集合性住宅成交案件單價比較(續)

分析變數：單價					
行政區	案件量	平均值	標準差	最小值	最大值
前鎮區	4790	47260.86	18105.56	7819	112054
新興區	2030	46994.07	18581.32	9926	106496
苓雅區	5220	46830.18	18911.47	8328	114173
仁武區	1101	45448.61	16007.98	7938	101339
三民區	12459	43730.65	15819.99	7882	113653
楠梓區	8672	39388.57	11922.58	7896	112764
鳳山區	11413	39245.12	13290.67	7971	114051

表 3.5 行政區集合性住宅成交案件屋齡比較

分析變數：屋齡					
行政區	案件量	平均值	標準差	最小值	最大值
橋頭區	295	7.427119	9.703808	-2	32
仁武區	1101	7.612171	11.51447	-2	37
鼓山區	9081	8.2383	9.361607	-3	44
楠梓區	8672	9.033441	11.71904	-3	45
左營區	9450	10.12434	8.687547	-2	46
鳳山區	11413	12.55402	12.56425	-1	47
鳥松區	787	15.15756	10.56041	-2	39
小港區	3207	15.40287	13.75503	-2	41
前鎮區	4790	15.78727	12.58312	-2	44
前金區	1535	15.87231	12.45977	-2	46
三民區	12459	15.91211	11.48756	-3	51

(四)模型建構

本研究資料運算分析工具使用 SAS 統計軟體。應用 SAS Enterprise Guide 於資料整理及基本統計，並以 SAS Enterprise Miner 於模型建構及資料分析，因為本研究不擬改變複迴歸與類神經網路演算法本身的結構，分析資料所用的複迴歸和類神經多層感知器(Multilayer Perceptron, MLP)之網路架構，兩模型的參數以 SAS 軟體之預設值為準。

研究選取的 17 個變數均源自政府實價登錄和公開資訊，其中總價為目標應變數，測試不同的自變數組合為輸入變數的多個網路，取其中預估精確度高的網路為結果；本研究不著力於說明輸入資訊與輸出預測值的關係，而且不擬對應變數和自變數間的關係作出特定假設，故選用類神經網路為主要的房價預測模型，SAS EM 類神經多層感知器網路架構的參數設定值如圖 3.1 所示。

本研究以 2013~2016 年間的交易案件為研究資料；分別以 2013~2015 三年的歷史資料推估 2016 房價，並以 2014~2015 二年的歷史資料推估 2016 房價，比較不同網路的推估精確度。第一個網路是 2013~2016 的資料全部投入，隨機打散分佈在訓練、驗證和測試三區，訓練和驗證資料內有 2016 的資料，運行的結果再去比對第二個網路和第三個網路之不含 2016 年資料的訓練模型結果，觀察預估的結果是否會有大幅差異；第二個網路是用 2013~2015 三年的歷史資料作為訓練和驗證，第三個網路則縮短為用 2014~2015 二年的歷史資料作為訓練和驗證；第二與第三個網路的測試資料集都是 2016 的房價資料，模型隱藏層的層數和神經元數量，沿用 SAS EM 的預設值。集合性住宅類神經網路資料分區設定與模型說明整理如表 3.6 與表 3.7 所示。

表 3.1 集合性住宅類神經網路資料分區設定

集合住宅類神經模型資料分區			
網路編號	樣本期間	樣本分割	測試
NET_1	2013~2016	40%訓練 30%驗證	30%測試
NET_2	2013~2015	60%訓練 40%驗證	2016 年 100%
NET_3	2014~2015	60%訓練 40%驗證	2016 年 100%

表 3.72 集合性住宅類神經網路模型說明

網路編號	說明
NET_1	2013~2016 所有資料均投入訓練驗證及測試,用以了解在訓練資料含有 2016 年的情形下,預估的精確度。
NET_2	以 2013~2015 三年的資料,作為訓練及驗證,來預估 2016 的房價。
NET_3	以 2014~2015 二年的資料,作為訓練及驗證,來預估 2016 的房價,對比 NET_2 的預估結果。

鑑於類神經網路演算方法沒有變數選擇能力,無論投入變數對目標變數是否具有顯著影響性,只要投入變數,類神經網路皆會將其帶入模型,進行各項權數訓練過程,因此為避免太多變數投入類神經網路,產生過度配適的模型結果,先藉由迴歸分析變數選擇方法,進行重要變數挑選,再將影響顯著的變數投入類神經網路中。

SAS EM 複迴歸的參數設定值如圖 3.2 所示。

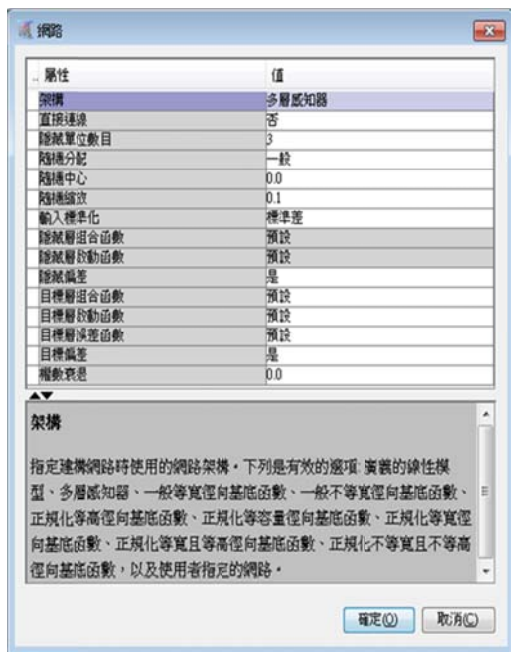


圖 3.1 SAS EM 類神經網路參數設定



圖 3.2 SAS EM 複迴歸的參數設定值參數設定

(資料來源：SAS EM 類神經網路架構設定畫面)

