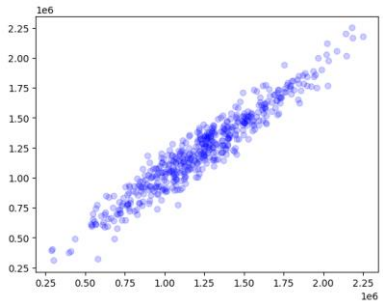
機器學習概論

<HW1> 房價預測修改

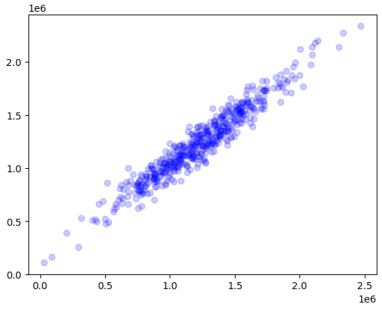
課堂中原始程式碼使用 LinearRegression 模型完成預測的值為 0.9216，值越接近 1 預測越準確，切割測試及訓練與隨機切割的參數為 test_size = 0.3，random_state = 54；欲將值提高。

(一)首先嘗試調整參數，

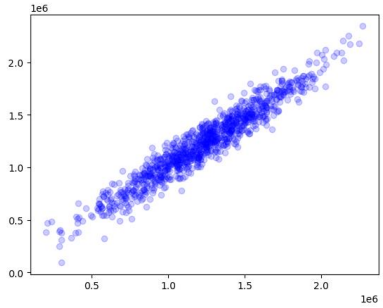
test_size = 0.1
random_state = 54
r2_score = 0.9187



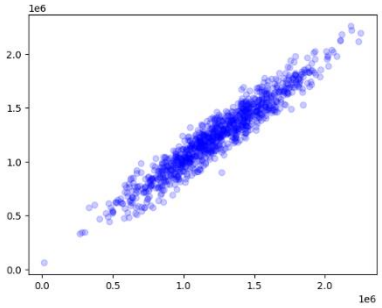
test_size = 0.1
random_state = 43
r2_score = 0.9270



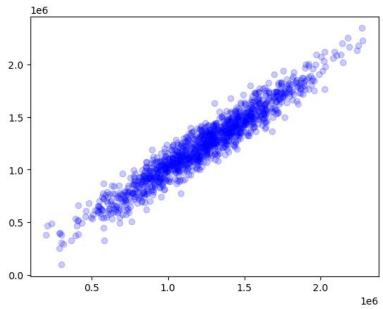
test_size = 0.2
random_state = 54
r2_score = 0.9218



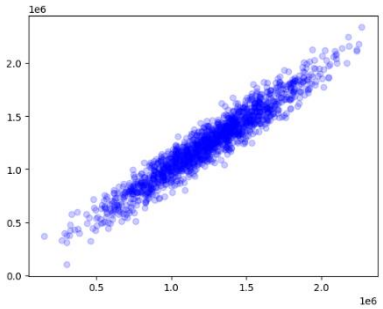
test_size = 0.2
random_state = 68
r2_score = 0.9194



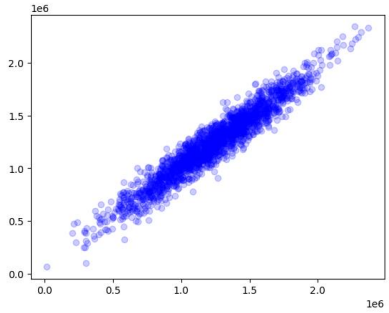
test_size = 0.25
random_state = 54
r2_score = 0.9218



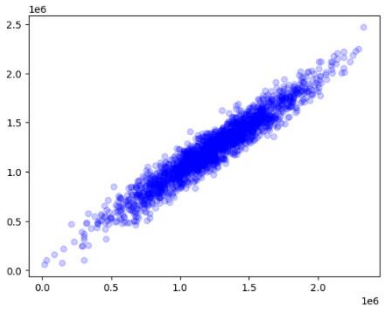
test_size = 0.3
random_state = 95
r2_score = 0.9215



