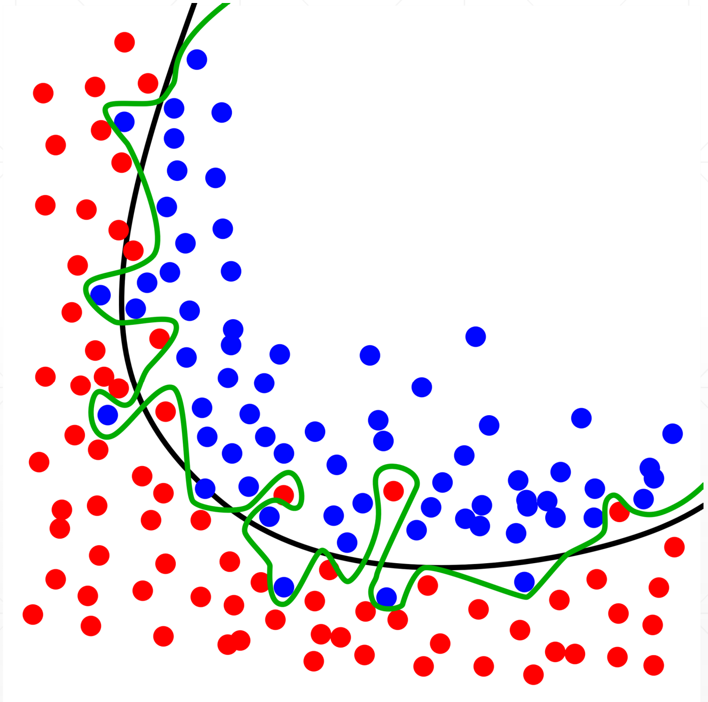
Lasso, Ridge及彈性網模型介紹

指導教授：　中山財管所 王昭文 教授

簡報製作：中山財管所 蘇彥庭 研究助理

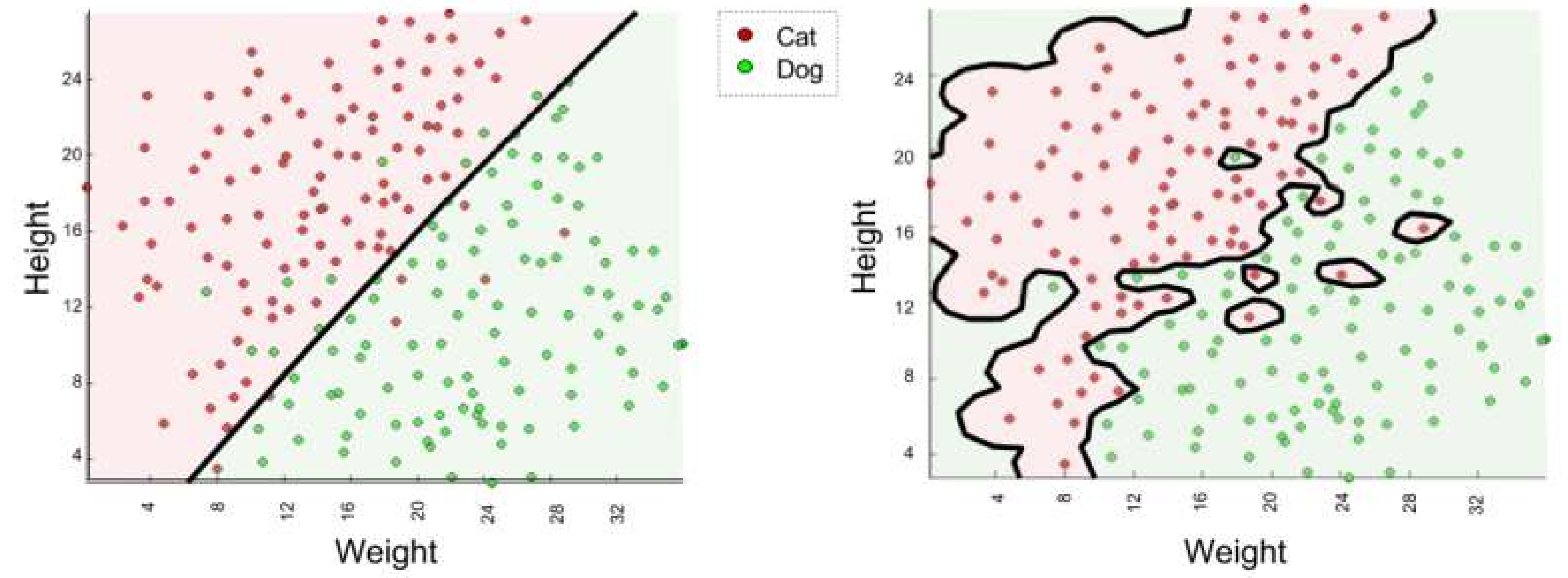