(五) 預估結果的評估方法

以 SAS EM 執行類神經網路時,會在檔案中增加一個欄位,存入模型運算的推估值;本研究利用 SAS EG 的皮爾森相關(Pearson Correlation)相關分析,探討樣本實際總價(X)和模型推估總價(Y)兩項變數之間的關聯性,皮爾森相關係數(r)除了可以看出兩變數間的正負相關外,還可以看出兩者之間直線關係的強度。

r 的正負號代表著 X 與 Y 的相關性,如果 $r > 0$,表示 X 和 Y 為正相關,亦代表 Y 值會隨 X 值變大而增大;反之,如果 $r < 0$,表示 X 和 Y 為負相關,亦代表 Y 值會隨 X 值變大而縮小,其中 $|r| \leq 1$ 。若 $|r| = 1$ 則表示 X 和 Y 在一直線上, $|r|$ 越靠近 1 表示 X 和 Y 的直線關係越強;反之,若 $|r|$ 越靠近 0 則表示 X 和 Y 的直線關係越弱。

相關係數的解釋依賴於具體的應用背景和目的,如果是應用在使用高性能的儀器來驗證一個物理定律,0.9 的相關係數可能是很低的,但如果是應用在社會科學中,由於社會科學受到各種複雜多變因素影響,0.9 的相關係數

是相當高的。

第四章 實證分析

(一) 高雄市全區域

2013~2016 年間，高雄市集合性住宅成交案件的總價平均值約為 584.9 萬元，標準差約為 425.3 萬元，+3 個標準差的最高總價約 2560 萬，統計量如表 4.1。表 4.2 為依案件有無購買車位分類的基本統計，顯示出有車位的案件平均總價高於高雄全區域的平均總價，而未購買車位的案件總價則低於高雄全區域的平均總價，且有車位的案件平均屋齡也遠低於無車位的案件屋齡，可見車位是新建案銷售的重要因素之一，另外政府將增設車位作為容積獎勵的手段，早期沒有這個獎勵因素，加上消費者對車位不重視，所以老屋的集合式產品多沒有車位，且依照登記方式與交易習慣多併入交易總價計算，新建案則把車位單獨出來，交易時再予以計入總價，有車位的案件，購買的平均建坪數比沒購買車位的案件建坪高約 67.5 平方公尺(約 20.4 坪)。

表 4.1 高雄市全區域基本統計量

2013~2016 高雄市全區去極端值後總計 71823 筆資料				
變數	平均值	標準差	最小值	最大值
總價	5849571.85	4253614.03	17345.00	25600000.00
單價	43846.54	16970.96	7814.00	99263.00
屋齡	13.7244894	11.9364343	-3.0000000	49.0000000
建坪	126.3657018	57.3741074	1.1900000	383.0700000

表 4.2 高雄市全區依車位分類基本統計量

車位	資料筆數	變數	平均值	標準差	最小值	最大值
有	34020	總價	8440158.63	4517418.57	300000.00	25600000.00
		單價	50956.84	15550.19	7898.00	99220.00
		屋齡	6.4032628	7.7664751	-3.0000000	41.0000000
		建坪	161.9004186	52.3243627	13.1200000	382.3400000
無	37803	總價	3518228.77	2129886.68	17345.00	25000000.00
		單價	37447.77	15593.54	7814.00	99263.00
		屋齡	20.3130704	11.1701443	-3.0000000	49.0000000
		建坪	94.3869947	40.3898436	1.1900000	383.0700000

使用監督式學習模型(例如決策樹和類神經演算法)作預估計算的特性，在於存在目標變數用以訓練模型，因此模型評估要求精確度；模型經由歷史資料學習，用以預估未知目標變數。為了解所設計模型預估未知目標變數時的精確程度，本研究先將所有資料，區分成 40%的訓練資料、30%的驗證資料和 30%的測試資料。訓練和驗證資料不包括 2016 年的資料，而將 2016 年資料用來測試模型的預測精確度。表 4.3 說明類神經網路資料分區的資料筆數分配。

高雄市全區域 2013~2016 年間的資料經過標準化、去除極端值之後，總計有 71,823 筆資料可供使用，將全部 17 個變數帶入 SAS EM 中，先利用迴歸分析過濾選擇變數，避免太多變數投入類神經網路，產生過度配適的模型結果。

迴歸的結果顯示顯著相關的變數有：是否靠近捷運系統、台股股價指數、土地移轉總面積、總樓層數、建物型態、M1b、屋齡、房間數、建物面積、衛浴數量等 10 個自變數。

將迴歸分析選取的變數分別投入三個不同學習期間的類神經網路學習，使用 SAS EG 的皮爾森相關(Pearson Correlatio)對模型運算的推估值進行相關分析，探討樣本實際總價(X)和模型推估總價(Y)兩項變數之間的關聯性。第一個網路訓練資料含 2016 年資料，實際值(X)和測試預估值(Y)的皮爾森相關數為 0.92940，表 4.4 為第一個網路計算結果統計。相同的變數帶入第二個和第三個模型，訓練資料不含 2016 年，分別以前二年的資料和前三年的資

料訓練網路，以預估 2016 年的房價。用前二年的資料訓練並用以推估 2016 年的房價，實際總價(X)和 2016 測試預估值(Y)的皮爾森相關係數為 0.92715；用前三年的資料訓練並用以推估 2016 年的房價，實際值(X)和 2016 測試預估值(Y)的皮爾森相關係數(r)為 0.92740。兩者與使用全部資料訓練網路的皮爾森相關係數 0.92940 差距不大，但使用三年資料比二年資料的訓練資料模型結果好一些，結果統計量如表 4.5 所示。

