



1.4

# Lasso & Ridge Regression

# 概念

- Linear Regression、Lasso Regression、Ridge Regression，同樣都是為了求出一組最佳 $X$ 來預測 $Y$ 。
- Lasso & Ridge Regression較Linear Regression而言，增加了對迴歸權重大小的懲罰值，進而降低模型過擬合的風險。

Linear Regression:

$$\min_w \frac{1}{m} \sum_i (y_i - w^T x_i)^2$$

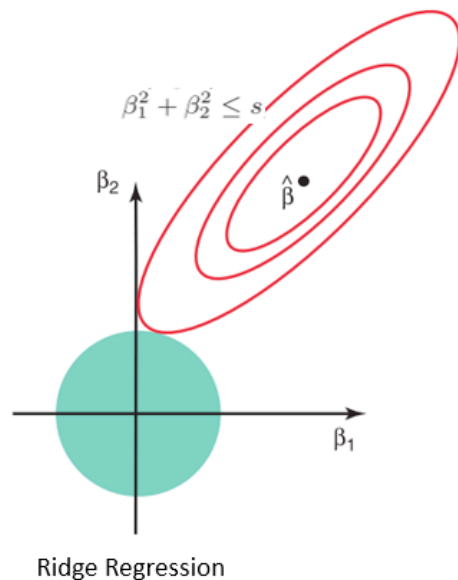
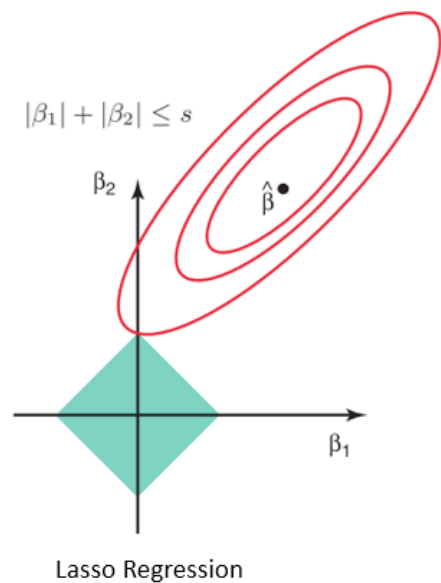
Lasso:

$$\min_w \frac{1}{m} \left( \sum_i (y_i - w^T x_i)^2 + \lambda \sum_j^n |w_j| \right)$$

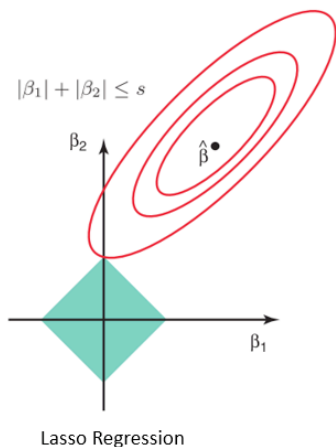
Ridge:

$$\min_w \frac{1}{m} \left( \sum_i (y_i - w^T x_i)^2 + \lambda \sum_j^n w_j^2 \right)$$

$\lambda$  is regularization penalty



Lasso (左) 和Ridge (右) 是綠色的幾何形狀，橢圓形 (紅色圓圈) 是每個模型的成本函數。放寬penalty factor的約束會導致受約束區域 (菱形、圓形) 的增加。不斷增大penalty將hit橢圓的中心。這兩種方法都是通過找到橢圓輪廓到達elliptical contours的第一個點來確定係數。



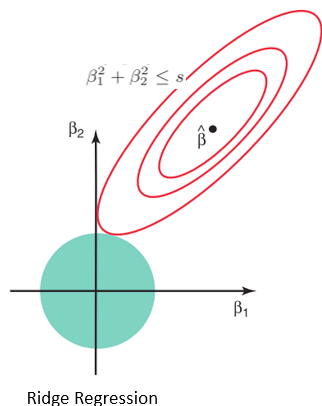
### 綠色區域是什麼？

綠色的幾何形狀（菱形或圓形）代表「模型可以接受的範圍」，也就是「對參數的限制」——我們希望模型的參數不要太大，這樣可以避免過度擬合。

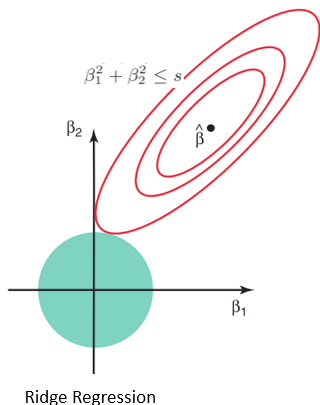
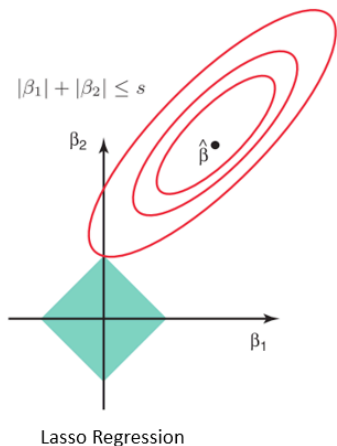
- **Lasso 的綠色區域是菱形**：表示對參數的限制是讓  $|\beta_1| + |\beta_2| \leq s$

假設  $s = 1$ ：

$|\beta_1| + |\beta_2| \leq 1$  這表示我們只能選擇那些  $\beta_1$ 、 $\beta_2$  的組合，它們的「絕對值加起來  $\leq 1$ 」。



- **Ridge 的綠色區域是圓形**：表示對參數的限制是讓  $\beta_1^2 + \beta_2^2 \leq s$   
所有在平面上，距離原點的距離不超過  $\sqrt{s}$  的點  
換句話說，這是在  $\beta_1$  與  $\beta_2$  的空間中，畫出一個「以原點為中心、半徑為  $\sqrt{s}$  的圓」，這個圓的內部就是 Ridge 所限制的「可接受的參數範圍」



