

基於天氣因素之共享單車 需求分析及多方法預測研究

Predicting the Riding Volume of Shared Bikes Based on Weather Factors

組員：陳玟卉、王柏元、江詠筑
組別：第 1 組

TABLE OF CONTENTS

01

Introduction

文獻回顧 問題設定 資料分析 / 江詠筑

02

Method_prediction model

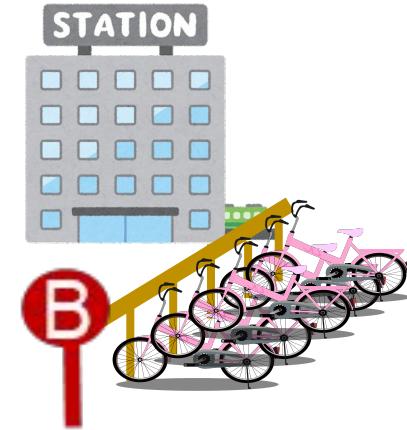
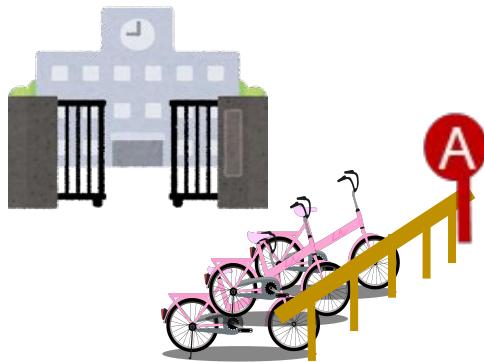
實驗結果最好的五個預測方法 / 王柏元

03

Discussion and conclusion

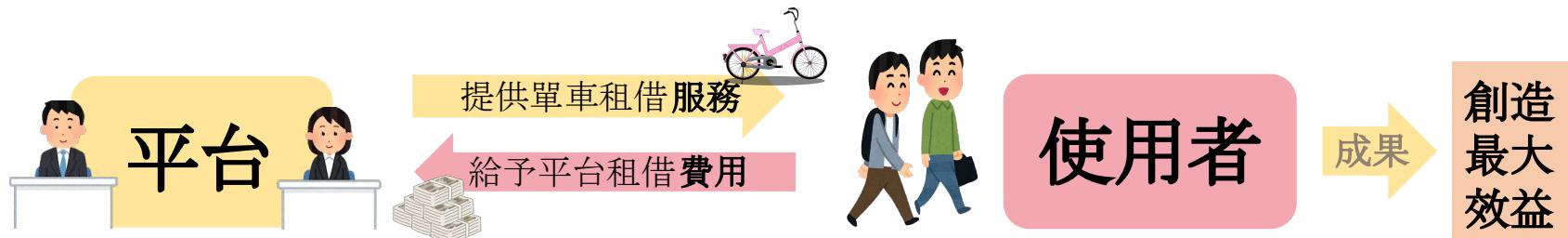
不同編碼方式的實驗結果 / 陳玟卉

共享單車系統 BBS (Bicycle Sharing System)



解決民眾「最後一公里」路程的單車租借平台

共享單車系統之優缺點



優點

1. 簡單租車手續
2. 不用受制於在店家定時定點交車、還車
3. 可利用**物聯網**確認租車相關即時資訊

缺點

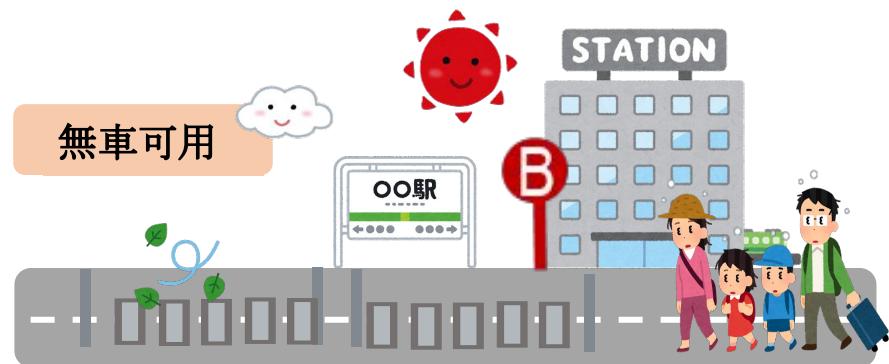
1. 大量單車的**停放**、**維護**等問題。
2. 易出現**沒車可用**或**沒樁可停**的情況。
3. 創建初期需投入**大量資金成本**



共享單車數量失衡問題之原因

地理：火車站出入口、上下班時間流量難以平衡

天氣：壞、好天氣下單車供需難以平衡



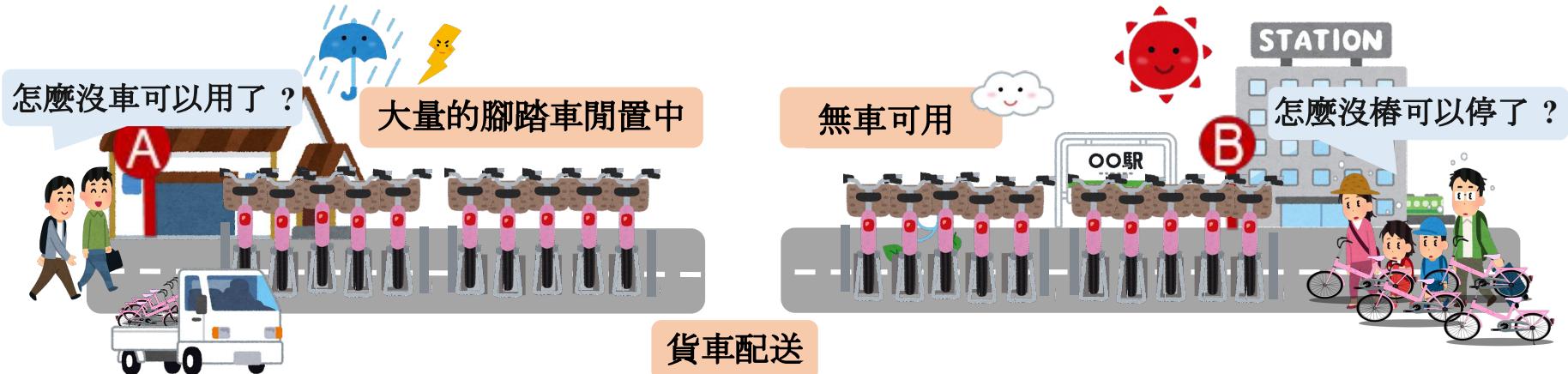
解決辦法



貨車再配送
(Vehicle redistribution)

