Linear Regression

Agenda

- Linear Regression
  - OLS
  - Ridge Regression
  - LASSO
- Demo

Linear Regression-OLS

- **核心精神**
  - 多增加一個x, 對y額外的解釋力
- **目標式**

$$\min_{w} ||Xw - y||_2^2$$

- **參數**
  - 沒有參數設定, 為係數估計, LR的Beta不能動
- **Linear Regression in Statistics v.s. Linear Regression in ML**
  - 模型假設。EX:共線性問題
  - 變數選擇。EX:Stepwise

Linear Regression-Regularization/Penalization

- **核心精神**
  - Avoid multicollinearity
  - Avoid overfitting
  - Feature selection: Beta皆可以動

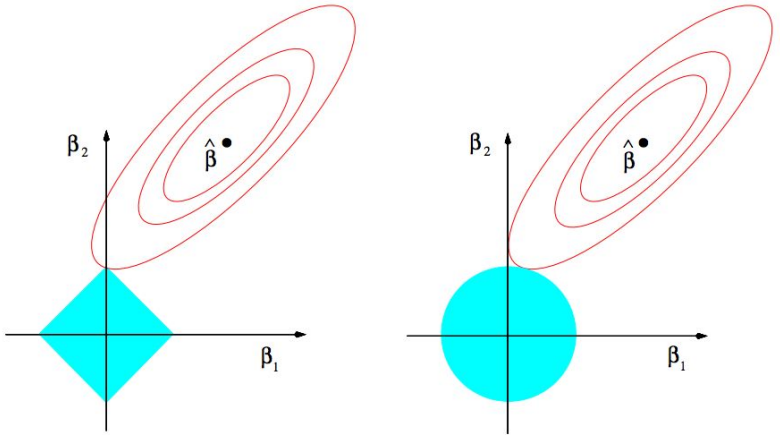
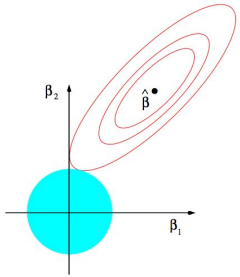


FIGURE 3.11. *Estimation picture for the lasso (left) and ridge regression (right). Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions* $|\beta_1| + |\beta_2| \leq t$ *and* $\beta_1^2 + \beta_2^2 \leq t^2$*, respectively, while the red ellipses are the contours of the least squares error function.*

Linear Regression-Regularization/Penalization

Ridge Regression

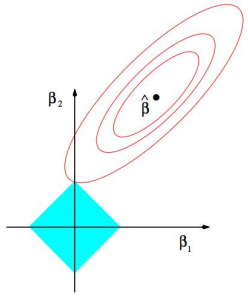


- **核心精神**
  - 加了 L2 term,
    sum of square of the magnitude of weights
  - Beta趨近於0
- **目標式**

$$\min_w ||Xw - y||_2^2 + \alpha||w||_2^2$$

- **參數**
  - alpha

LASSO



- **核心精神**
  - 加了L1 term,
    sum of absolute value of the magnitude of weights
  - Beta可能會有0
- **目標式**

$$\min_w \frac{1}{2n_{samples}}||Xw - y||_2^2 + \alpha||w||_1$$

- **參數**
  - alpha

Demo

Data & Packages

https://archive.ics.uci.edu/ml/datasets/Las+Vegas+Strip

```
import numpy as np
import pandas as pd
from sklearn import cross_validation
from sklearn import linear_model
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
```

Process

- Data Preparation
  - Missing values handling
  - Categorical variable handling (EX:One hot encoding)
  - Normalization
  - Feature combination
- Modeling
  - Partition: training/testing
  - Avoid overfitting: k fold validation, regularization, simple model
  - Parameter tuning: grid search
- Evaluation
  - RMSE/MSE
  - F1 Score/Precision/ROC Curve

Model Comparison

- 模型選擇可再調整(EX:SVM, 預測最適平面)
- 比較三個模型後, 加入Regularization結果有變好

```
------------------------------------------------
Compare Score:
 -0.2366563347013808 -0.1949494041577460l8 -0.07871713952384263
Compare MSE:
 1.2315178734768597 1.1899838793548392 1.0742349353429956
```

Appendix

Linear Regression Algorithm

1. 要計算的Weight:利用X之假的反矩陣(pseudo inverse)，再透過和y相乘
2. y hat:要計算的Weight乘上X
3. 如何找到梯度為0的證明，其中提到pseudo inverse

Linear Regression Algorithm

很快算出結果, 看不見學習的過程, 稱為analytic solution。

a. 過程當中計算的pseudo inverse, 即為學習過程。
b. 而再加上最後出來的Eout是不錯的, 沒有理由說他不是機器學習演算法。

Linear Regression Algorithm

利用幾何空間去詮釋y-y hat (error), 最後證明E(in)和E(out)當n很大時, 會收斂至相等, 而保證機器是有學到東西(VC dimension)

- both converge to σ^2 (**noise** level) for $N \to \infty$