Elastic Net
(Lasso + Ridge)

Ridge

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boxed{\lambda \|w\|^2}$$

$\lambda(w_1^2 + w_2^2 + w_3^2 \ldots)$

$$m = \frac{(y_i - \bar{y})(x_i - \bar{x})}{(x_i - \bar{x})^2 \boxed{+\lambda}}$$

⇒ we use Ridge reg. in scenario when all features are imp. for us. we don't want to remove any of the feature by making their co-eff = 0

Lasso

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boxed{\lambda \|w\|}$$

$\lambda(|w_1| + |w_2| + \ldots + |w_n|)$

m

$m > 0$

$$m = \frac{(y_i - \bar{y})(x_i - \bar{x}) \boxed{-\lambda}}{(x_i - \bar{x})^2}$$

$m = 0$

$$m = \frac{(y_i - \bar{y})(x_i - \bar{x})}{(x_i - \bar{x})^2}$$

$m < 0$

$$m = \frac{(y_i - \bar{y})(x_i - \bar{x}) \boxed{+\lambda}}{(x_i - \bar{x})^2}$$

⇒ If any particular input columns are not imp. or all features are not important for us, in that case we will use Lasso reggression

Scenario-1

* unlike these above two situation, if we have huge no. of cols where not able to decide which should i use Ride or Lasso?

in that case we will use *** $\boxed{\text{Elastic Net}}$

$$L = \sum (y_i - \hat{y}_i)^2 + \boxed{a} \|w\|^2 + \boxed{b} \|w\|$$

scenario 2:-
If we have data having multicollinearity
then also we can use * Elasticnet.

in sklearn implementation

$\lambda = a + b$

$l_1\_ratio = \dfrac{a}{a+b}$

⇒ $l_1 = \dfrac{a}{\lambda}$

$a = l_1 * \lambda$

$b = \lambda - a$

Ex:- If $\lambda = 1$, $l_1\_ratio = 0.5$
then, automatically $a = 0.5$, $b = 0.5$

In this case we have to apply lasso, ridge in equal amount.

Scanned with OKEN Scanner

- Scenario 2:-
  - If we have data having multicollinearity
    - then also we can use ElasticNet.

Q1. If by applying GD → we can get best parameter $\beta_1; \beta_2 \dots$ then why we need regularization?

Gradient descent will give best parameter only on seen (sample) data, but not necessarily give best of unseen leads to overfitting to reduce overfitting we require Regularization

Q2 why we dont use intercept (β_0) in L_1 or L_2 Norm or in regularization

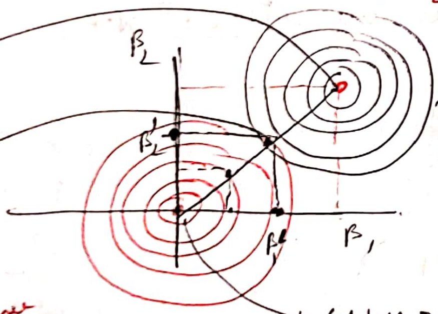
∵ we are using only m to change regression line.

Q3. which one is more robust to outlier? Ridge, Lasso, ElasticNet?

Q Explain diagramatically as $\lambda\uparrow$ how, β_1 & β_2 reduces.

Actual solution was this

but we are taking this solution.



Loss function

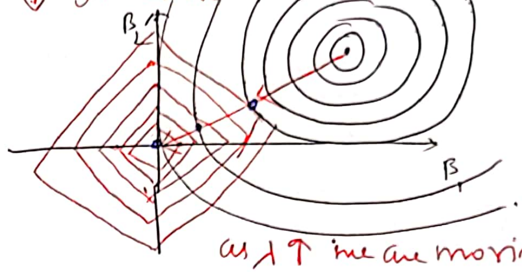
as we reduce λ ↑ circle

to bring up to this solution we are reducing (β_1, β_2) → (β_1', β_2') by increasing $\lambda\uparrow$.

⇓

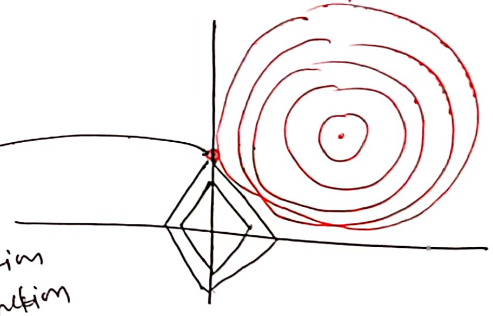
as we are increase λ ↓ it is going towards smaller concentric circle which is decreasing β_1 & β_2 as well as moving away from min Loss

Q Explain diagramatically as $\lambda\uparrow$ how ~~Loss function~~ minima of Loss function shifting back and β_1 & β_2 is also ↓ in Lasso



as $\lambda\uparrow$ we are moving ~~too~~ towards smaller concentric square. due to which min[ma] Loss is moving in upward direction. and $(\beta_1, \beta_2) > (\beta_1', \beta_2')$

Q why does Lasso creates sparsity?



Probability of intersection of Loss function with contour with diamond shaped Lasso concentric diamond is more on sharp-corner.

due to this one of the co-eff becomes 0, and other becomes Non-zero.