高雄市全區域的類神經網路估價結論如下：不含 2016 年資料的訓練模型的精確度與含 2016 年的訓練模型的精確度不會相去太多，而且訓練資料三年期的模型優於訓練資料二年期的模型。總結高雄市全區域預測房價的模型，以三年期訓練資料模型之判斷係數 R^2 等於 0.86 (0.92740×0.92740)較佳。

表 4.3 類神經網路資料分區

高雄市全區去極值後總計 71823 筆資料			
類神經模型使用資料	訓練資料筆數	驗證資料筆數	測試資料筆數
2013~2016 全部資料	28729	21547	21547
2013~2015 估 2016	33889	22593	15160
2014~2015 估 2016	20760	13840	15160

表 4.4 高雄全區域預估房價訓練模型結果統計

高雄市全區域預估房價訓練模型結果簡單統計 測試資料 21547 筆資料					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
測試資料實際總價(X)	-0.11828	0.67057	-1.03909	2.99971	
模型預估結果(Y)	-0.11994	0.62385	-1.12423	2.39990	0.92940

表 4.5 高雄市全區域預估 2016 房價模型結果統計

高雄市全區域預估 2016 房價模型結果簡單統計 測試資料 15160 筆					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
2016 測試實際總價(X)	-0.02903	0.68885	-1.02604	2.99971	
2013~2015 估 2016(Y)	-0.00485	0.64540	-0.98176	3.38524	0.92740
2014~2015 估 2016(Y)	-0.03144	0.63658	-0.97162	3.25931	0.92715

(二) 三民區

本文取交易頻繁之前五個行政區域，依成交量及成交金額所占比率，依序為三民區、鳳山區、左營區、鼓山區、楠梓區，執行相同的模型，期望分區估價能提高模型的精確度。三民區為高雄縣市合併前人口最多的行政區，因緊鄰原高雄市中心，所以發展甚為迅速，高雄市最早的市地重劃即是從後火車站開始發展，三民區開發亦以輻射方式向外發展，目前則以與鳳山區交界本館(高應大週邊)為主要發展地段，表 4.6 的基本統計量顯示三民區的平均房價和高雄市全區域的平均房價接近，近四年的成交案件量係全高雄市最多，資料最具代表性，選擇作為帶入模型研究測試的首選行政區。表 4.7 說明購買車位的案件平均比未購買車位的案件多約 68.5 平方公尺(約 20.7 坪)的建坪，和高雄市全區域的統計量相近。三民區的資料經過標準化、去除極端值之後，總計有 12,065 筆資料可供使用，資料分區如表 4.8 所示，利用迴歸分析變數選擇方法的結果顯示，顯著相關的變數有：是否靠近捷運系統、土地移轉總面積、總樓層數、建物型態、M1b、屋齡、房間數、建物面積、衛浴數量等 9 個自變數，取上述 9 個變數帶入 CART 決策樹和類神經網路，並比較三者的預估效果。CART 決策樹的預估結果(Y)和實際值(X)皮爾森相關係數為 0.94593，比第一個類神經網路結果的皮爾森相關係數 0.95859 低，而且決策樹資料的屬性是類別型態，而房價變數的屬性是連續型態數據，必須先作離散化過後才可以開始建立決策樹，結果如圖 4.1 所示，可觀察到同一個規則的預測房價是箱型的，相較之下類神經網路是比較理想的預估模型，因此後續僅採用類神經網路模型不再測試 CART 決策樹模型，測試其餘四個行政區。

表 4.6 三民區的基本統計量

三民區去極值後總計 12065 筆資料				
變數	平均值	標準差	最小值	最大值
總價	5393582.10	3818693.19	150000.00	25500000.00
單價	43211.50	15082.24	7882.00	97928.00
屋齡	16.0697886	11.4161226	-3.0000000	49.0000000
建坪	119.1706067	59.6041852	10.2700000	383.0700000

表 4.7 三民區依車位分類的基本統計量

車位	資料筆數	變數	平均值	標準差	最小值	最大值
有	5133	總價	8043138.96	4007103.10	300000.00	25500000.00
		單價	50185.43	14207.98	11796.00	97928.00
		屋齡	8.8885642	8.5469254	-3.0000000	41.0000000
		建坪	158.5256400	56.7511333	13.1200000	382.3400000
無	6932	總價	3431641.04	2108074.87	150000.00	23000000.00
		單價	38047.44	13556.45	7882.00	97060.00
		屋齡	21.3873341	10.3094734	-3.0000000	49.0000000
		建坪	90.0290335	42.4568876	10.2700000	383.0700000

表 4.8 三民區類神經網路模型資料分區

三民區去極值後總計 12065 筆資料			
類神經模型	訓練資料筆數	驗證資料筆數	測試資料筆數
2013~2016 全部資料	4826	3620	3619
2013~2015 估 2016	9351	5611	2714
2014~2015 估 2016	3343	2229	2714

針對三民區的分析中，第一個含 2016 年資料的類神經網路的預估結果(Y)，和實際值(X)皮爾森相關係數為 0.95859，比同模型的高雄全區域預估皮爾森相關係數 0.92715 明顯提高，二年期訓練資料及三年期訓練資料模型結果顯示，預估數值(Y)，和實際值(X)皮爾森相關係數分別為 0.94502 和 0.94976，三年期的模型仍然優於二年期訓練模型。總結三年期訓練資料模型之判斷係數 R^2 等於 0.90 (0.94976 × 0.94976) 較佳。

圖 4.2 為三民區三年期訓練資料模型的實際房價(X)對預測房價(Y)散佈圖，二者的相關係數為 0.9498，圖中散佈在橢圓範圍內是實際總價和預測總價相關較強的 95% 的點，顯示模型對於三民區房屋總價的預估值，比較精確的範圍約在+1 個標準差以下。

