Advance Regression

Sunday, 18 April 2021 11:53 AM

Generalized vs Regularized Regression

GLR :- Refers to conventional linear models for a continuous response variable given continuous/categorical predictors.
eg :- Simple linear Regression,
⇒ Binary logistic Regression.

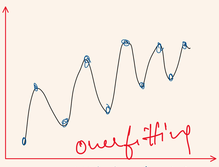
Regularized Regression :-
Are type of Regression where the coefficient estimates are constrained to zero. The size of Coefficient and the size of low terms are penalized.
(Complex models are unwarranted due to overfitting.)

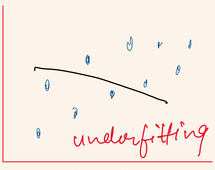
Types of Regularizers
1. L_1
2. L_2
3. Elastic Net

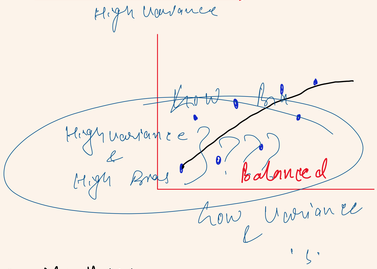
Recap :-

Bias :- Biases are the underlying assumptions that our model makes.
↓ Training time due to less sensitive nature
may lead to underfitting

Variance :-
→ Happens due to model's sensitivity to small fluctuations of data.
→ Commonly Refered as overfitting
→ The model basically learns every single point and doesn't do well on novel dataset that it hadn't seen before.





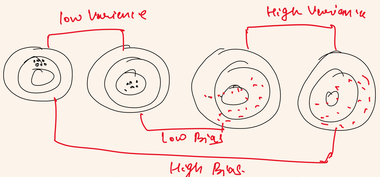


Consider this :-

$Y = \beta * X + \epsilon$

$Y = \hat{\beta} X + \beta_1$

Goal is to get $\hat{\beta}$ that produces lowest RSS.



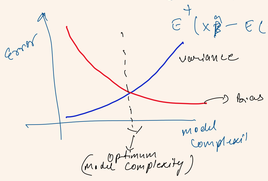
$$L(\hat{\beta}) = \sum_{i=0}^{n} || y_i - x_i * \hat{\beta}_i ||^2 = ||Y - X\hat{\beta}||^2$$

$$(\hat{\beta}) = (X^T X)^{-1} (X^T Y)$$

Bias = $E(\hat{\beta}) - \beta$
Variance = $\sigma^2 (X^T X)^{-1}$

Error - term = $(E(X\hat{\beta}) - X\beta)^2$

$E^2(X\hat{\beta} - E(X\hat{\beta}))^2 + \sigma^2$



If Predictions are ⟶ scattered → Yes → High Variance
↓ No
⟶ far from your target → Yes → (Crappy Model)
↓ Yes
High Bias

Ridge Regression :- (L^2)

↳ we add penalty term which is equal to the square of the coefficient.
The L^2 term is equal to the square of the magnitude of the coefficients.
We also add λ to control the penalty.

$$\text{Ridge} = \underset{\hat{\beta}}{\operatorname{argmin}} \left(||Y - \beta * X||^2 + \lambda * |\beta|_2^2 \right)$$

Limitation :- Reduces the model complexity but doesn't make the coeff zero meaning it doesn't reduce the number of parameters.
(Not good for feature Reduction)

Lasso Regression :- (L_1)

(Least abs. Shrinkage and Selection Operator.)

→ It adds penalty term to the cost function.
→ This term is the absolute sum of coefficient.

$$\text{Lasso} = \underset{\hat{\beta}}{\operatorname{argmin}} \left(||Y - \beta * X||^2 + \lambda * ||\beta||_1 \right)$$

Can make coefficient '0' hence suitable for feature Reduction.

Limitation :- Sometimes struggles with some specific type of data where n of predictors > no. of observation.
→ If there two highly colinear variable, Lasso will select 1 randomly which is bad for interpretability.

Elastic Net :-
Sometimes the lasso produces very simple model where the prediction is too dependent on 1 variable.
In that case we use Elastic Net because it combines the best of both.