共享單車再配送失衡問題之原因

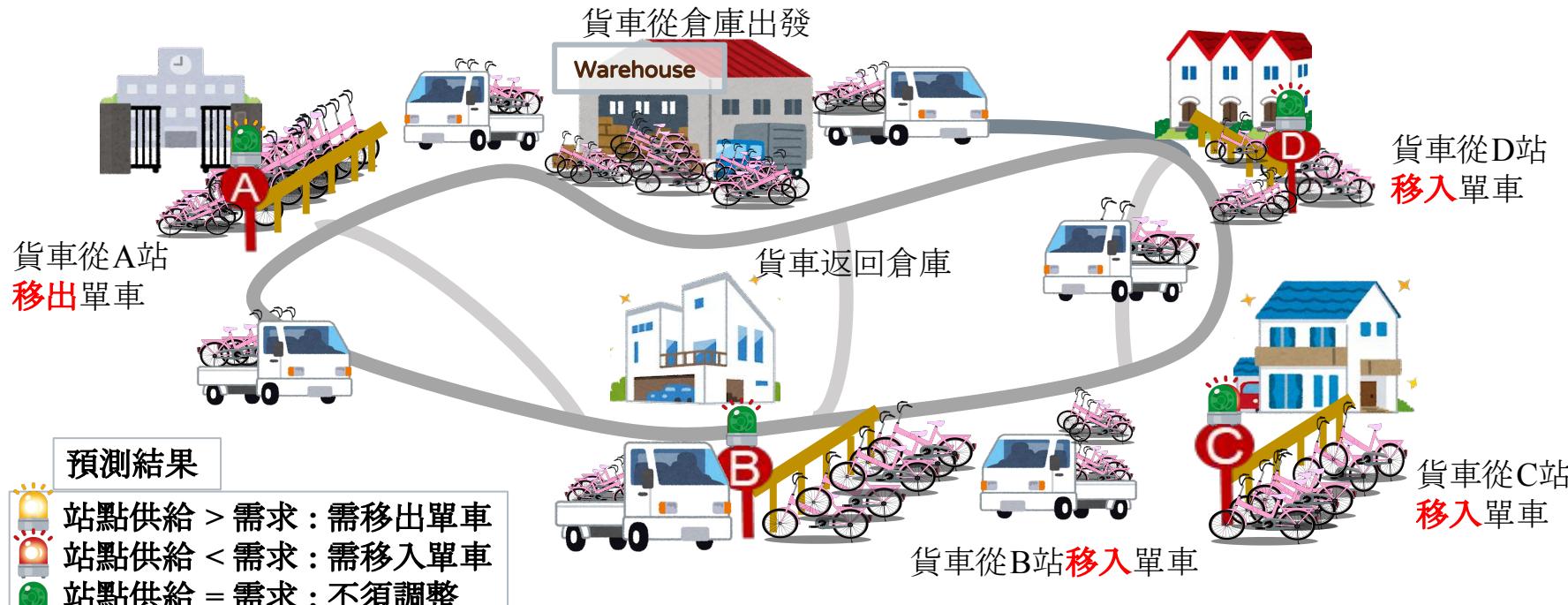
未來變動：若僅依照車站當前資料進行再配送難以符合真實狀況



解決辦法

共享單車需求預測
(Demand forecasting)

共享單車需求預測(Demand forecasting)



研究動機

過去 研究

1. 共享單車站點需求預測 **問題** → 沒有比較不同預測方法的成效
2. 共享單車配送路徑規劃 **問題** → 時空因素摻雜難以實際運用
過去環境 ≠ 未來環境 → 舊有的最佳化路徑無法套用

本篇 研究



研究目的

最佳化目標

基於天氣因素下之共享單車需求預測

有效特徵整理分析

建立**多方法**預測模型

最佳站點設置特徵條件

運用**交叉驗證**求最佳模型

找到最佳預測模型有效解決單車需求不確定問題

共享單車系統需求預測及數據分析之文獻回顧

文獻(作者及年分)	共享單車問題研究	重新定位 (靜態/動態)	考量外部因素 (天氣、地理...)	研究方法
Labadi et al. (2016)	最小化共享單車數量失衡問題的時間	靜態	✗	Branch-and-bound algorithm
Gao and Lee (2018)	單車需求數量預測問題	動態	✓	FCM-based GA, Back Propagation Network
Yao et al. (2019)	單車需求數量預測問題	動態	✓	Recurrent Neural Network, Long Short-term Memory
Xu et al. (2020)	單車需求數量預測問題	動態	✓	Self-organizing map Regression tree
Tsai (2018)	共享單車系統顧客抵達分布	動態	✓	Test of Goodness-of-Fit (follow Poisson distribution)
Liu et al. (2017)	共享單車使用特性分析	靜態	✗	Big data analysis
Yan et al. (2020)	共享單車站點騎行量分析	動態	✓	Data visualization Hypothesis testing
Benlic et al. (2020)	共享單車構建與營運策略改善	靜態	✗	few-space hot spot
本研究	共享單車最佳站點設置條件 +單車需求數量預測方法比較	動態	✓	hot spot, SVR, XGBoost, Regression tree

需求預測

數據分析

數學規劃

機器學習

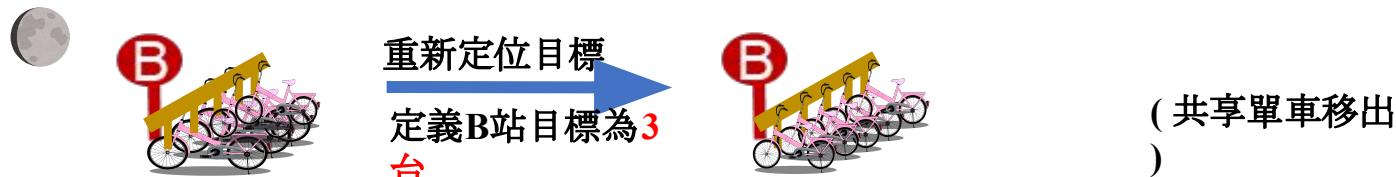
統計分析

動態、靜態重新定位

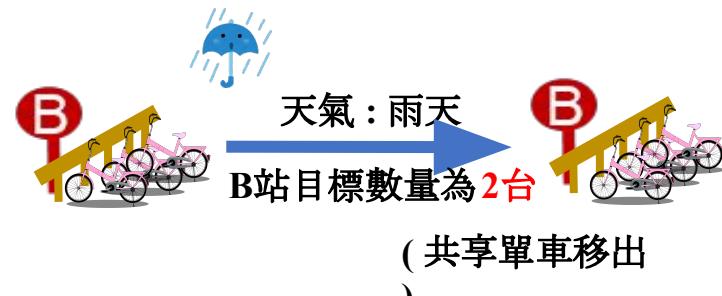
重新定位：調整各站點共享單車數量，目的是為了讓各站點共享單車數量平衡。



- 靜態重新定位：發生在夜間交通量低的時候。重新定位的目標通常已預先定義。

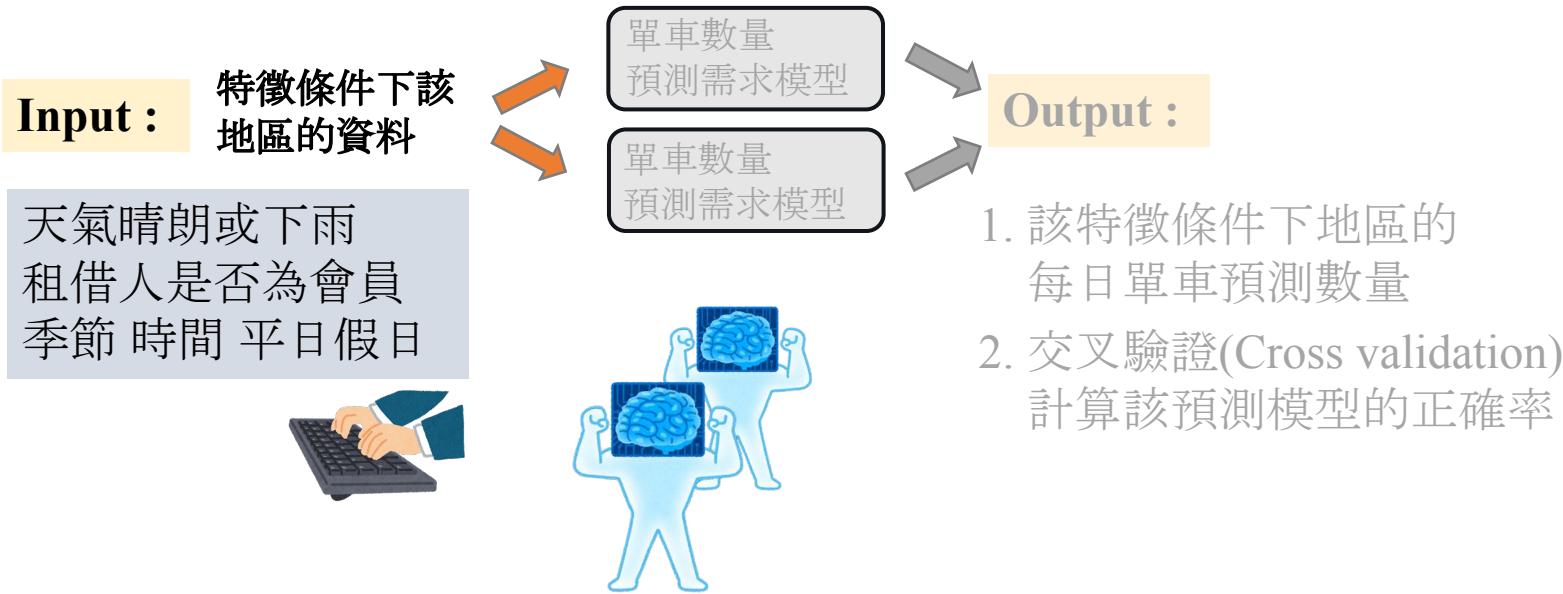


- 動態重新定位：與共享單車系統同時運行。重新定位的目標隨著實際情況(天氣、地理)不同而更動。



共享單車需求預測問題設定

Step 1 資料預處理及視覺化(數據分析)