表 4.9 三民區預估房價訓練模型結果統計

三民區預估房價訓練模型結果簡單統計 測試資料 3619 筆資料					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
測試資料實際總價(X)	-0.18882	0.60881	-1.01815	2.89709	
決策樹估算值(Y)	-0.19082	0.58520	-0.90700	2.19738	0.94593
全部資料類神經(Y)	-0.19129	0.58455	-0.97361	2.48665	0.95859

表 4.10 三民區預估 2016 房價模型結果統計

三民區預估 2016 房價模型結果簡單統計 測試資料 2714 筆					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
2016 實際標準化總價	-0.07558	0.60472	-0.97079	2.65555	
2013~2015 估 2016	-0.01856	0.59617	-0.91699	2.52018	0.94976
2014~2015 估 2016	0.06813	0.65572	-0.9975	2.63635	0.94502

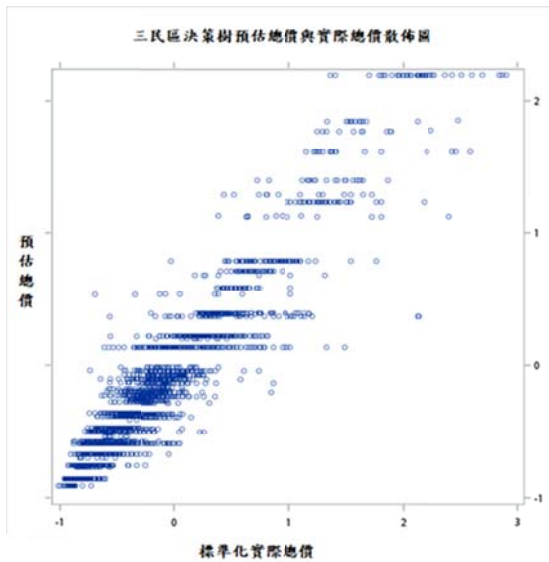


圖 4.1 三民區決策樹模型 2016 預估房價

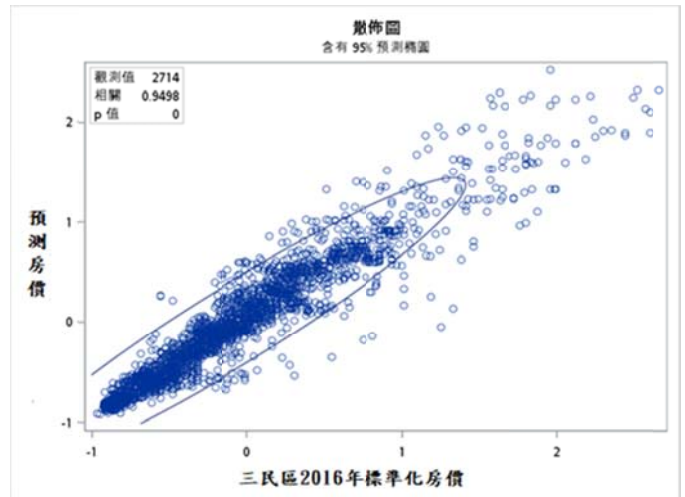


圖 4.2 三民區決策樹模型 2016 預估房價

(三) 鳳山區

鳳山區在縣市合併後是大高雄市的地理位置中心，即取代三民區成為高雄市人口密度最高的行政區，晚近在由於捷運紅線貫穿並串聯高雄市區與鳳山核心，加以大型公共建設(縱貫鐵路地下化、市議會遷址至鳳山、衛武營、大東文化中心之開發、眷村改建)以及隣近澄清湖地區等市地重劃等條件加持下，其房地開發與交易甚為熱絡，原鳳山舊市區幾近飽和沒有新建案和重劃區大量的新建房屋，形成強烈對比，同時反映在依車位分類的表 4.12 統計量上面，案件有無車位的建坪大約相差 57.6 平方公尺(17.4 坪)，但總價、單價、屋齡卻相差甚多。

鳳山區的資料經過標準化、去除極端值之後，總計有 11,303 筆資料可供使用，資料分區如表 4.13 所示，利用迴歸分析變數選擇方法的結果顯示顯著相關的變數有：是否靠近捷運系統、車位、土地移轉總面積、移轉樓層、總樓層數、建物型態、M1b、屋齡、建物面積等 9 個自變數，據以帶入三種類神經網路模型。

表 4.11 鳳山區的基本統計量

鳳山區去極值後總計 11303 筆資料				
變數	平均值	標準差	最小值	最大值
總價	5227097.68	2759539.41	300000.00	24900000.00
單價	39143.69	13016.87	7971.00	98814.00
屋齡	12.5214545	12.5636886	-1.0000000	47.0000000
建坪	131.9984650	46.4230782	15.6000000	376.8100000

表 4.12 鳳山區依車位分類的基本統計量

車位	資料筆數	變數	平均值	標準差	最小值	最大值
有	6196	總價	6746693.36	2564929.67	1620000.00	24900000.00
		單價	43089.47	12018.79	11392.00	95526.00
		屋齡	4.6941575	7.1818243	-1.0000000	35.0000000
		建坪	158.0264187	36.2791506	48.7700000	376.8100000
無	5107	總價	3383468.38	1634621.04	300000.00	22800000.00
		單價	34356.52	12568.79	7971.00	98814.00
		屋齡	22.0178187	11.0573708	-1.0000000	47.0000000
		建坪	100.4203955	36.7958036	15.6000000	374.8900000

表 4.13 鳳山區類神經網路資料分區

鳳山區去極值後總計 11303 筆資料			
類神經模型	訓練資料筆數	驗證資料筆數	測試資料筆數
2013~2016 全部資料	4521	3391	3391
2013~2015 估 2016	5398	3598	2307
2014~2015 估 2016	3337	2224	2307

