

L1 & L2 Regularization -

Lasso -

- ① Regularization tech. used in feature select using a shrinkage method also referred to as penalized regression method.
 - ② Lasso is short for Least Absolute Shrinkage & Selection operator used for both regularization & model selection. Referred as L1 Regularization.
 - ③
$$L_1 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^p |B_i|$$

Coeffs. determined in linear model from this eqn. are shrunk towards the central pt. & as mean by introducing a penalization factor called the α (or λ) values.
 - ④ ' α ' penalty term that denotes the amt. of shrinkage (or constraint) that will be applied implemented in eqn. α set to 0, equivalent of the lin. reg. model. Larger val. penalizes optimization fn.
 - ⑤ So, Lasso shrinks coeffs & help to reduce model complexity & multi-collinearity.
 - ⑥ α can be any real valued no. b/w 0 & ∞ ; Larger the val., more aggressive penalization is.
- ~~Draw~~ Due to fact that coeffs. will be shrunk towards a mean of 0, less imp. features in a dataset are eliminated when penalized.

- Shrinkage of these coeffs. based on some α val. provided leads to some form of automatic feature selectⁿ, as i/p vars are removed in an effective approach.
- λ is regularizatiⁿ param that controls amt. of regularizatiⁿ applied.
- $\beta_1, \beta_2, \dots, \beta_p$ are coeffs.

① Linear reg. model:

LASSO reg. starts with lin. reg. model that assumes a linear relationship b/w independent vars. (features) & dependent var. (target).

② L_1 regularizatiⁿ:

LASSO reg. intro. ~~absolute~~ additional penalty term based on abs. vals. of coeffs. L_1 reg. term is sum of abs. vals. of coeffs. mul. by tuning param. λ .

$$L_1 = \lambda * (|\beta_1| + |\beta_2| + \dots + |\beta_p|)$$

③ Objective fn:

Obj. of LASSO is to find vals. of coeffs. that min. sum of squared diff. b/w pred. & actual vals.

④ Shrinking coeffs:

Adding L_1 reg. term, LASSO can shrink coeffs. towards 0. When λ sufficiently large, some coeffs are driven to exactly 0. This prop. of LASSO makes it useful for feature selectⁿ as vars. with zero coeffs. are effectively removed from model.

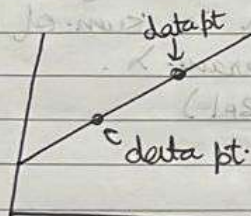
⑤ Tuning Param λ :

Denote amt. of shrinkage. Larger λ val. \uparrow amt. of regularization while leading to more coeffs. being pushed towards zero. Smaller λ val. reduces regularization effect leading to allowing more vars. to have non-zero & coeffs.

Ridge Reg:

- Similar to LASSO, puts similar constraint on coeffs. by introducing penalty factor. While Lasso takes mag. ridge takes square. L2 Reg.
- Reduce model complexity by coeff. shrinkage.
- Shrinks parameters, so used to avoid multi-collinearity.

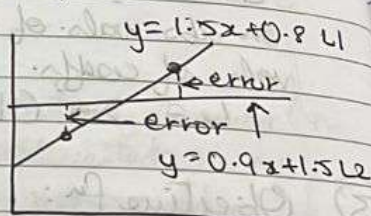
$$L_2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^p \beta_i^2$$



Overfitting

Loss for line 1

$$\lambda=1 \rightarrow 0 + (1.5)^2 = 2.25$$



Loss for line 2

$$\lambda=1 \rightarrow$$

- From above fig model will choose line 2 as best fit line cuz although it is giving error in training, due to ridge λ added & cost is reduced.