by
batch
wise

29th April

Regularization

Overfitting - low bias high variance.
Underfitting - high bias low variance.

* What are technique we have to reduce overfitting technique.

(1) PCA (2) Regularization (3) Ensamble Learning

(4) Cross validation (5) drop out.

Regularization # Regularize the coefficient of independent variable.

* Train the model with more attributes.
* Train the model with high coefficient

] overfit

($y = m_1 x_1 + m_2 x_2$)

3 steps :- Interview

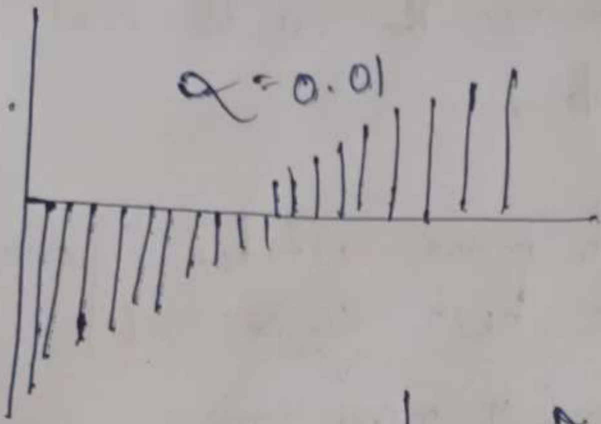
→ Lasso regularization | L_1 regression
→ Ridge regularization | L_2 regression
→ elasticnet regularization | $L_1 + L_2$

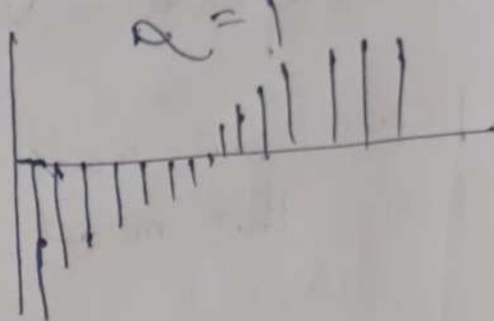
Ridge Regression

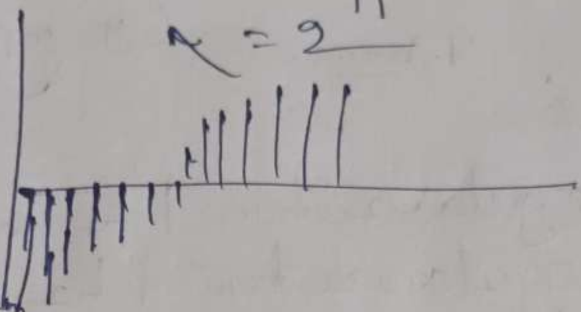
Regularization
- Ridge
- Lasso

→ Ridge R = Loss + $\alpha \|w\|^2$ (Penalty)

$$\|w\|^2 = w_1^2 + w_2^2 + w_3^2 + \cdots + w_n^2$$







scaled down so, problem

$y = 0.9 + 1.2x_1 + 20x_2 + 39x_3$

$y = 0.9 + 0.7x_1 + 2x_2 + 5x_3$

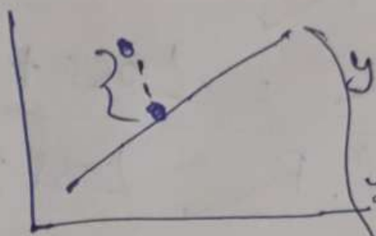
overfitting

* L_2 = is the technique to scale down high co-efficient to low coefficient to reduce overfitting.

* L_1 =

Lasso Regression (L_1)

Regularization
- Ridge
- Lasso



scaled down

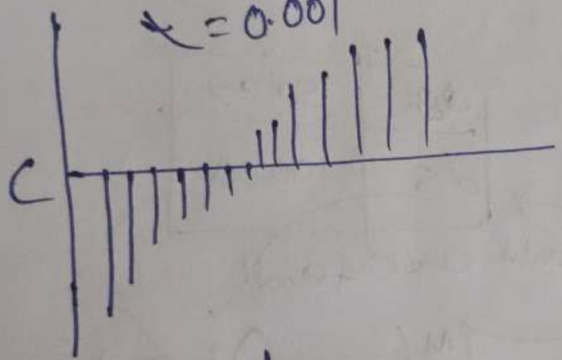
$y = 0.1 + 1.2x_1 + 20x_2 + 39x_3$

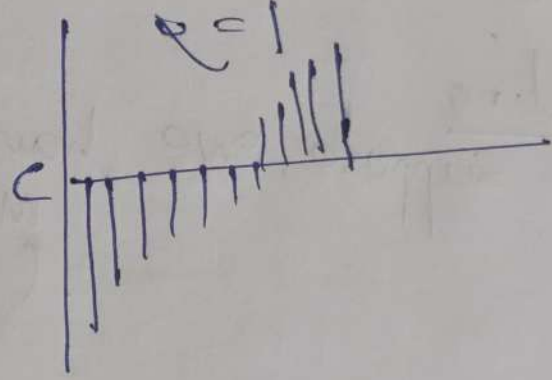
$y = 0.9 + 0 \cdot x_1 + 0x_2 + 5x_3$

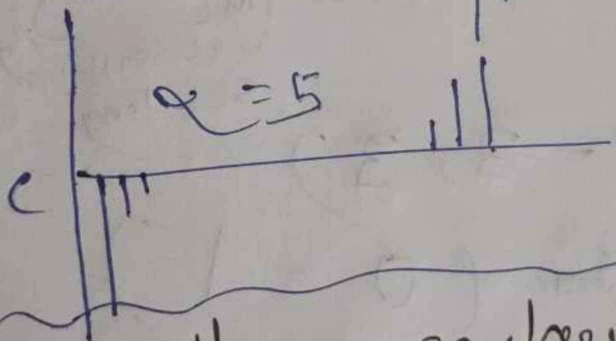
2 variable ellemenated (overfitting gone)

Lasso R = Loss + $\alpha \|w\|$ (penalty)

$\|w\| = w_1 + w_2 + w_3 + \dots + w_n$







L_1 = It is the regularization technique to **scale down high coef to 0** thats l_1 also called **feature selection technique / feature ellemenation**.