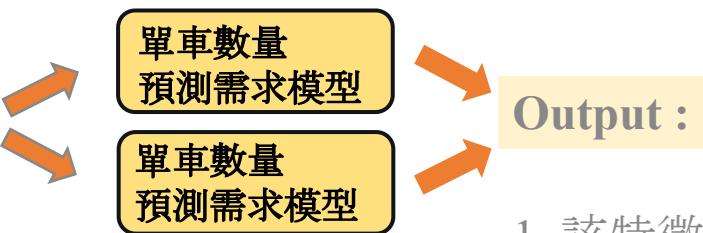
共享單車需求預測問題設定

Step 2 預測單一站點單車數量

Input :

特徵條件下該
地區的資料

天氣晴朗或下雨
租借人是否為會員
季節 時間 平日假日



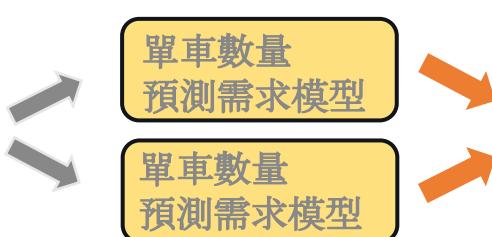
- Output :
1. 該特徵條件下地區的
每日單車預測數量
 2. 交叉驗證(Cross validation)
計算該預測模型的正確率

共享單車需求預測問題設定

Step 3 比較各預測方法的正確率

Input : 特徵條件下該地區的資料

天氣晴朗或下雨
租借人是否為會員
季節 時間 平日假日



Output :

1. 該特徵條件下地區的
每日單車預測數量
2. 交叉驗證(Cross validation)
計算該預測模型的正確率

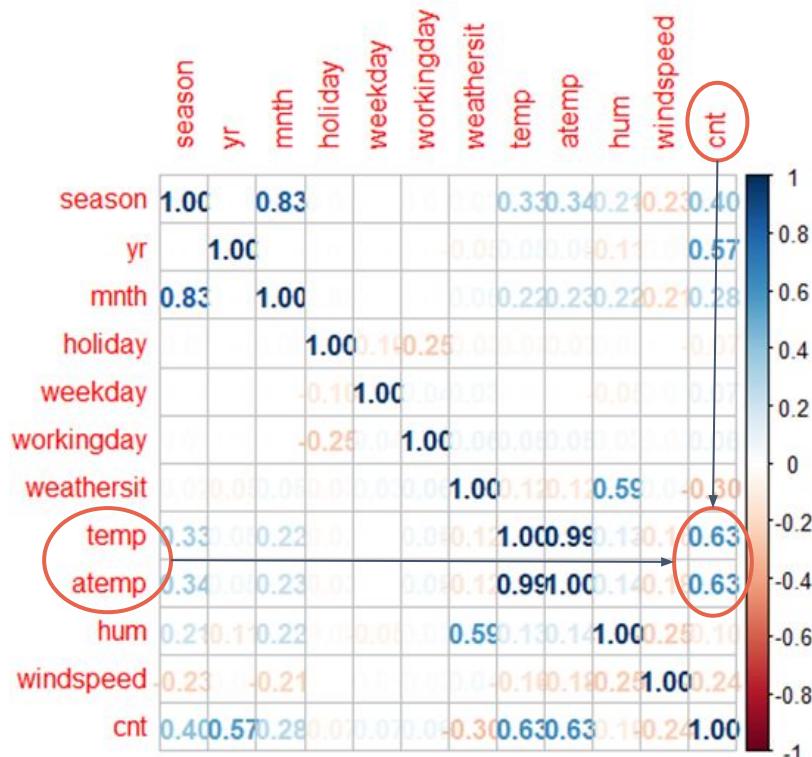


多變量資料的各特徵相關性熱圖(heat map)

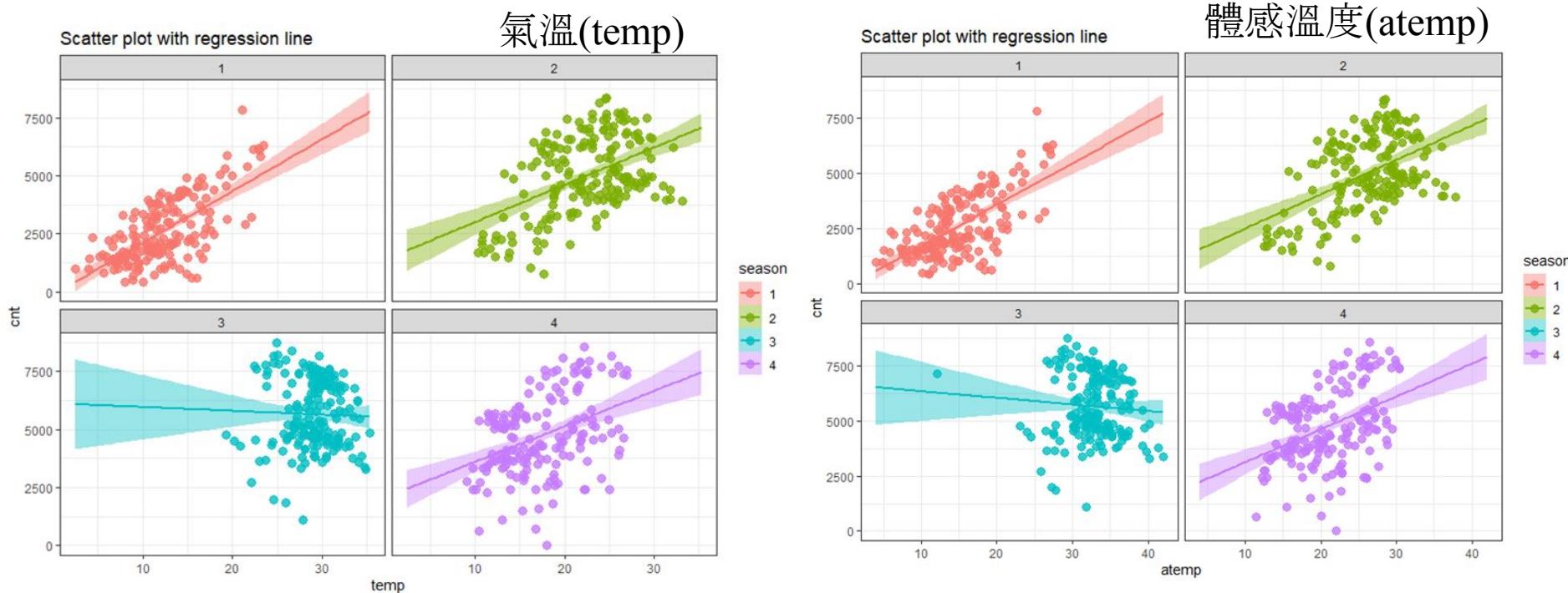
- 由Correlation plot知：
 - 影響總租借人數的前三大因子：
 - 溫度、體感溫度(temp、atemp)相關性最高
→ 相關性為 0.63
 - 月份(yr), 相關性為 0.57
 - 季節(season), 相關性為 0.4

