

# 機器學習HW1 補充資料

## Bayesian Linear Regression

1. 定義與勘誤
2. 題目的積分式的推導
3. 題目的機率分佈的選取與計算

Ans:

以上問題主要呈現在討論區的內容裡，因此這邊附上對應的網址讓大家觀看整個脈絡

勘誤:

<https://e3new.nctu.edu.tw/mod/dccforum/discuss.php?d=72572>

定義與各式推導和計算:

<https://e3new.nctu.edu.tw/mod/dccforum/discuss.php?d=72200>

# Linear Regression

## 1. Validation 比例相關：

Ans:

如何調整比例，其實沒有固定的標準，但有一個原則是，你的 validation 建構出的分佈必須夠有代表性，能夠用來驗證你的模型，而只要夠有代表性，其他都可以用來做訓練來更增進你模型的效能。(一般是選擇 8 成資料當訓練，2 成資料當驗證)。

而本題沒有限制你的比例，所以請你依照你的喜好進行設定。(但要能合理解釋得到的結果)

## 2. 如何設計 basis 函數

Ans:

- ① Basis 函數的設計是不限定的，我們只是需要有一個函數能夠將  $X$  映射到新的基底上，因此我們可以對不同的 feature 使用不同的基底函數，或者是使用兩個以上的函數如:  $f(g(x))$  這樣的形式做疊加。
- ② 本題的重點在於讓同學知道 basis function 的用途為何，以及該如何使用 basis function 來求出解，故哪一種 basis function 不會是本題的重點，您可以只選定一種。(basis function 的參數也是自由設定)

## 3. 跑回歸的時候，要把它轉換成類別型變數還是直接當成連續變數？

Ans:

助教實作的時候是將其當成連續型變數去做回歸的，換言之，即為假設那些類別變數它的 scale 是等距來將它視為是連續型。

但處理方式是無限的，同學可以將其視為類別變數然後轉為 dummy variable 再進行處理。

我們這題主要確認的點會是 polynomial 部份的程式碼是否實作正確，以及  $m_2$  error 相較於  $m_1$  error 發生什麼現象，不會單純根據 error 的高低來打分數。

#### **4. Pandas 可不可以用？**

Ans:

可以，若尚有不確定的套件可以來信詢問助教。