針對鳳山區的分析中，第一個含 2016 年資料的類神經網路的預估結果(Y)，和實際值(X)皮爾森相關係數為 0.91956，比同模型的高雄全區域預估皮爾森相關係數 0.92715 低。二年期訓練資料及三年期訓練資料模型結果顯示，預估數值(Y)和實際值(X)皮爾森相關係數分別為 0.92161 和 0.90203，反而是二年期訓練模型優於三年期的模型，和高雄全區域、三民區的模型結果大不相同。總結鳳山區房價預估模型，以二年期訓練資料模型之判斷係數 R^2 等於 0.85 (0.92161×0.92161) 較佳。圖 4.3 為鳳山區二年期訓練資料模型的實際房價(X)對預測房價(Y)散佈圖，二者的相關係數為 0.92161，散佈在橢圓範圍內的較強相關 95% 的點，顯示模型對於鳳山區房屋總價的預估值，也是在房價的 +1 個標準差之內。

表 4.14 鳳山區預估房價訓練模型結果統計

鳳山區預估房價訓練模型結果簡單統計 測試資料 3391 筆資料					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
測試資料實際總價(X)	-0.21447	0.43600	-0.99447	2.55767	
全部資料類神經(Y)	-0.21864	0.39564	-0.79640	2.28028	0.91956

表 4.15 鳳山區預估 2016 房價模型結果統計

鳳山區預估 2016 房價模型結果簡單統計 測試資料 2307 筆					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
2016 實際標準化總價	-0.13770	0.48021	-0.97868	2.77869	
2013~2015 估 2016	0.01657	0.55560	-0.80087	2.64001	0.90203
2014~2015 估 2016	-0.07297	0.43472	-0.89709	2.39310	0.92161

(四) 鼓山區

鼓山區內凹仔底(高雄美術館)等鄰近左營區拜高雄市地重劃以及農 16 區段徵收之賜，一舉蛻變為高雄新興的住宅區段，近三年來配合高雄市政府的容積獎勵措施以及鐵路地下化建設，迅速改變了區域內的土地開發利用，由早期的高級透天建築推向超高樓層的高級住宅大樓，新建案如雨後春筍，快速佈滿天際線。表 4.17 顯示，舊建物的買賣數量遠不及新建案買賣數量，使得鼓山區的屋齡在目前高雄都會區裡的屋齡是最低的。鼓山區的資料經過標準化、去除極端值之後，總計有 8,159 筆資料可供使用，資料分區如表 4.18 所示，利用迴歸分析變數選擇方法的結果顯示顯著相關的變數有：是否靠近捷運系統、土地移轉總面積、移轉樓層、總樓層數、建物型態、M1b、屋齡、建物面積等 8 個自變數，沒有選擇車位變數，表 4.17 顯示案件有無車位的建坪差 86.3 平方公尺(約 26 坪)。

表 4.16 鼓山區的基本統計量

鼓山區去極值後總計 8159 筆資料				
變數	平均值	標準差	最小值	最大值
總價	9371096.69	5850430.27	100000.00	25600000.00
單價	59532.00	15809.46	8216.00	99220.00
屋齡	8.9620051	9.4731153	-1.0000000	44.0000000
建坪	149.5599412	69.8677670	6.2700000	381.2600000

表 4.17 鼓山區依車位分類的基本統計量

車位	資料筆數	變數	平均值	標準差	最小值	最大值
有	5081	總價	12136812.30	5629465.56	1571898.00	25600000.00
		單價	65020.66	12679.34	12921.00	99220.00
		屋齡	5.1068687	5.6475723	-1.0000000	31.0000000
		建坪	182.1451761	65.6857753	40.8100000	381.2600000
無	3078	總價	4805599.28	2223747.67	100000.00	20950000.00
		單價	50471.61	16290.46	8216.00	99040.00
		屋齡	15.3258609	10.9642691	0	44.0000000
		建坪	95.7699545	34.2251529	6.2700000	359.5600000

表 4.18 鼓山區類神經網路資料分區

鼓山區去極值後總計 8159 筆資料			
類神經模型	訓練資料筆數	驗證資料筆數	測試資料筆數
2013~2016 全部資料	3264	2448	2447
2013~2015 估 2016	3860	2573	1716
2014~2015 估 2016	2011	1304	1716

鼓山區第一個含 2016 年資料的類神經網路的預估結果(Y)，和實際值(X)皮爾森相關係數為 0.97336，比同模型的高雄全區域預估皮爾森相關係數 0.92715 高。二年期訓練資料及三年期訓練資料模型結果顯示，預估數值(Y)和實際值(X)皮爾森相關係數分別為 0.97717 和 0.97433，二年期訓練模型優於三年期訓練模型模型，和高雄全區域與三民區的模型結果不相同，但差距很小，幾乎相同。總結鼓山區房價預的模型，二年期訓練資料模型之判斷係數 R2 等於 0.95 (0.97717 × 0.97717) 較佳。圖 4.4 為鼓山區二年期訓練資料模型的實際房價(X)對預測房價(Y)散佈圖，二者的相關係數為 0.97717，圖中散佈在橢圓範圍內的較強相關 95% 的點，顯示模型對於鼓山區房屋總價的預估效果良好，可以涵蓋到接近房價的 +3 個標準差。

表 4.19 鼓山區預估房價訓練模型結果統計

鼓山區預估房價訓練模型結果簡單統計 測試資料 2447 筆資料					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
測試資料實際總價(X)	0.44212	0.90681	-0.99412	2.95273	
全部資料類神經(Y)	0.44680	0.92616	-1.00721	2.99971	0.97336

4.20 鼓山區預估 2016 房價模型結果統計

鼓山區預估 2016 房價模型結果簡單統計 測試資料 1716 筆					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
2016 實際標準化總價	0.66036	0.96750	-1.02604	2.99971	
2013~2015 估 2016	0.59574	0.93197	-1.34308	2.73647	0.97433
2014~2015 估 2016	0.68532	0.96131	-0.92740	2.96954	0.97717