結論

租借人數受溫度影響最大

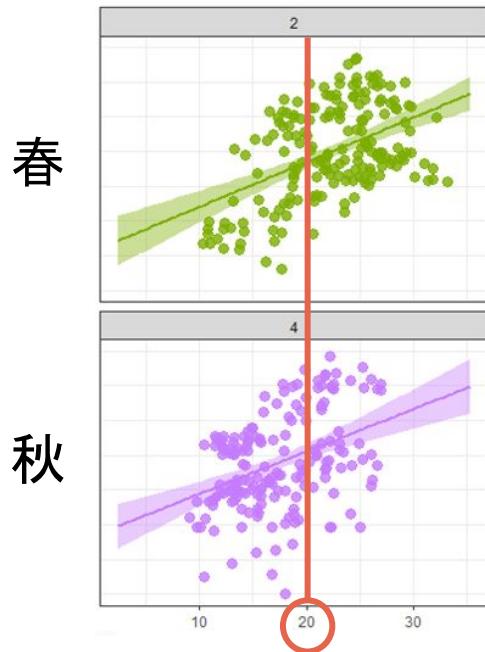


溫度、季節和總租借人數的統計分析

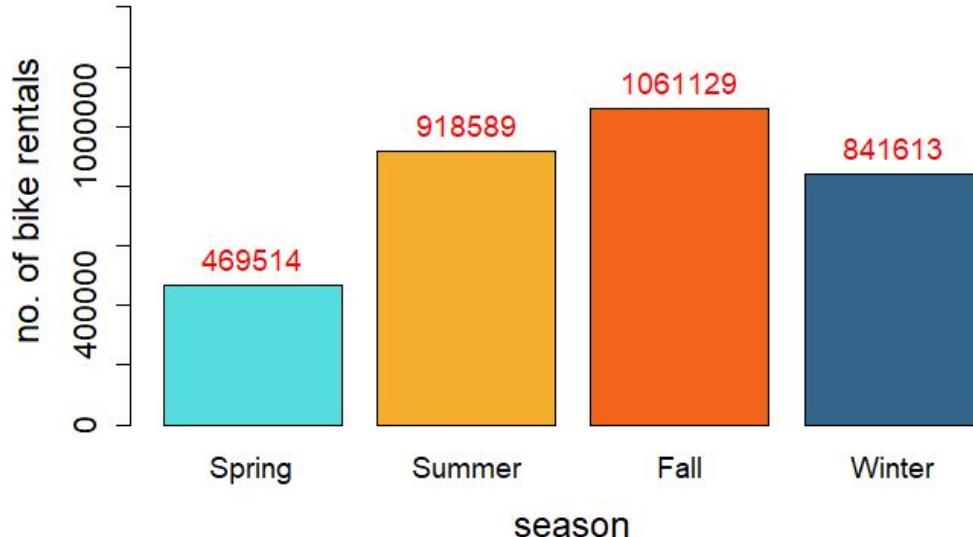


因為由Correlation plot得知，總租借人數和氣溫(temp)和體感溫度(atemp)和季節相關性較大，因此我們scatter plot+regression line加以分析。

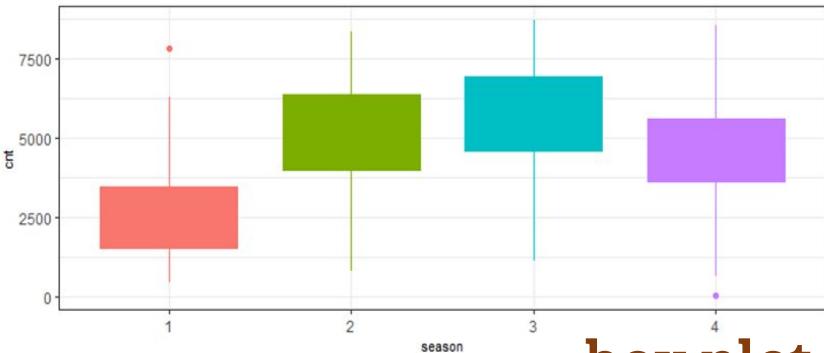
溫度、季節和總租借人數的統計分析



Bike rentals vs. season



季節和總租借人數的統計分析

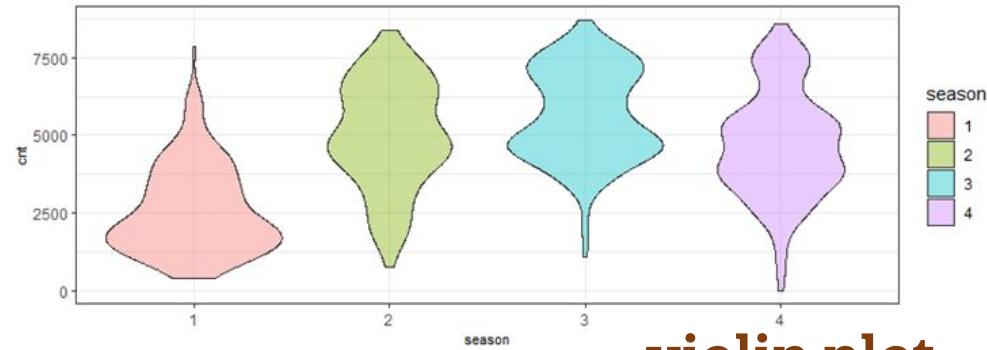


season	Mean of cnt	Median of cnt
1	2608.411	2222.5
2	4992.332	4941.5
3	5644.303	5353.5
4	4728.163	4634.5

1 : 春
2 : 夏
3 : 秋
4 : 冬

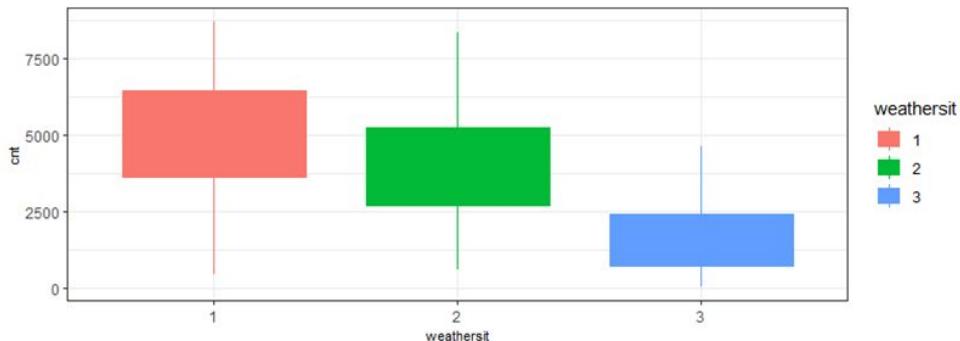
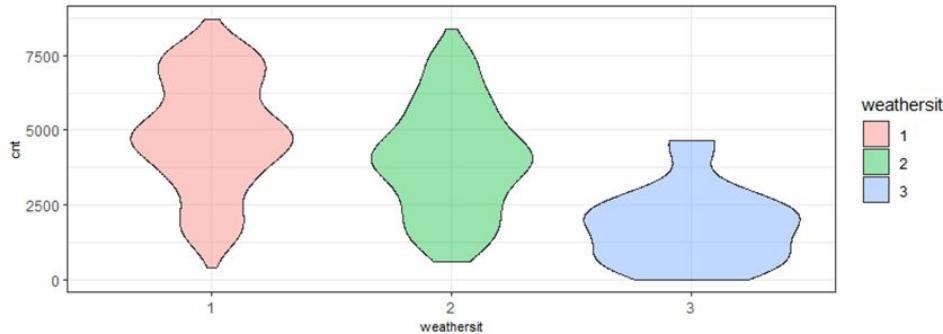
夏季和秋季的中位數和平均數都顯著高於其他季節

- Highest demand in fall (season 3)
- 秋 > 夏 > 冬 > 春 in total demand count
(春季下雨 冬季下學)
- 使用者傾向在秋季和夏季租借並騎乘共享單車。