過度配適

- 黑色分界線和綠色分界線哪個比較好？



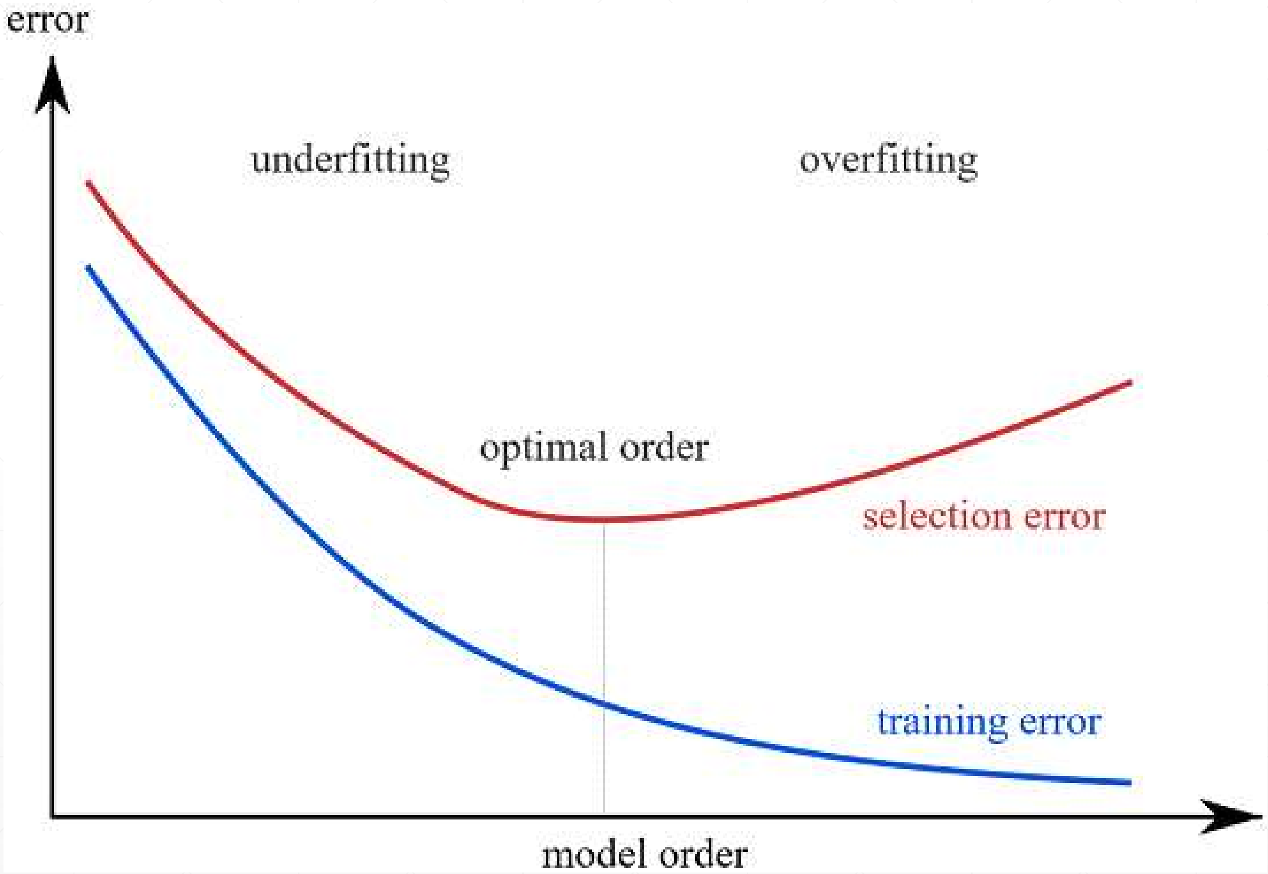
圖片來源：https://en.wikipedia.org/wiki/Overfitting

