

健康資料加值暨統計中心



[首頁](#) > [二級單位](#) > [健康資料加值暨統計中心](#) > [eNews](#)

類別

單位簡介

業務職掌

服務內容

[健保資料庫專區](#) >

學習資源-軟體操作>

eNews

平台系統

檔案下載 >

2020.06.29

eNews第36期



脊迴歸模型

(Ridge Regression Model)

在大數據分析應用時，當反應變數為連續性變數時，常用線性迴歸模型進行解釋變項及反應變項之相關性討論，並利用最小平方方法(OLS)進行參數之推論。在以下狀況時，在參數估計值將產生偏誤(bias)，其參數估計的標準差也會高估的情形：(1)多元共線性(multicollinearity)：當解釋變數間具高度相關性時，容易使模型之參數估計的正負號和實際相反，其標準差估計也會異常大，可利用VIF檢查其共線性的狀況($VIF_j > 5$ or $VIF_j > 10$ ，表示變數間具共線性， $j=1,2,\dots,p$ ， p 為解釋變數個數)，剔除共線性之變數，以避免模型的參數錯誤的被估計及檢定，從重配適模型，以納入合適之變數。此時，也可以利用正規化迴歸線，脊迴歸或Lasso迴歸，克服上述的缺點。(2)模型預測上：若樣本數 $n > p$ 時，利用一般之迴歸線並無問題，但若遇到 $n < p$ 時，則估計值的變異數偏高(趨近無窮大)且預測能力會變差，而且一般的OLS的估計結果非唯一。(3)模型解釋：當模型同時納入大量解釋變數時，則無法真實呈現對反應變數的影響，而進行模型篩選及變數取舍，以正確解釋模型之相關性；而脊迴歸或Lasso迴歸也可以找出重要變數的模型，使得部份係數估計趨近於0，也是機器學習常討論的方法之一。

[read more...](#)

臺北醫學大學 大數據科技及管理研究所 陳錦華 副教授

R 資料分析應用：無母數方法

我們先前介紹過的方法大多皆是建立在母體分配為常態分配或某一特定分配的假設下。若是當樣本資料太少、母體分配不明或是違反分配假設時，我們即必須改考慮採用「無母數(Non-Parametric)方法」進行分析。無母數分析方法最大的特點顧名思義即為不需假設母體的分配，不必推論其中位數、適合度、獨立性、隨機性，因為缺乏分配機率的訊息及檢定以等級(rank)為主要統計量，導致推論較不精確，但對於資料型態適用性的限制也能保有相對較大的彈性。在eNews第12期已經介紹利用「雲端資料分析暨導引系統」(R-web)作為分析工具，本期將跟大家介紹使用R來進行無母數分析。

[read more...](#)

邱義翔 副統計分析師

統計中心諮詢服務

1. 個人預約諮詢

諮詢師	統計中心老師及研究員
地點	大安校區11樓會議室
預約電話	(02)6638-2736 #1106
預約平台	http://biostat.tmu.edu.tw/consult/Reserve/



2. 三院駐點諮詢

因應疫情因素，三院駐點服務暫停舉行，鼓勵同仁多利用統計諮詢平台預約線上諮詢。

醫院駐點服務	時間	聯絡窗口
北醫附設醫院	請洽附設醫院聯絡窗口	Tel: 2737-2181 #3748
萬芳醫院	請洽醫院之研究部	Tel: 2930-7930 #7206
雙和醫院	每週二 08:30-11:30 每週四 14:00-17:30 (需先預約)	Tel: 2249-0088 #8861



[上一則](#) | [回上頁](#) | [下一則](#)

📍 235 新北市中和區圓通路301號教學研究大樓4樓（雙和校區）

☎ (02)6620-2589

Google Map →



退出(Opt out)TMUCRD利用

最後更新時間：2024.05.03

© 2020 Office Of Data Science, Taipei Medical University.

All Rights Reserved, Please see [Terms of use](#)