天氣和總租借人數的統計分析

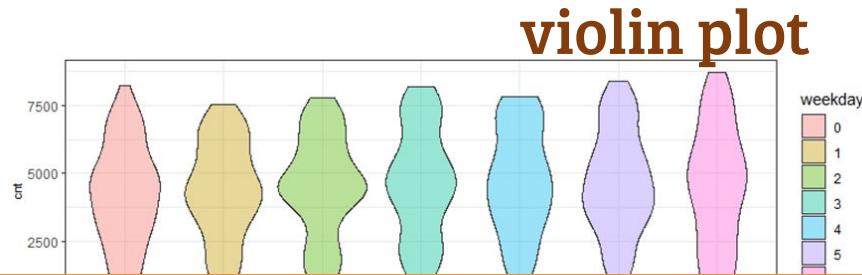
由圖可明顯得知，在weathersit第三種（下小雨或下小雪或陰天）的租借量明顯少於其他兩種天氣狀態。



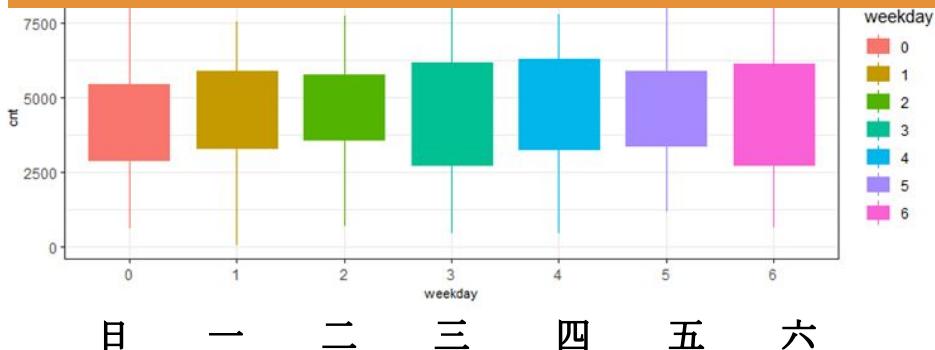
而在weathersit第一種（晴朗無雲）的資料相較於其他兩種分散，需求變異大，因此對於預測的需求較高，需要為站點的需求多加監控。

其他因素和總租借人數的統計分析

- 各星期的總租借量大致相同
- 星期二資料**集中**且整體而言租借量較**固定**, 所以對於預測的需求可能會較小



結論：其他因素對總租借人數影響很小



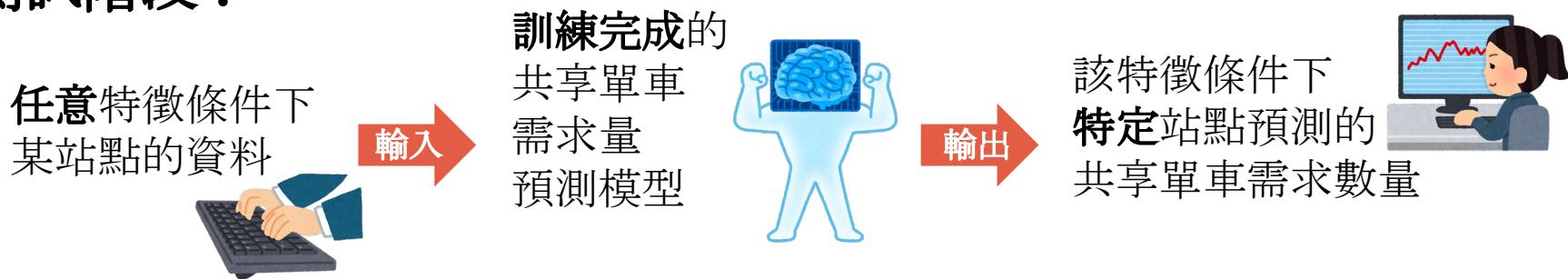
- 周一、二、三、四、五、六的分布較寬，也就是在這兩星期的租借量較不固定，需求變異大。
- 需求變異對於站預測特別重要，避免腳踏車供應有時過多，有時過少的失衡問題。

研究方法 1/2：交叉驗證共享單車需求預測模型

訓練階段：



測試階段：



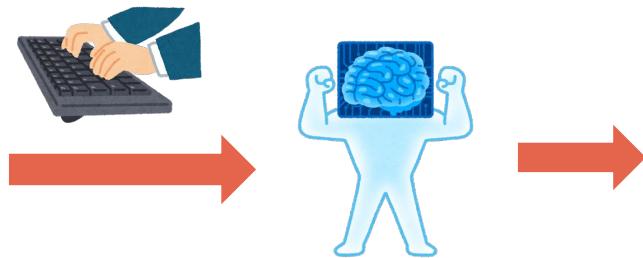
研究方法 2/ 2：建構多個共享單車需求預測模型

預測方式 1/5 :

Multiple linear regression

Input : 特徵資料

- 溫度
- 天氣狀況
- 相對溼度
- 風速大小
- 季節 (春/夏/秋/冬)
- 特殊節日 (假日 / 平日)
- 月份



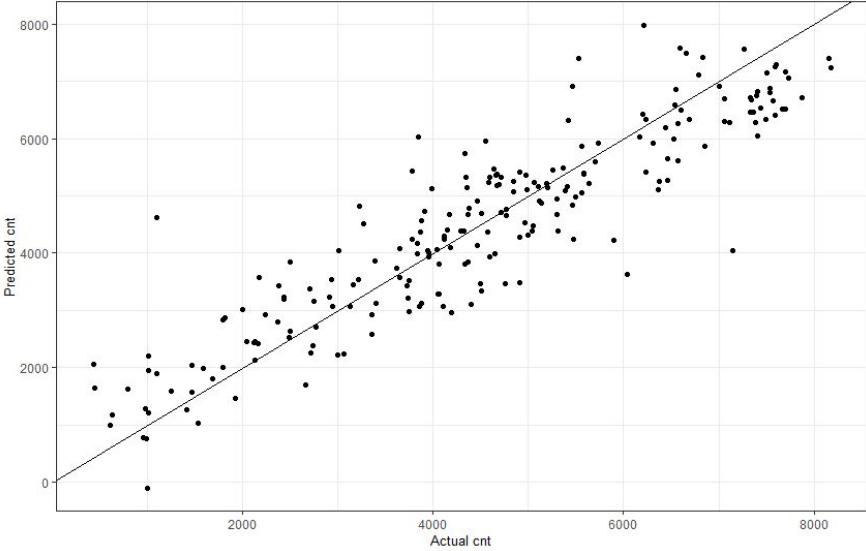
Model 1
Use R language to
implement

Output : 數值
(單車需求預測數量)

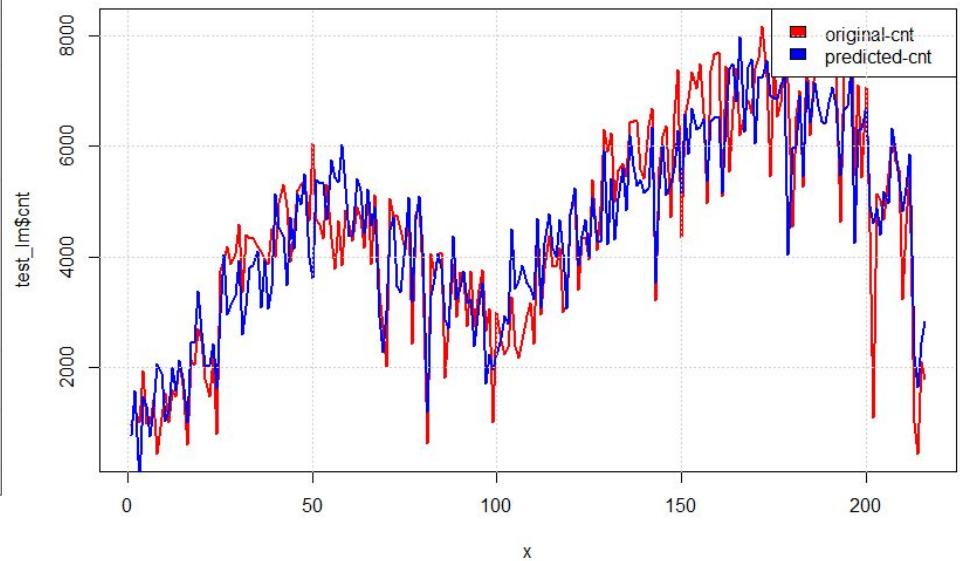
MSE: 634975.9
MAE: 587.6497
RMSE: 796.8537
accuracy: 0.8196657