過度配適



圖片來源：https://kevinbinz.com/tag/overfitting/

過度配適



圖片來源：https://www.neuraldesigner.com/blog/model-selection

過度配適的生活例子

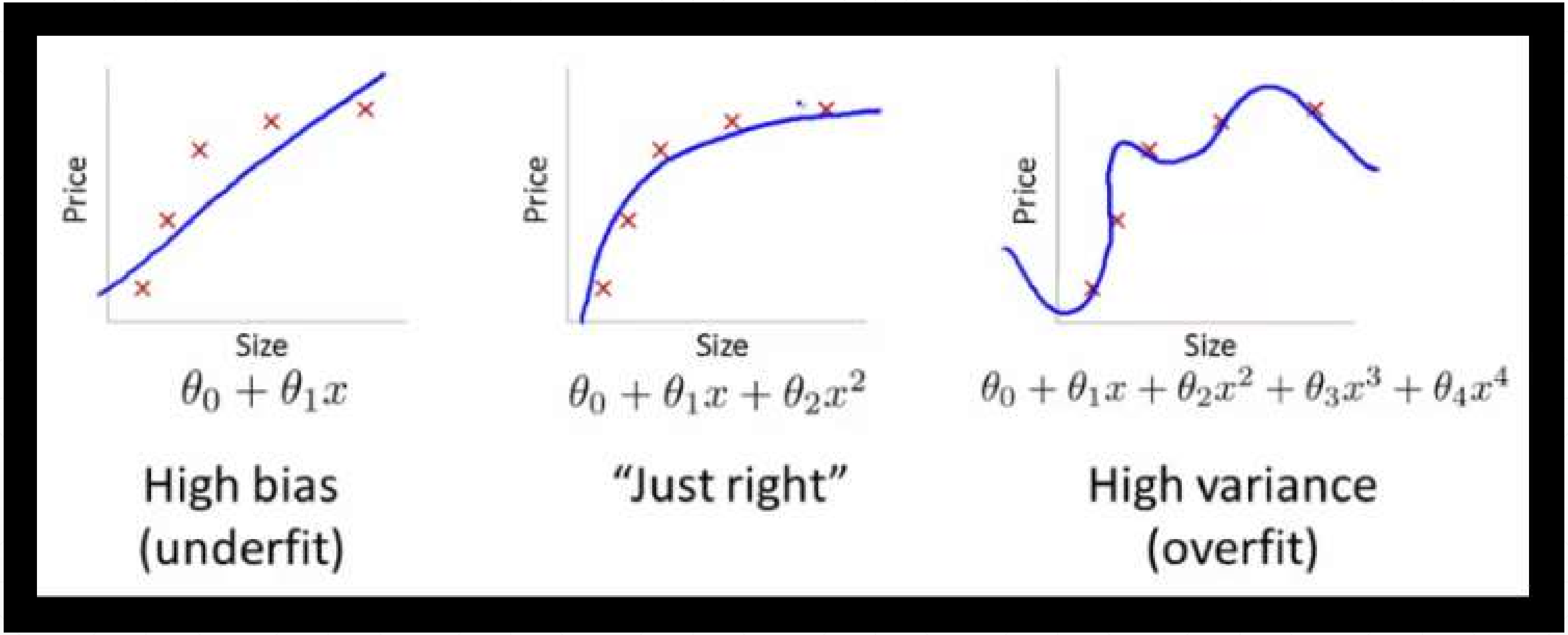
考試

駕照

圖片來源：https://i-chentsai.innovarad.tw/2018/04/education_policy.html
http://www.isinchu.com/blog/post/

