Lasso Regression (L1)

Sunday 14 April 2024 12:34 AM

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \|\beta\| \text{, where } \lambda \geq 0$$

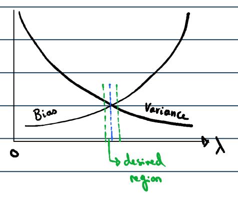
$\hookrightarrow \lambda(|\beta_1| + |\beta_2| + |\beta_3| + \dots + |\beta_n|)$

* coeffiencient <u>may become</u> 0 for higher λ values, unlike in Ridge
  - useful in feature selection while dealing with high dimensionality data.

* higher coefficients impacted more compared to lower coefficient values

* feature selection happens for intermediate value of λ.

* Bias variance tradeoff



* Effect on Loss function

→ As $\lambda \uparrow$
- loss function tends to move towards origin & will sit on top of origin for very large λ value.
- loss function is translated upwards
- it also shrinks But after a certain value of λ we tend to see an angle created at the origin.
- Won't go below zero.
- curve loses it shape.

kink

* Sparsity [coefficients become 0]

<u>Lasso</u>

for $m > 0$

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

for $m = 0$

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

for $m < 0$

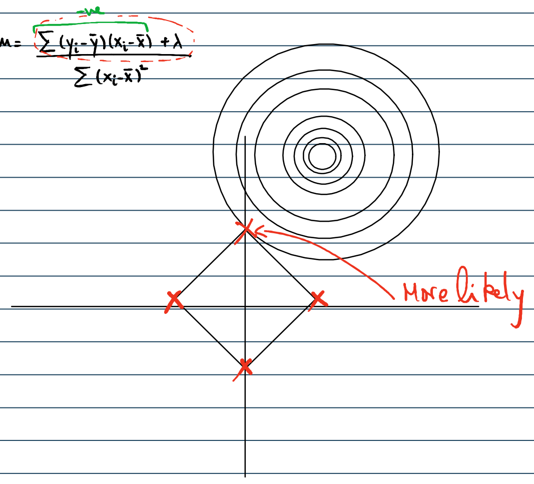
$$m = \frac{\overbrace{\sum (y_i - \bar{y})(x_i - \bar{x})}^{-ve} + \lambda}{\sum (x_i - \bar{x})^2}$$

* as λ increases

start m(-ve) → 0 ← start m(+ve)

stops at zero

* Ridge λ in denominator
  Lasso λ in numerator



* Can deal with multicolinearity, it will completely decrease the coefficient of one of the colinear features.

class `sklearn.linear_model.`**`Lasso`**(*alpha=1.0, *, fit_intercept=True, precompute=False, copy_X=True, max_iter=1000, tol=0.0001, warm_start=False, positive=False, random_state=None, selection='cyclic'*) [source]