預測結果 (Multiple linear regression)

Multiple Linear Regression



bike sharing test data prediction(ml)



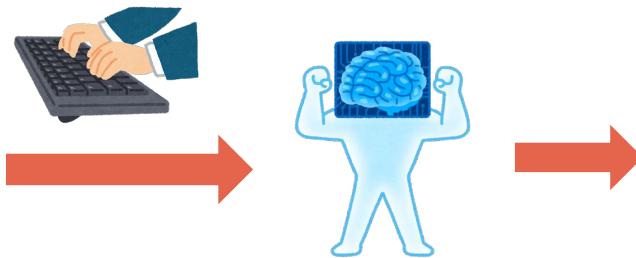
研究方法 2/2：建構多個共享單車需求預測模型

預測方式 2/5：
Support Vector Regression

建立在機器學習上的
支持向量迴歸模型
時間序列分析 (time series analysis)

Input : 特徵資料

- 溫度
- 天氣狀況
- 相對溼度
- 風速大小
- 季節 (春/夏/秋/冬)
- 特殊節日 (假日 / 平日)
- 月份



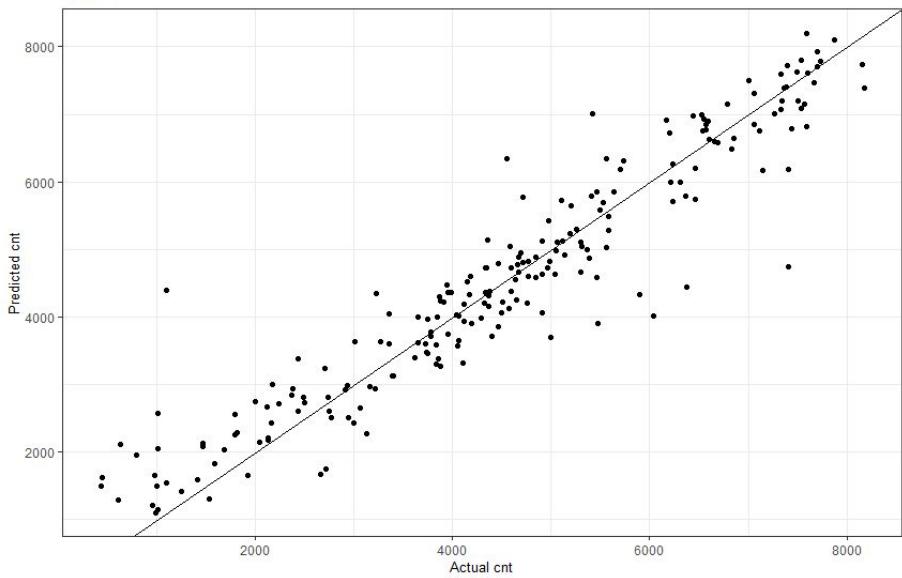
Model 2
Use R language to
implement

Output : 數值
(單車需求預測數量)

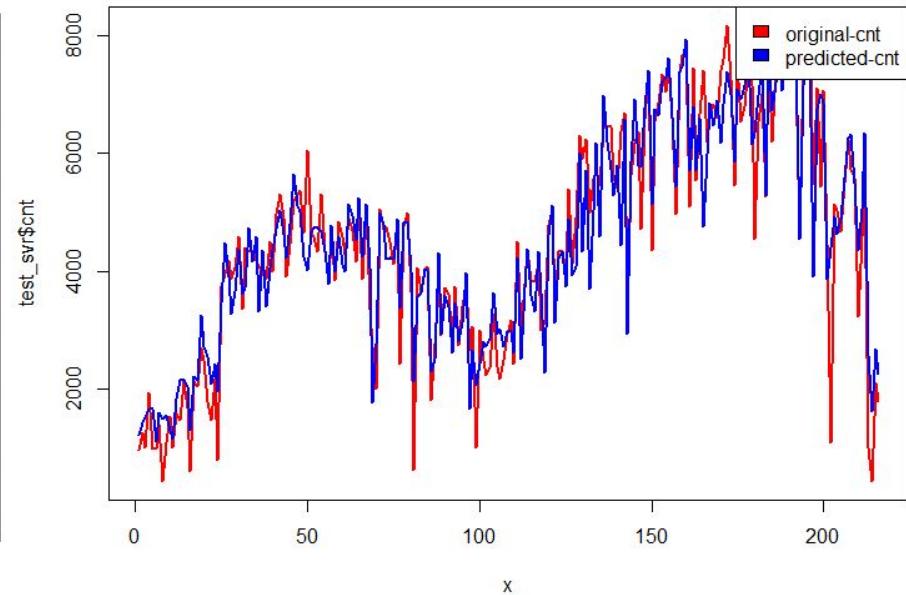
MSE: 392856.1
MAE: 442.0154
RMSE: 626.7823
accuracy: 0.8242527

預測結果 (Support Vector Regression)

SVR



bike sharing test data prediction(knn)