過度配適

解釋變數愈多，模型複雜度愈高，發生過度配適的機會愈大



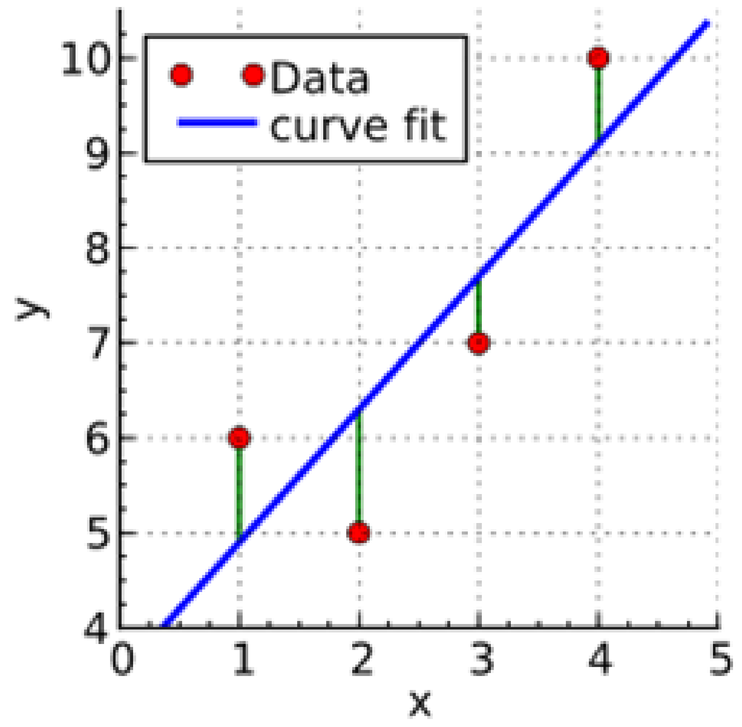
圖片來源：https://kevinbinz.com/tag/overfitting/

簡單線性迴歸模型

$$Y_i = \widehat{\beta_0} + \widehat{\beta_1} X_{i,1} + \widehat{\beta_2} X_{i,2} + \cdots + \widehat{\beta_p} X_{i,p}$$

最小平方法：

$$\min_{\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \ldots, \widehat{\beta_p}} \sum_{i=1}^{n} \widehat{Y}_i - {Y_i}^2$$



圖片來源：https://zh.wikipedia.org/wiki/%E6%9C%80%E5%B0%8F%E4%BA%8C%E4%B9%98%E6%B3%95

L1 / L2 正規化

- 最小平方法

$$\min_{\widehat{\beta_0},\widehat{\beta_1},\widehat{\beta_2},\dots,\widehat{\beta_p}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2}$$

- Lasso迴歸

$$\min_{\widehat{\beta_0},\widehat{\beta_1},\widehat{\beta_2},\dots,\widehat{\beta_p}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2} + \lambda\,[|\widehat{\beta_0}| + |\widehat{\beta_1}| + |\widehat{\beta_2}| + \cdots + |\widehat{\beta_p}|]$$

- Ridge迴歸

$$\min_{\widehat{\beta_0},\widehat{\beta_1},\widehat{\beta_2},\dots,\widehat{\beta_p}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2} + \lambda\,[\widehat{\beta_0}^{\,2} + \widehat{\beta_1}^{\,2} + \widehat{\beta_2}^{\,2} + \cdots + \widehat{\beta_P}^{\,2}]$$

L1 / L2 正規化

- 最小平方法

$$\min_{\widehat{\beta_0},\widehat{\beta_1},\widehat{\beta_2},\dots,\widehat{\beta_p}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2}$$

- Lasso迴歸

$$\underset{\widehat{\beta_0},\widehat{\beta_1},\widehat{\beta_2},\dots,\widehat{\beta_p}}{\operatorname{argmin}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2} + \underbrace{\boxed{\lambda}}_{\text{懲罰參數}} \overbrace{\left[\left|\widehat{\beta_0}\right| + \left|\widehat{\beta_1}\right| + \left|\widehat{\beta_2}\right| + \cdots + \left|\widehat{\beta_p}\right|\right]}^{\text{L1正規化}}$$

