LASSO : LEAST ABSOLUTE SHRINKAGA & SELECTION OPERATOR.

1. Linear Regression and Ridge regression
   → suitable for low-dimensional situation.
   → ~~A~~ $n > d$

2. LASSO → suitable for high-dimensional situation
   → d can be exponentially greater than 'n'.

How do we achieve this? → make assumptions!

Assumption : Only 's' of the 'd' components of β are non-zero!

So in $Y = \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_d X_d + \varepsilon$,

$\beta_1 X_1$ ↓ non-zero, $\beta_2 X_2$ ↓ zero, ... $\beta_d X_d$ ↓ nonzero

→ the above is an ~~sparse~~ example.

Scaling for Lasso (ie) conditions under which LASSO works:

$\boxed{n > s \cdot (\log \cdot d)}$ but $\boxed{d > n}$.

Ideal LASSO :

$$\min_{\beta} \|Y - X\beta\|_2^2 \quad \text{such that} \quad \|\beta\|_0 \le s.$$

where ~~$\|\beta\|$~~ $\|\beta\|_0 \rightarrow$ ~~number~~ number of non-zero entries in a vector β.

Issues : Combinatorial optimization problem !

Time exponential in 'd' is needed !

SOLUTION $\rightarrow$ Just Relax !

Relax the $\|\beta\|_0 \longrightarrow \|\beta\|_1$

$$\|\beta\|_1 = \cancel{\|\beta\|} \; |\beta_1| + |\beta_2| + \cdots \cancel{|\beta_i|} + |\beta_d|$$

$\rightarrow$ sum of absolute values of 'β'.

FACT : $\|\beta\|_1$ is the TIGHTEST convex relaxation of $\|\beta\|_0$.

Advantage : $\min_{\beta} \|Y - X\beta\|_2^2$ such that $\|\beta\|_1 \le z$

$\rightarrow$ This problem is convex & could be solved in polynomial time (ie. faster)

Unconstrained LASSO :

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \| Y - X\beta \|_2^2 + \lambda \|\beta\|_1 \quad ①$$

(→ called as Lagrangian form.)

for some $(\lambda > 0)$.

~~To gain~~ ~~sorry~~

→ Compared to Ridge Regression that has a closed form solution, LASSO does not have a closed form solution.

→ 'glmnet' command in 'R' solves it algorithmically.

To gain intuition : Assume $X^T X = I$.

In this case, recall $\hat{\beta} = X^T y$

where $\hat{\beta}$ → solution of linear regression.

Equation ① above becomes,

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \{ y^T y - 2y^T X\beta + \beta^T\beta + \lambda \|\beta\|_1 \}.$$

$$= \underset{\beta}{\operatorname{argmin}} \{ -2\hat{\beta}\beta + \beta^T\beta + \lambda \|\beta\|_1 \}.$$

[as $y^T y$ has no part in minimization]

$$= \underset{\beta}{\operatorname{argmin}} \sum_{j=1}^{d} \left\{ -2\hat{\beta}_j \beta_j + \beta_j^2 + \lambda |\beta_j| \right\}.$$

So the optimization can be done separately for each co-ordinate of β.

Fix a 'j'. Then

$$(\hat{\beta}_L)_j = \underset{\beta}{\operatorname{argmin}} \; -2\hat{\beta}_j \beta_j + \beta_j^2 + \lambda |\beta_j| .$$

Claim 1: If $\hat{\beta}_j > 0$ then $(\hat{\beta}_L)_j \geq 0$

claim 2: If $\hat{\beta}_j \leq 0$ then $(\hat{\beta}_L)_j \leq 0$.

Using the above claims it is easy to show

$$(\hat{\beta}_L)_j = \operatorname{sign}(\hat{\beta}_j) \left[|\hat{\beta}_j| - \lambda/2 \right]^+$$

where ① $\operatorname{sign}(a) = \begin{cases} +1 & \text{if } a \geq 0 . \\ -1 & \text{if } a < 0 \end{cases}$

② $[a]^+ = \max(0, a)$

INTERPRETATION: LASSO pushes the Linear Regression solution to zero for some co-ordinates by selection 'λ' appropriately!.