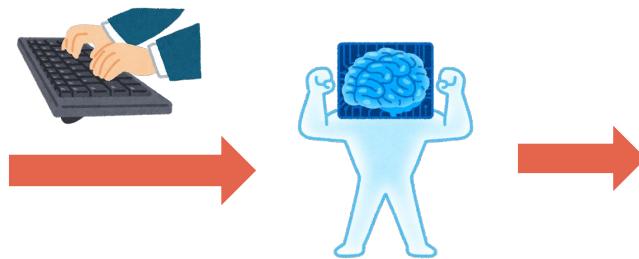


研究方法 2/ 2：建構多個共享單車需求預測模型

預測方式 3/5：
neural networks

Input : 特徵資料

- 溫度
- 天氣狀況
- 相對溼度
- 風速大小
- 季節 (春/夏/秋/冬)
- 特殊節日 (假日 / 平日)
- 月份

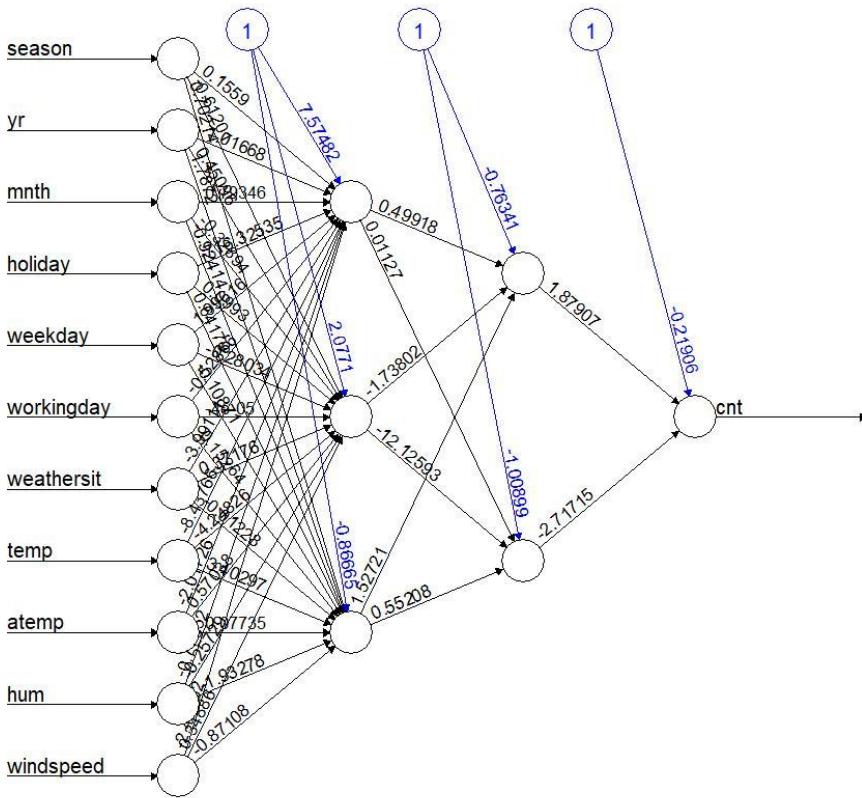


Model 3
Use R language to
implement

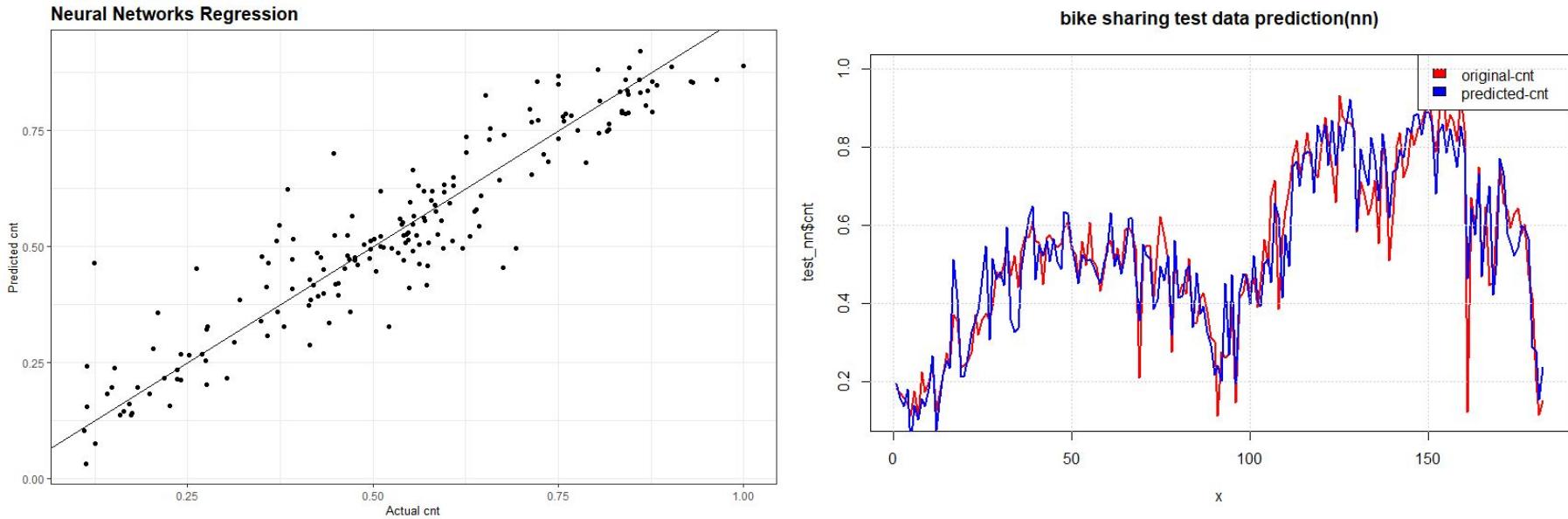
預測 (prediction)
類神經網路

Output : 數值
(單車需求預測數量)
accuracy: 0.8536584

input-hidden layer-output



預測結果 (neural network)



研究方法 2/ 2：建構多個共享單車需求預測模型

預測方式 4/5：

Random Forest Regression

Input : 特徵資料

- 溫度
- 天氣狀況
- 相對溼度
- 風速大小
- 季節 (春/夏/秋/冬)
- 特殊節日(假日 / 平日)
- 月份



Model 4

Use R language to implement

Bagging (Bootstrap Aggregation)

多顆CART決策樹(decision tree)
(CART樹為使用GINI算法的決策樹)

Output : 數值

(單車需求預測數量)

MSE: 78126.28

MAE: 175.6793

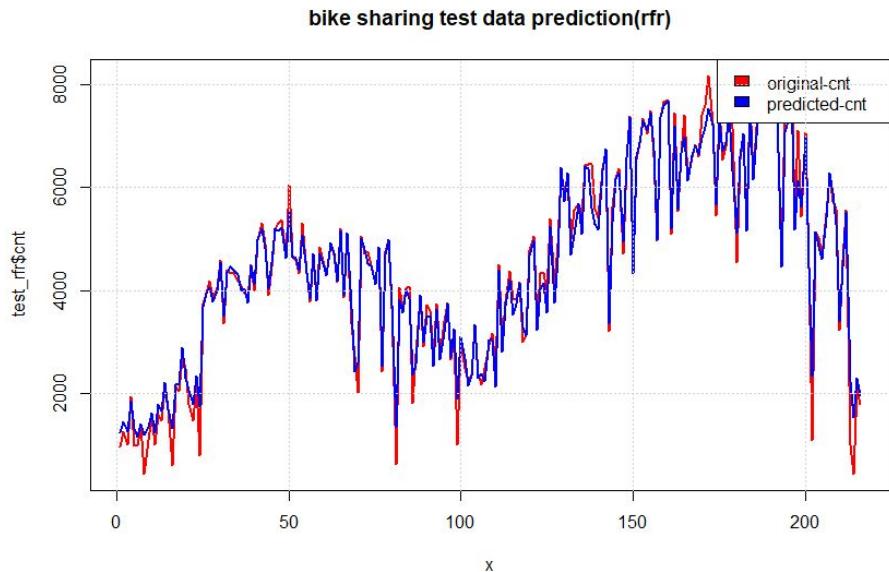
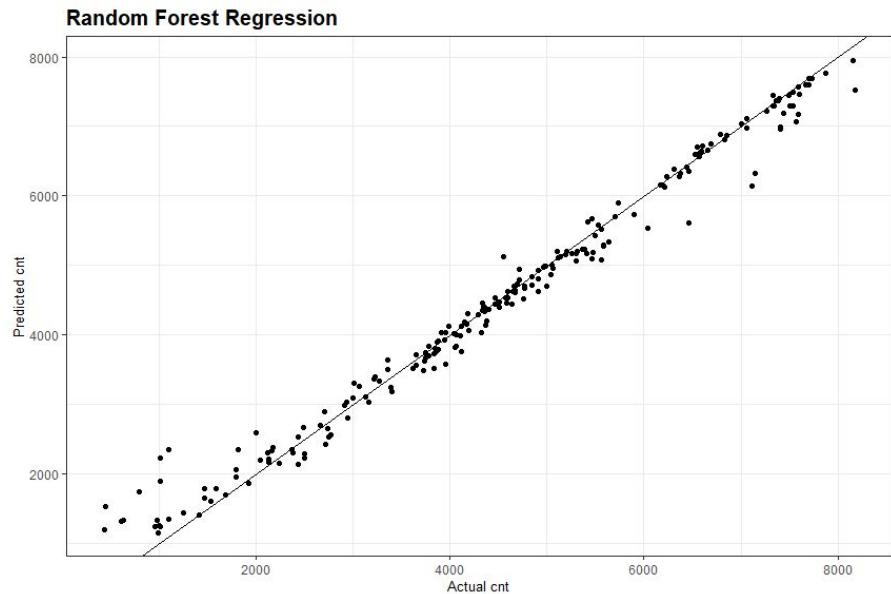
RMSE: 279.5108

accuracy: 0.907943

以多數決的方式
來決定分類結果



預測結果 (Random Forest Regression)



研究方法 2/2：建構多個共享單車需求預測模型

預測方式 5/5：

eXtreme Gradient Boosting

以 Gradient Boosting 為基礎

機器學習競賽神器 高效、靈活且可移植

Input : 特徵資料

- 溫度
- 天氣狀況
- 相對溼度
- 租借人是否為會員
- 季節 (春/夏/秋/冬)
- 特殊節日(假日 / 平日)



Model 5

Use R language to implement

Output : 數值

(單車需求預測數量)