- Ridge迴歸

$$\underset{\widehat{\beta_0},\widehat{\beta_1},\widehat{\beta_2},\dots,\widehat{\beta_p}}{\operatorname{argmin}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2} + \underbrace{\boxed{\lambda}}_{\text{懲罰參數}} \overbrace{\left[\widehat{\beta_0}^{\,2} + \widehat{\beta_1}^{\,2} + \widehat{\beta_2}^{\,2} + \cdots + \widehat{\beta_P}^{\,2}\right]}^{\text{L2正規化}}$$

L1 / L2 正規化

- 若懲罰參數λ=0，則為一般的迴歸式。
- 懲罰係數設定愈大，模型複雜度愈低。但愈大不一定愈好，因為有可能變為配適不足問題(under-fitting)。
- L1正規化(Lasso迴歸)可將不重要的解釋變數(即係數項值接近0)的係數收縮至0，達成特徵挑選(Feature Selection)的效果，降低模型複雜度，防止過度配適(Over-fitting)的問題。
- L2正規化(Ridge迴歸)可讓解釋變數權重變小(Weight Decay)，降低模型複雜度，防止過度配適問題發生，但係數項值不會如Lasso迴歸收縮至0。

L1 / L2 正規化

Lasso Regression:

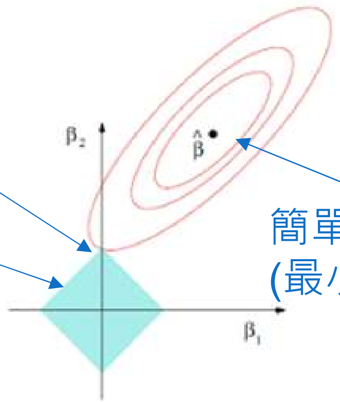
$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2$$

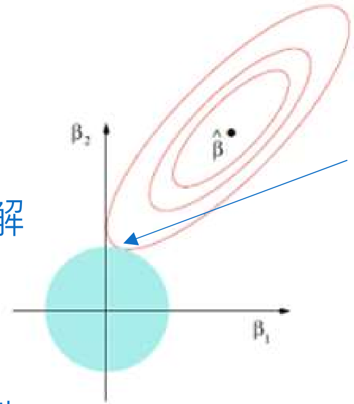
$$\text{subject to} \sum_{j=1}^{p} |\beta_j| \leq t. \quad \text{(L1 term)}$$

Ridge Regression:

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2,$$

$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 \leq t, \quad \text{(L2 term)}$$





t值愈小=> 正方形愈小=> 模型愈簡單 => 極端狀況為只有截距項的模型

t值與λ值成反比

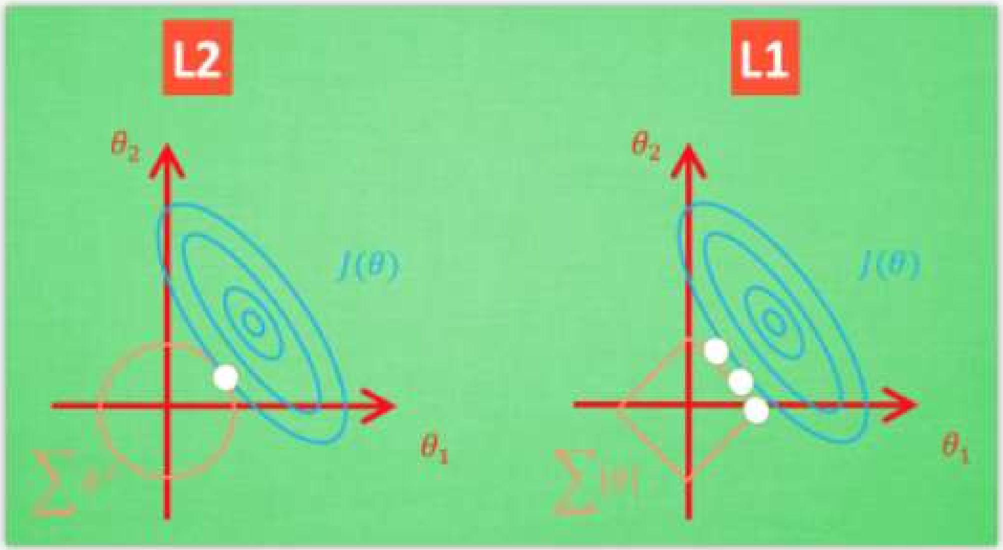
圖片來源：https://rpubs.com/skydome20/R-Note18-Subsets_Shrinkage_Methods