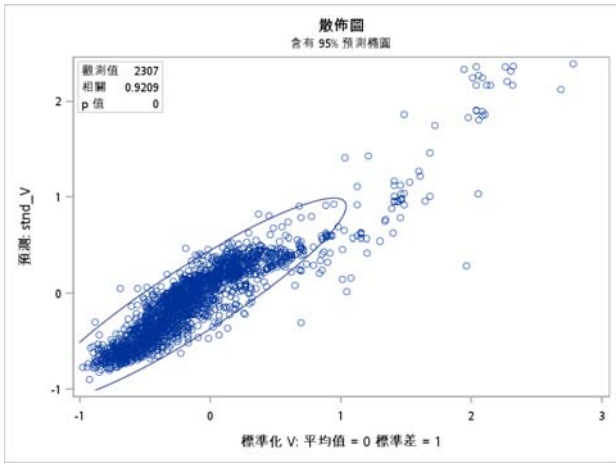


圖 4.3 鳳山區 2016 預估房價與實際房價散佈圖

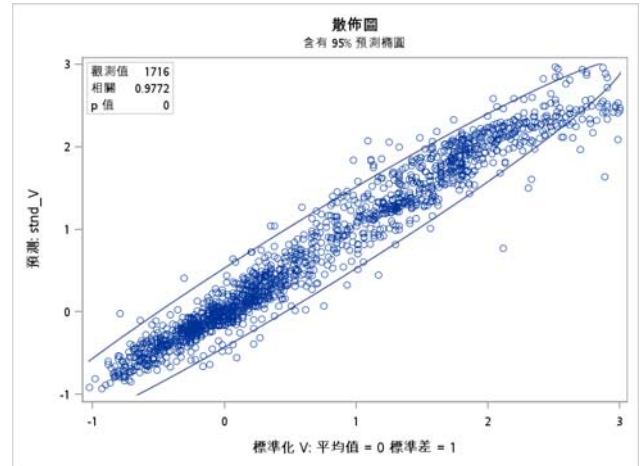


圖 4.4 鼓山區 2016 預估房價與實際房價散佈圖

(五)左營區

左營區新社區則因與鼓山凹仔底重劃區以及農 16 區段徵收區相連，成為左營區房地發展之重點，目前在新社區內主要有高鐵、捷運以及鐵路地下化等重大建設，且因為人口快速進駐，形成了新的北高雄商圈，其腹地與南高雄的三多、中央公園商圈相較而言絲毫不遜色，重劃區大量的新建案，使得屋齡統計值相對拉低，是目前高雄市房屋總價最高的行政區，有無車位的建坪大約相差 55 平方公尺(約 16.6 坪)。

左營區的資料經過標準化、去除極端值之後，總計有 9,140 筆資料可供使用，資料分區如表 4.23 所示，利用迴歸分析變數選擇方法的結果顯示顯著相關的變數有：是否靠近捷運系統、土地移轉總面積、移轉樓層、總樓層數、建物型態、M1b、屋齡、建物面積等 8 個自變數，和鼓山區相同。

表 4.21 左營區的基本統計量

左營區去極值後總計 9140 筆資料				
變數	平均值	標準差	最小值	最大值
總價	7358087.09	4294917.51	500000.00	25500000.00
單價	51426.10	15173.74	8247.00	99203.00
屋齡	10.3794311	8.6145286	-2.0000000	46.0000000
建坪	139.6929978	55.7290326	24.7200000	381.8500000

表 4.22 左營區依車位分類的基本統計量

車位	資料筆數	變數	平均值	標準差	最小值	最大值
有	5560	總價	8981771.49	4502220.20	680000.00	25500000.00
		單價	54668.37	14242.54	12372.00	99203.00
		屋齡	8.2996403	7.5579288	-2.0000000	27.0000000
		建坪	161.1499227	54.4746623	30.8600000	381.8500000
無	3580	總價	4836387.31	2272126.99	500000.00	24500000.00
		單價	46390.60	15203.91	8247.00	96280.00
		屋齡	13.6094972	9.1441848	-1.0000000	46.0000000
		建坪	106.3688352	38.6686074	24.7200000	381.4200000

表 4.23 左營區類神經網路資料分區

左營區去極值後總計 9140 筆資料			
類神經模型	訓練資料筆數	驗證資料筆數	測試資料筆數
2013~2016 全部資料	3656	2742	2742
2013~2015 估 2016	4545	3030	1565
2014~2015 估 2016	2790	1860	1565

左營區第一個含 2016 年資料的類神經網路的預估結果(Y)和實際值(X)的皮爾森相關係數為 0.95361，比同模型的高雄全區域預估皮爾森相關係數 0.92715 高。二年期訓練資料及三年期訓練資料模型結果顯示，預估數值(Y)和實際值(X)的皮爾森相關係數分別為 0.93365 和 0.93185，也是二年期訓練模型優於三年期的模型，和鼓山區的模型結果相同也是差距很小，幾乎相同。總結左營區房價預的模型，以二年期訓練資料模型之判斷係數 R^2 等於 $0.87(=0.93365 \times 0.93365)$ 較佳。圖 4.5 為左營區二年期訓練資料模型的實際房價(X)對預測房價(Y)散佈圖，二者的相關係數為 0.93365，散佈在橢圓範圍內的較強相關 95%的點，顯示模型對於左營區房屋總價的預估值效果也不錯，涵蓋範圍接近 +2 個標準差的房價範圍。

表 4.24 左營區預估房價訓練模型結果統計

左營區預估房價訓練模型結果簡單統計 測試資料 2447 筆資料					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
測試資料實際總價(X)	0.11579	0.67115	-0.93447	2.98392	
全部資料類神經(Y)	0.11107	0.63774	-1.04594	2.76385	0.95361