### 紅色同心圓是什麼？

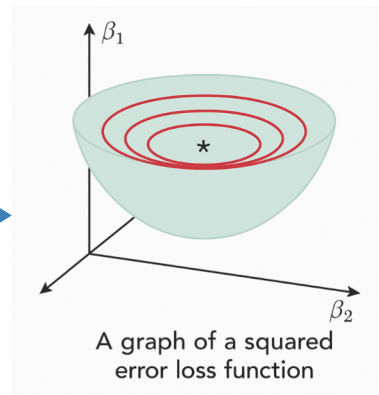
紅色是橢圓，用的是「平方誤差」來衡量模型準不準，我們把所有可能的  $\beta_1$ 、 $\beta_2$  組合一一試一遍，然後畫出「誤差值一樣大」的那些點，會發現它們連成的形狀就是一個**橢圓圈**，像地圖上的「等高線」一樣。

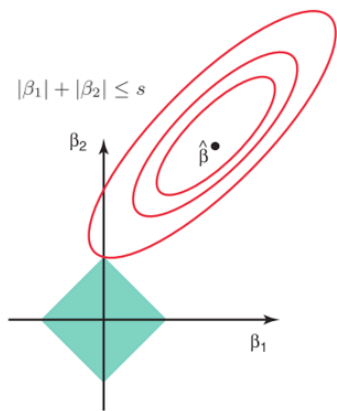
$$\text{Loss} = (\text{模型預測} - \text{真實值})^2$$

我們希望模型預測得越準越好，所以會想辦法找到一個紅色橢圓裡的最小值點。

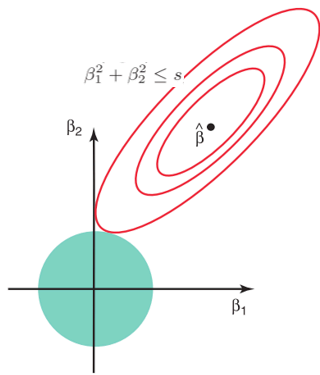
### 「最小值點」是什麼？

就是紅色橢圓最中心的那個點，代表模型的預測誤差最小、最準確的參數組合  $(\beta_1, \beta_2)$ 。





Lasso Regression



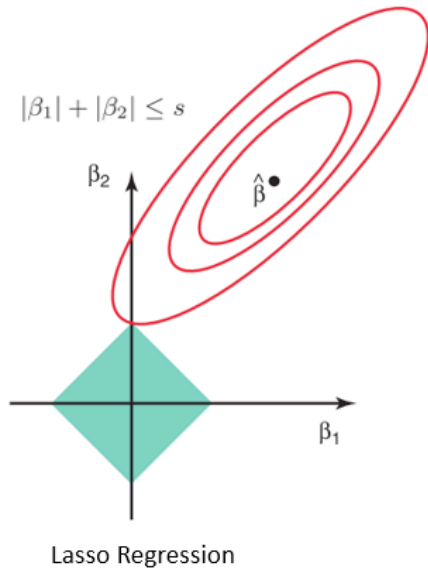
Ridge Regression

## 模型在做什麼？

模型的目標是找到紅色橢圓跟綠色限制區域交會的第一個點。  
這個點就是在「盡量準確」的同時，也「符合參數限制」的最佳平衡點。

想像你在一個下陷的山谷裡  
谷底是模型最準的地方  
但你只能在綠色的地板（限制區域）上走  
你就會停在「最靠近谷底、又不超出地板」的地方  
這就是**最小值點**

# Lasso regression

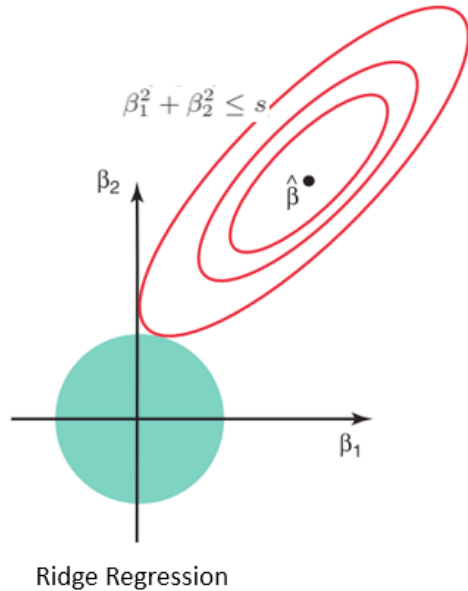


Lasso regression stands for “Least Absolute Shrinkage and Selection Operator”. It adds penalty term to the cost function. This term is the absolute sum of the coefficients.

As the value of coefficients increases from 0 this term penalizes, cause model, to decrease the value of coefficients in order to reduce loss.



# Ridge regression



Ridge regression decreases the complexity of a model but **does not reduce the number of variables** since it never leads to a coefficient been zero rather only minimizes it.

Hence, this model is not good for feature reduction.

# Lasso & Ridge regression

方法	限制形狀	幾何特性	結果
Lasso (L1)	菱形	有尖角	減少特徵，把參數變成0。 可以用來做重要Feature Extraction。(降維)
Ridge (L2)	圓形	光滑無角	參數只會縮小，不會歸零。 減少特徵權重差異，讓某些特徵的權重不要太突出。



Lasso 的菱形像是有「角落」，模型會被壓在角落 → 某些參數變 0  
Ridge 的圓形像是「光滑的碗」，模型只能滑到邊邊，參數只會縮小，不會消失

請問若要進行變數篩選的話，你要選擇哪種方法？

WHY ?

# Thanks!