L1 / L2 正規化

- 差異比較
- 可閱讀此篇文章：http://www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/

L2 loss function	L1 loss function
Not very robust	Robust
Stable solution	Unstable solution
Always one solution	Possibly multiple solutions

L2 regularization	L1 regularization
Computational efficient due to having analytical solutions	Computational inefficient on non-sparse cases
Non-sparse outputs	Sparse outputs
No feature selection	Built-in feature selection



表格來源：http://www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/

圖片來源：https://morvanzhou.github.io/tutorials/machine-learning/ML-intro/3-09-l1l2regularization/

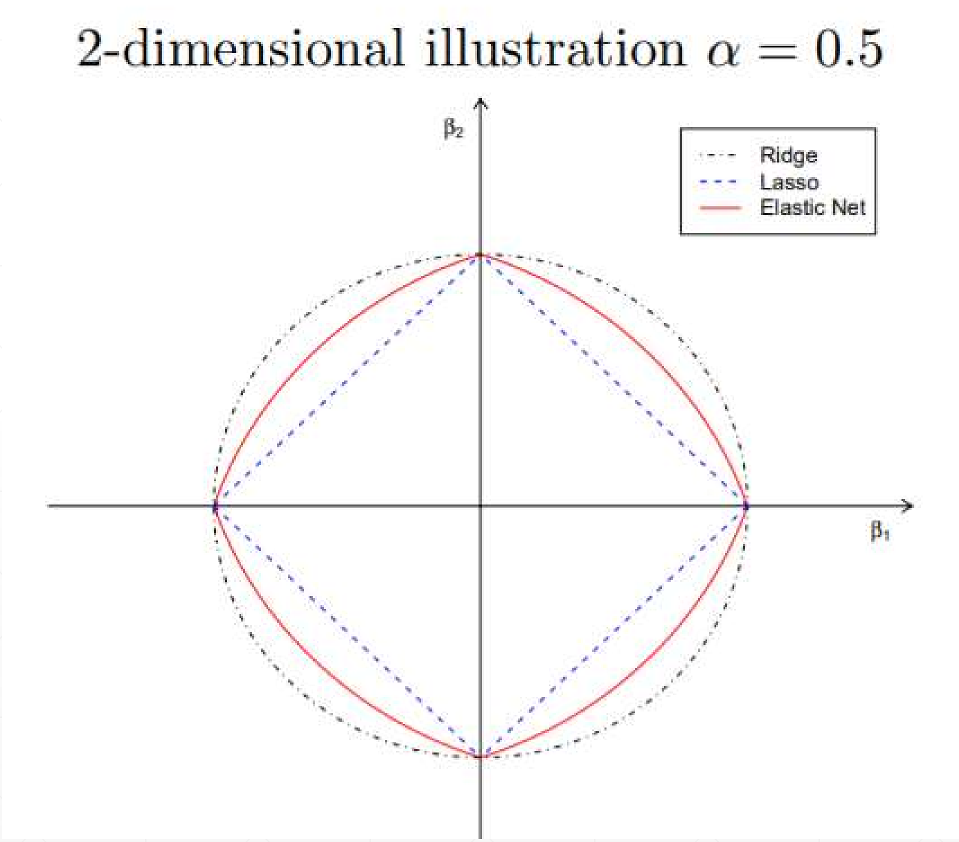
彈性網(Elastic Net)