test_size = 0.4
random_state = 54
r2_score = 0.9209



test_size = 0.4
random_state = 36
r2_score = 0.9195

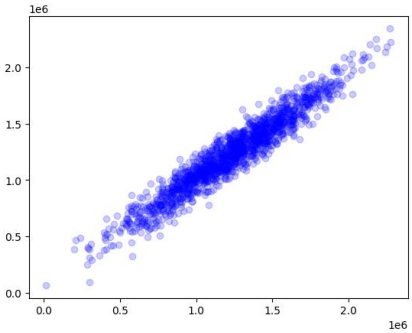


發現並無明顯改變，亦無使值增加，即無法提高預測準確度。

(二)再嘗試使用別種預測模型，同樣使用回歸預測模型，
參考在監督式學習中，sklearn 的模型。

(1) Lasso

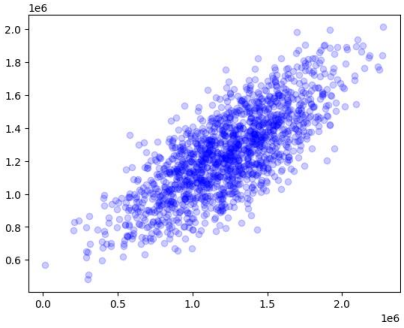
test_sizz = 0.3
random_state = 54
r2_score = 0.9216



預測的值和使用 LinearRegression 模型相同。

(2) SVR

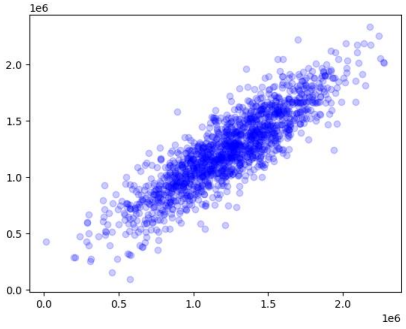
test_sizz = 0.3
random_state = 54
r2_score = 0.5746



預測結果更為發散，改變參數亦同。

(3) DecisionTreeRegressor

test_size = 0.3
random_state = 54
r2_score = 0.7523



預測的結果並無較 LinearRegression 模型佳。

無法有更高的結果，可能沒有使用到適用的模型，資料預測結果不適用嘗試使用的模型，計算方法無法達到預期效果，導致預測結果未能超過原模型的值。

<HW2> Kaggle 自選競賽

1. 比賽簡介/為什麼選擇這個比賽/資料集、目標介紹

選擇使用 spaceship titanic 是否將乘客送到另一個維度的資料預測，上課的主題為預測船上乘客存活與否，而選擇的主題為預測飛船乘客被送至另一維度與否，概念與上課所學相似，資料集內容也較完整，資料集有乘客個人紀錄，約使用三分之二的資料做機器學習訓練，有乘客 ID、姓名、年齡、原居住星球、是否在艙內處於冷凍狀態、艙室號碼、欲登陸之星球、是否使用 VIP 服務及使用太空船內設施費用等項目。目標為預測乘客是否有被異常傳送至另一個維度。

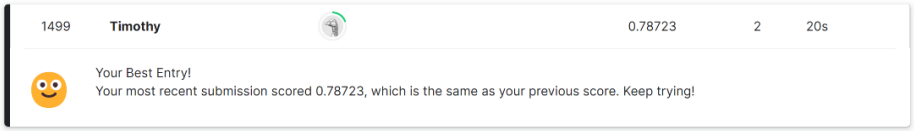
2. 如何實作(附上程式碼與結果)/跟上課內容的關聯性/延伸學習了那些

程式碼如下：
(https://github.com/lct1452/HW_ML/blob/main/HW2_competition_spaceship.ipynb)。

首先觀察資料後將無相關的姓名和艙位去掉，再觀察目標與各欄位項目的關聯性，發現在使用特殊服務中，FoodCourt 和 ShoppingMall 消費金額高的乘客皆有被傳送，而 RoomService、Spa 和 VRDeck 使用費用越高越沒有被傳送。接著處理各欄位中的空值，從觀察中看出過半乘客皆未消費，因此以最大值填補空值，較不會因極端資料產生誤差，再將 HomePlanet 及 Destination 轉為是否為該項目。處理完資料即開始機器學習，先將預測的目標丟掉放入 X，集合所有目標放入 y，最後使用 Logistic regression 邏輯回歸分類預測模型進行預測，再將結果以表格方式呈現。將完成的程式碼導入比賽格式。

3. 比賽結果說明/推測還可能從那些方式改善/不同的嘗試與結果分析

比賽結果為 0.78723，並沒有達到較高的精確度，可能可以調整參數、更換預測模型及方法，或是將不必要的項目捨棄，避免影響機器學習結果。



參考資料

1. https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
2. https://www.kaggle.com/competitions/spaceship-titanic/