表 4.25 左營區預估 2016 房價模型結果統計

左營區預估 2016 房價模型結果簡單統計 測試資料 1565 筆					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
2016 實際標準化總價	0.15368	0.64393	-0.91553	2.90499	
2013~2015 估 2016	0.17704	0.64954	-0.79799	3.07483	0.93185
2014~2015 估 2016	0.15297	0.62074	-1.18313	2.70519	0.93365

(六)楠梓區

楠梓與仁武區相鄰的地區原是高雄市邊陲，早期因受煉油廠以及仁大工業區之影響，房地發展甚為遲緩，晚近因為汙染之改善、加工處口區轉型、五輕停工、高雄大學設立以及捷運等因素，現在的楠梓區相較於整個高雄市房地發展區域而言約當屬於所謂的蛋白區，一躍而為新興之房地發展區域，近年才有大量的新建集合住宅，大體來說新屋較多，屋齡也低。楠梓區的資料經過標準化、去除極端值之後，總計有 8,604 筆資料可供使用，資料分區如表 4.28 所示，利用迴歸分析變數選擇方法的結果顯示顯著相關的變數有：車位、股價指數、土地面積、建物型態、主要建材、M1b、屋齡、建物面積、衛浴數等 9 個自變數。

表 4.26 楠梓區的基本統計量

楠梓區去極值後總計 8604 筆資料				
變數	平均值	標準差	最小值	最大值
總價	4800041.84	2413478.94	300000.00	19400000.00
單價	39378.14	11767.75	7896.00	93759.00
屋齡	8.9575779	11.6951351	-3.0000000	45.0000000
建坪	119.1586192	42.5007120	10.8100000	372.5800000

表 4.27 楠梓區依車位分類的基本統計量

車位	資料筆數	變數	平均值	標準差	最小值	最大值
有	4506	總價	6324664.22	1972875.17	430000.00	19400000.00
		單價	43579.67	9003.85	9400.00	93493.00
		屋齡	3.3224589	6.5112635	-3.0000000	27.0000000
		建坪	145.2202308	31.0038365	30.2200000	372.5800000

表 4.28 楠梓區依車位分類的基本統計量(續)

車位	資料筆數	變數	平均值	標準差	最小值	最大值
無	4098	總價	3123626.89	1607430.37	300000.00	18000000.00
		單價	34758.29	12683.77	7896.00	93759.00
		屋齡	15.1537335	12.9329691	-3.0000000	45.0000000
		建坪	90.5022938	34.1699125	10.8100000	363.6100000

表 4.29 楠梓區類神經網路資料分區

楠梓區去極值後總計 9140 筆資料			
類神經模型	訓練資料筆數	驗證資料筆數	測試資料筆數
2013~2016 全部資料	3442	2581	2581
2013~2015 估 2016	4166	2778	1660
2014~2015 估 2016	2682	1788	1660

楠梓區第一個含 2016 年資料的類神經網路的預估結果(Y)，和實際值(X)皮爾森相關係數為 0.94202，比同模型的高雄全區域預估皮爾森相關係數 0.92715 高。二年期訓練資料及三年期訓練資料模型結果顯示，預估數值(Y)和實際值(X)皮爾森相關係數分別為 0.90644 和 0.92589，三年期訓練模型優於二年期的模型。總結楠梓區房價預的模型，以三年期訓練資料模型之判斷係數 R^2 等於 $0.86(=0.92589 \times 0.92589)$ 較佳。圖 4.6 為楠梓區三年期訓練資料模型的實際房價(X)對預測房價(Y)散佈圖，二者的相關係數為 0.92589，散佈在橢圓範圍內的較強相關 95% 的點，顯示模型對於楠梓區房屋總價預估較精確的範圍，大約是在房價的 +1 個標準差之內。

表 4.30 楠梓區預估房價訓練模型結果統計

楠梓區預估房價訓練模型結果簡單統計 測試資料 2581 筆資料					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
測試資料實際總價(X)	-0.28452	0.38706	-0.99447	2.02090	
全部資料類神經(Y)	-0.28922	0.36236	-1.03462	1.37050	0.94202

表 4.31 楠梓區預估 2016 房價模型結果統計

楠梓區預估 2016 房價模型結果簡單統計 測試資料 1660 筆					
變數	平均值	標準差	最小值	最大值	測試資料與估算資料相關係數
2016 實際標準化總價	-0.19076	0.41240	-0.97789	2.02090	
2013~2015 估 2016	-0.27888	0.34057	-0.99530	1.97179	0.92589
2014~2015 估 2016	-0.26443	0.32600	-0.98515	1.48513	0.90644