- 融合L1正規化與L2正規化。
- 除原本的懲罰參數λ外，再多一個α 參數。
- 若$\alpha = 1$，則為Lasso迴歸模型。若$\alpha = 0$，則為Ridge迴歸模型。

$$\underset{\widehat{\alpha},\widehat{\beta_1},\widehat{\beta_2},\ldots,\widehat{\beta_p}}{\operatorname{argmin}} \sum_{i=1}^{n} \widehat{Y_i} - Y_i^{\,2} + \underbrace{\boxed{\lambda}}_{\text{懲罰參數}} \sum_{j=0}^{p} \alpha \underbrace{\widehat{\beta_j}^{\,2}}_{\text{L2正規化}} + \ (1-\alpha) \underbrace{\left|\widehat{\beta_1}\right|)}_{\text{L1正規化}}$$

圖片來源： https://web.stanford.edu/~hastie/TALKS/enet_talk.pdf

彈性網(Elastic Net)

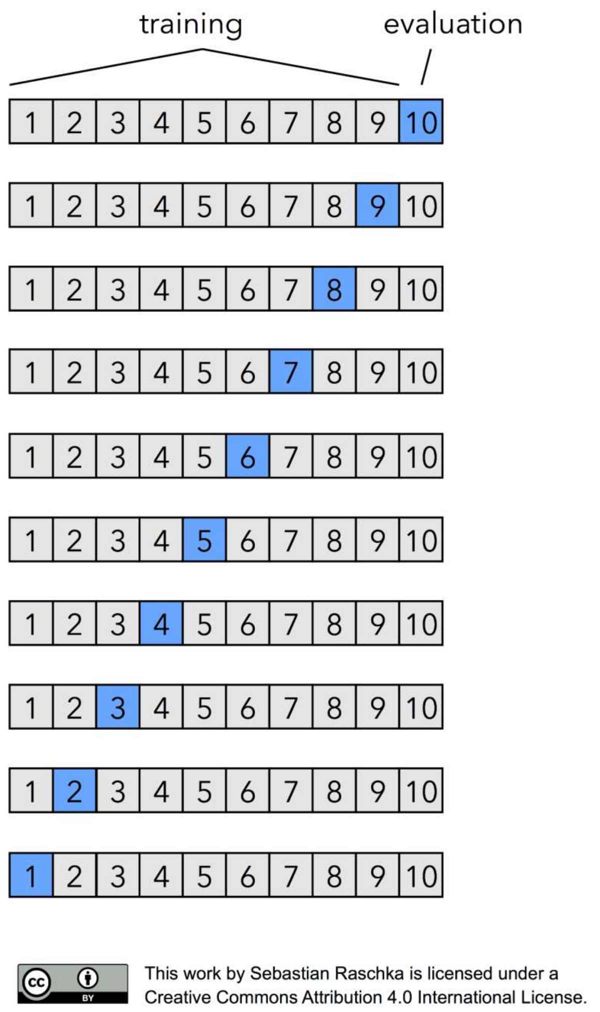


圖片來源： https://web.stanford.edu/~hastie/TALKS/enet_talk.pdf

留一驗證

- 留一驗證(Leave-one-out cross-validation; LOOCV)
- 假設有10筆訓練集樣本
- 每次依序取1筆樣本做為驗證集
- 剩餘9筆樣本進行模型訓練
- 將訓練好的模型來預測驗證集
- 每個樣本皆會得到模型的預測，計算留一驗證準確率