MSE: 26660.58

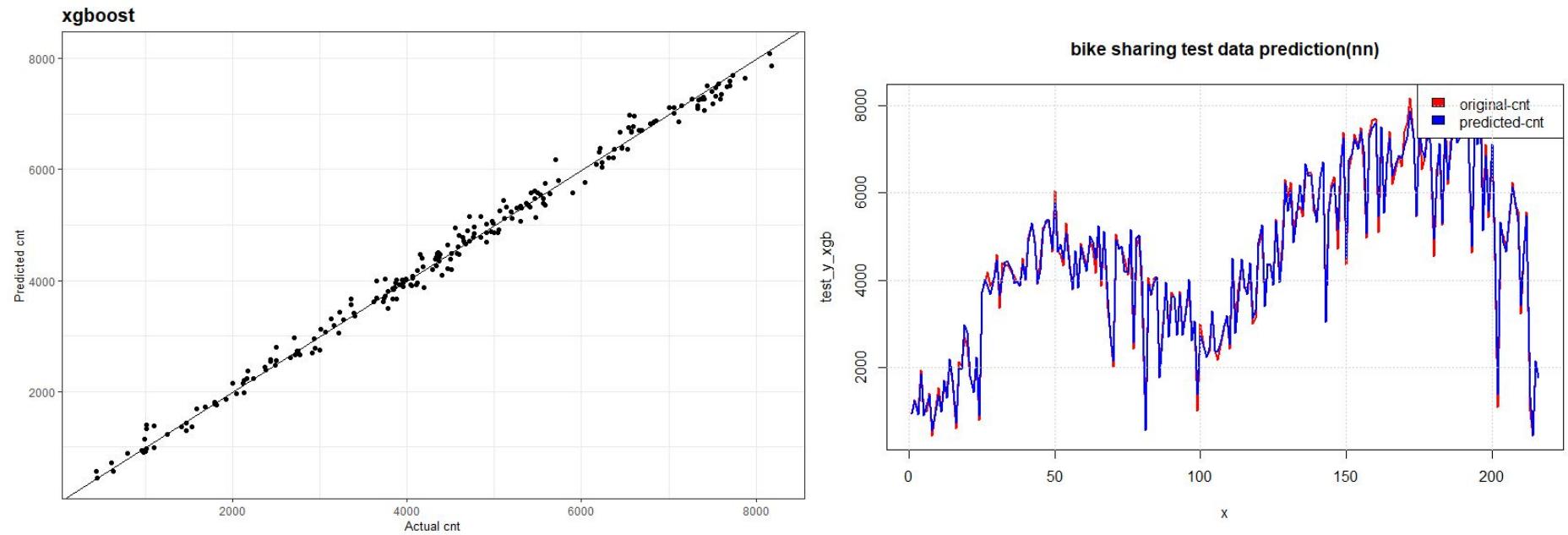
MAE: 124.7811

RMSE: 163.2807

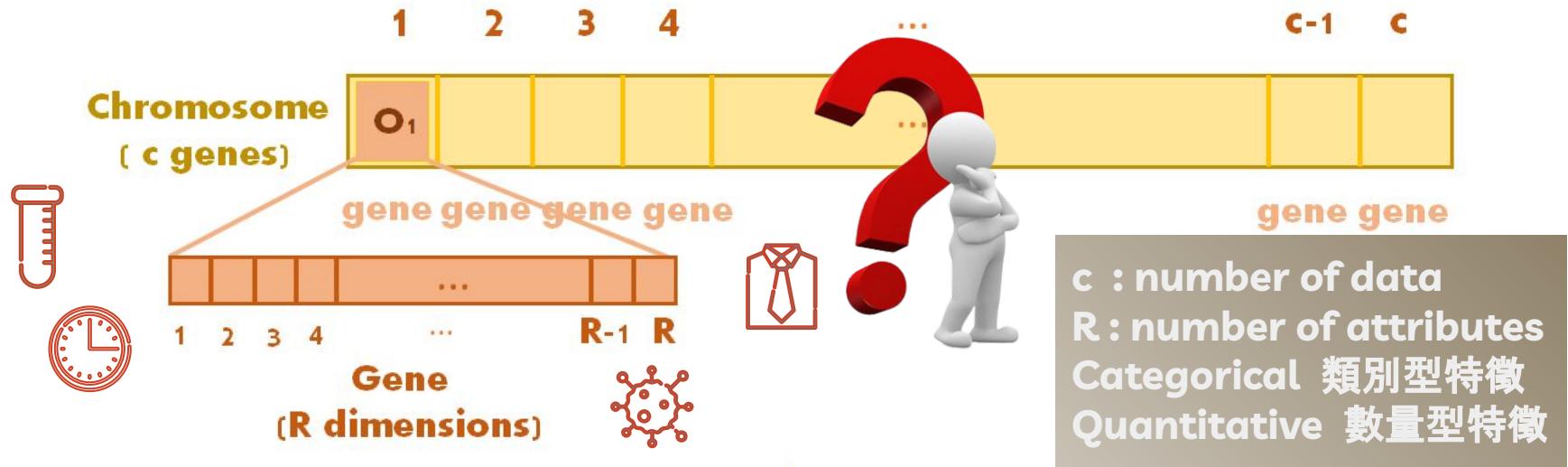
accuracy: 0.9632193



預測結果 (eXtreme Gradient Boosting)



編碼討論 Encoding discussion



Trade off between Label Encoding ,
One hot encoding & frequency encoding

參考資料 Reference

- Moment-based rental prediction for bicycle-sharing transportation systems
<https://www.sciencedirect.com/science/article/pii/S0360835218306260>
- Datasets
<https://www.kaggle.com/gauravduttakiit/bike-sharing>
- 共享單車系統顧客抵達分布分析
<https://www.airitilibrary.com/Publication/alDetailedMesh1?DocID=U0001-2202201814065200>
- 共享單車騎行量預測
[http://www.social.uestc.edu.cn/article/doi/10.14071/j.1008-8105\(2021\)-3014](http://www.social.uestc.edu.cn/article/doi/10.14071/j.1008-8105(2021)-3014)
- 大數據分析 臺北市公共自行車使用特性
<https://www-ws.gov.taipei/Download.ashx?u=LzAwMS9VcGxvYWQvMzY3L3JlbGZpbGUvNDUwMDAvNzY2MzlzOS85OTM3NTUzM0zYzIxLTRhZmQtOWY5Ny1mNjNkMzM0ZmEzMzcucGRm&n=MTA2MTjljkfmlbjmk5rlilbmnpDoh7rljflulIhazlhbHoh6rooYzou4rkvb%2FnIKnjibnmgKcucGRm&icon=.pdf>
- 公共自行車時空分析法之構建與營運策略改善 -以台北微笑自行車為例
<https://core.ac.uk/download/pdf/225228704.pdf>
- Short-term Demand Forecasting of Shared Bicycles Based on Seasonal Grey Markov Model
<https://iopscience.iop.org/article/10.1088/1742-6596/1903/1/012059>
- An efficient optimization approach
<https://link.springer.com/article/10.1007/s00521-020-05035-x>

參考資料 Reference

- 支持向量迴歸(Support Vector Regression)

<https://medium.com/r-%E8%AA%9E%E8%A8%80%E8%87%AA%E5%AD%B8%E7%B3%BB%E5%88%97/r%E8%AA%9E%E8%A8%80%E8%87%AA%E5%AD%B8%E6%97%A5%E8%A8%98-20-%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E4%B8%80-%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E8%BF%B4%E6%AD%B8-support-vector-regression-30cd834a918>

- 隨機森林(Random Forest)

<https://medium.com/chung-yi/ml%E5%85%A5%E9%96%80-%E5%8D%81%E4%B8%83-%E9%9A%A8%E6%A9%9F%E6%A3%AE%E6%9E%97-random-forest-6afc24871857>

- 神經網路

<https://www.geeksforgeeks.org/how-neural-networks-are-used-for-regression-in-r-programming/>

- 機器學習常勝軍 - XGBoost

<https://ithelp.ithome.com.tw/articles/10273094>

<https://www.datatechnotes.com/2020/08/regression-example-with-xgboost-in-r.html>

- 資料前處理(Label encoding、One hot encoding and frequency encoding)

<https://medium.com/@PatHuang/%E5%88%9D%E5%AD%B8python%E6%89%8B%E8%A8%98-3-%E8%B3%87%E6%96%99%E5%89%8D%E8%99%95%E7%90%86-label-encoding-one-hot-encoding-85c983d63f87>

THANKS!