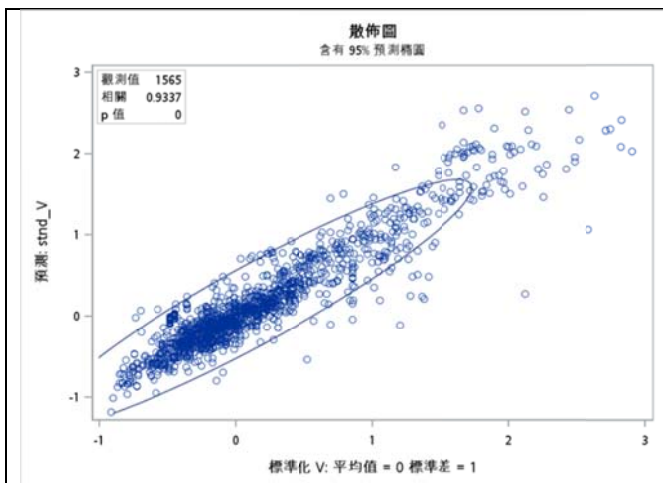


圖 4.1 左營區 2016 預估房價與實際房價散佈圖

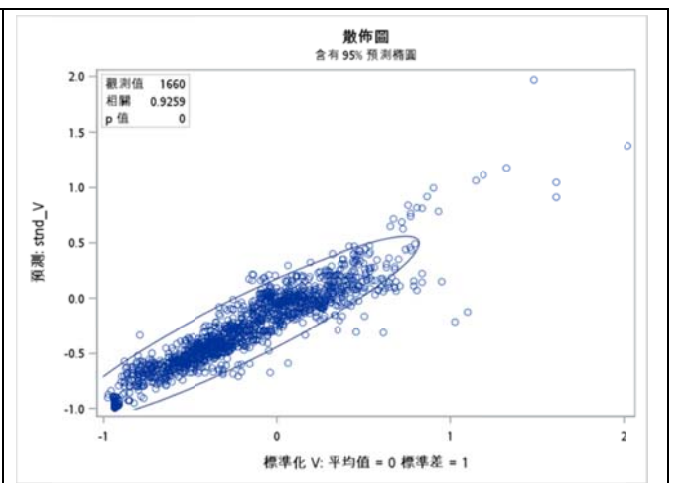


圖 4.2 楠梓區 2016 預估房價與實際房價散佈圖

(七) 集合性住宅實證總結

集合性住宅的類神經網路房價預測模型，依據表 4.31 顯示，三年期訓練資料比二年期訓練資料精確，區域範圍較小比全區域精確，推論是因為小區域範圍的資料同質性高，資料訓練的結果較精確。根據預估房價與實際房價的判斷係數和散佈圖顯示，高雄市全區預測較精確的房價總價範圍約在 1,500 萬元以下，三民區約在 1,400 萬元以下，鳳山區約在 1,300 萬元以下，鼓山區約在 2,560 萬元以下，左營區約在 1,900 萬元以下，楠梓區約在 1,300 萬元以下。

表 4.32 集合性住宅 2016 房價推估模型結果

行政區	相關係數較高的模型	判斷係數	預測較精確的房價範圍
高雄市	三年訓練資料	0.86	1,500 萬以下
三民區	三年訓練資料	0.90	1,400 萬以下
鳳山區	二年訓練資料	0.85	1,300 萬以下
鼓山區	二或三年訓練資料相近	0.95	2,560 萬以下
左營區	二或三年訓練資料相近	0.87	1,900 萬以下
楠梓區	三年訓練資料	0.86	1,300 萬以下

第五章 研究結論

由三種不同類神經網路模型實證結果發現，高雄市全區的預測值精確度不及各分區模型，顯示縮小區域範圍使得不動產同質性高，並拉長訓練資料期間，為影響預估精確度的重大因子。自變數的選取則可先以多元迴歸演算法所推算之相關係數較高者，作為類神經演算的模型自變數，再經由反覆調整自變數個數、不斷運算來提升預估的精確度。

大量估價在實價登錄制度實施之前，除難以取得具公信力的交易資料之外，電腦的軟硬體設備也有一定門檻，所以主要應用於公部門的財評或銀行業者放款估價參考，但公部門和銀行的估價結果礙於資料隱私，也無法公諸大眾。

本研究證實使用源自內政部提供的公開資料，實價登錄資料及影響不動產價格的相關經濟指標，運用統計軟體的倒傳遞類神經網路方法預測房價，在自住型集合住宅判斷係數可達 0.85 以上，顯示實價登錄實施的時間雖僅有四年餘，但自住型集合住宅所提供的成交資料，已經足供類神經網路模型作推算，隨著資料量持續累積，將逐漸提升不動產交易資訊透明度，有利於計劃購屋者的決策訂定。

本研究中，整體經濟變數的五大行庫平均房貸利率，始終不受迴歸的變數選擇青睞，可能是近年來利率一直在低檔而且沒有大幅變動，因此被模型列為不相關，但是購屋時貸款畢竟是大多數購屋者的主要資金來源，放款利率仍然是重要的考量因素，即使在本次研究中都沒派上用場，往後的研究還是建議列入為宜。

透天厝因為土地為屋主全部持有，未依建管法規申請的增建或改造等屋況，附加價值並未如實反映在目前實價登錄的欄位上，未來，實價登錄若增加未保存登記部分的欄位，有助於瞭解透天厝房屋的實際附加價格。

本研究實證資料只限於高雄市實價登錄的部分，並考慮物件與高捷系統的距離以及可能影響房價的整體經濟指標等因素，未來研究者可以加入其它地區因地制宜的特徵值，以適用在其它分區或縣市房地交易上。

未來若能發展不動產大量估價軟體系統來整合不同網站的資訊(實價登錄和不動產相關網站)，建構軟體及 APP，計算出預估數值及判斷係數，提供使用者自行判斷應用，並增設使用者回饋系統，利用所蒐集的回饋資料，提供估價軟體不斷更新學習，推估數值參考的價值自然會逐步提升。

本研究主要使用倒傳遞類神經網路計算預估房價，預期實價登錄制實施年期更長之後，也適合搭配時間序列演算法來觀察不動產交易價格的走勢，有助於進一步提升消費者購屋決策判斷，除了預估價格因素，還能選擇較佳的購買時機點。

參考文獻

- 105 年住宅需求動向調查(2016) ，中華民國住宅協會 第十四卷 第一期
- 花敬群(2010)，電腦大量估價模型於實務應用之探討，金融聯合徵信雙月刊 第十二期，p27-p36
- 賴碧瑩(2007)，應用類神經網路於電腦輔助大量估價之研究，住宅學報 第十六卷第二期，p43-65
- 蔡瑞煌、高明志、張金鶚(1997)，類神經網路應用於房地產估價之研究
- 魏如龍、陳奉瑤(2003)，類神經網路於不動產價格預估之研究，國立政治大學地政學系研究所碩士論文
- USPAP, Uniform Standards of Professional Appraisal Practice 美國估價基金會擬定的標準估價作業準則