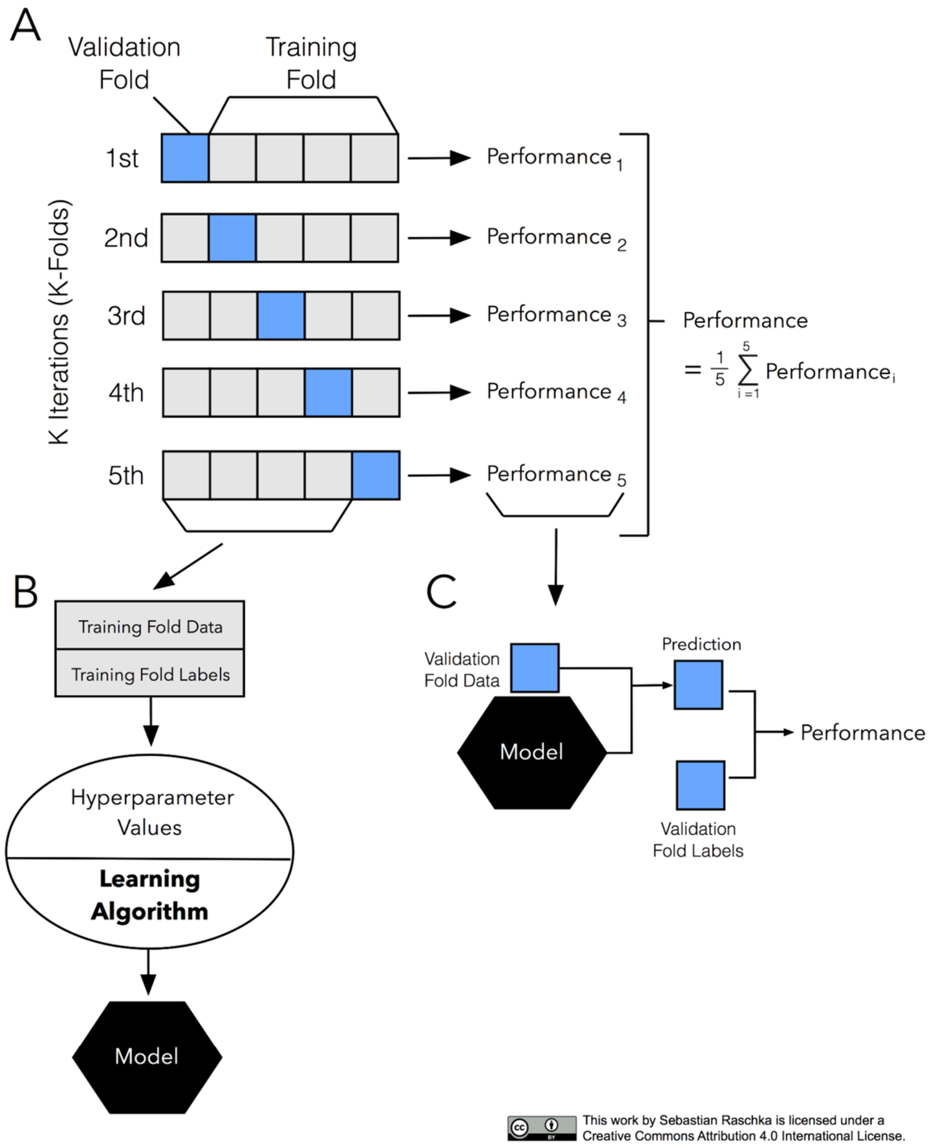
- 留一驗證方法缺點：
- 當訓練期資料樣本非常大時，速度太慢



圖片來源：https://sebastianraschka.com/blog/2016/model-evaluation-selection-part3.html

K次交叉驗證

- K次交叉驗證(K-fold Cross-validation)
- 實務上常用作法
- 使用者設定K值，將訓練期樣本分成K組
- 每次依序拿其中1組做為驗證集
- 其餘的 K-1組進行模型訓練
- 將訓練好的模型來預測驗證集
- 每組驗證集皆會得到準確率
- 平均每組的準確率得到 K次交叉驗證準確率



圖片來源：https://sebastianraschka.com/blog/2016/model-evaluation-selection-part3.html

HW1 作業

Homework (3 weeks)

- 運用**S&P500(注意時差)、台灣50、黃金價格、美元對台幣**等四個價格
- Predict the returns of 4 assets (For 2013/1/1-2018/8/31)
- For predictor variables, we choose **lagged 1M, 3M, 6M and 12M** returns of these same 4 assets, yielding a total of 16 variables.
- To calibrate the model, we used a rolling window of 500 trading days (~2y); re-calibration was performed once every 3 months.
- **Strategies:**
  - **The model was used to predict the next day's return.**
  - **If the next day predicted return R>0 => long the asset**

    R<0 => short it

=>請做出運用Linear Regression、Lasso與Elastic Net之績效為何??