Elastic Regression

Ridge R = Loss + $\alpha ||w||^2$

+

Lasso R = Loss + $\alpha ||w||$

$$\boxed{\text{Elastic net. R} = \text{Loss} + \alpha_1 ||w||^2 + \alpha_2 ||w||}$$

Ml algorithm (l_1, l_2 as system parameter automatically)

(like mean, median)

Scaling

Suppose, we have

age	salary
25	1000000

ML → don't understand

ML understand 0,1 0, 1

scaling - certain Number

z - score (- 3, 3)

Normalization (0 - 1)

Feature scaling :-

Normalization (Min Max scaler) (min - 0, max - 1)
Standardization (Standard Scaler)/Z-score

* Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging betwn 0 to 1. It is also known as Min-Max scaling.

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Let 1, 2, 3, Let x=1

$\frac{1 - 1}{3 - 1}$

$= \frac{0}{2} = 0$

X = min^m value in column, numerator = 0
so X = 0

X = max, in column, numerator = denominator
X = 1

X betwn min $ max, value of X', betwn 0-1

Standardization :-

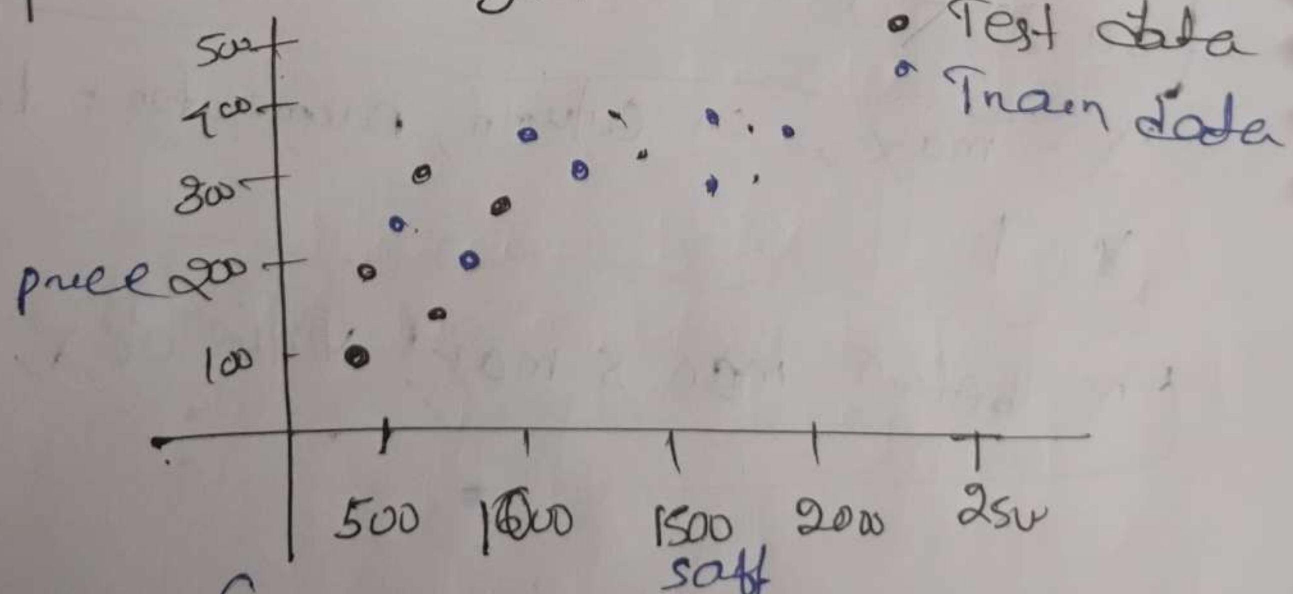
→ Another scaling technique where values are centered around mean with a unit standard deviation. This means that mean of attribute becomes zero and resultant distribution has a unit standard deviation.

$$X' = \frac{X - \mu}{\sigma}$$

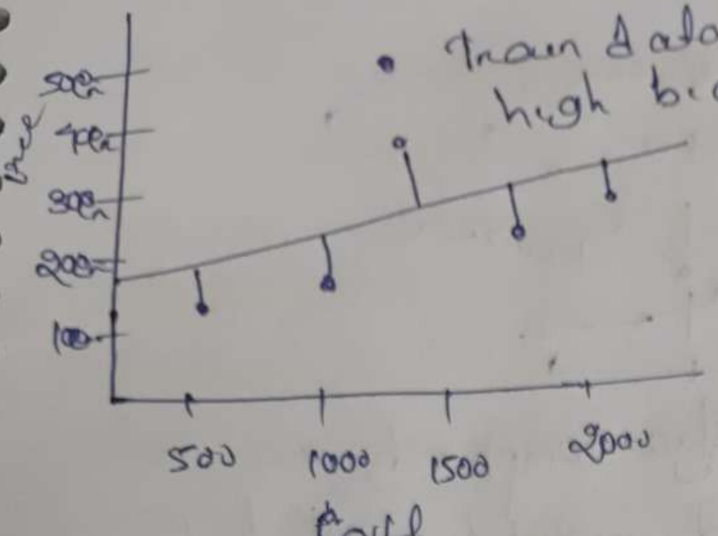
μ → mean of feature value

σ → standard deviation of feature value

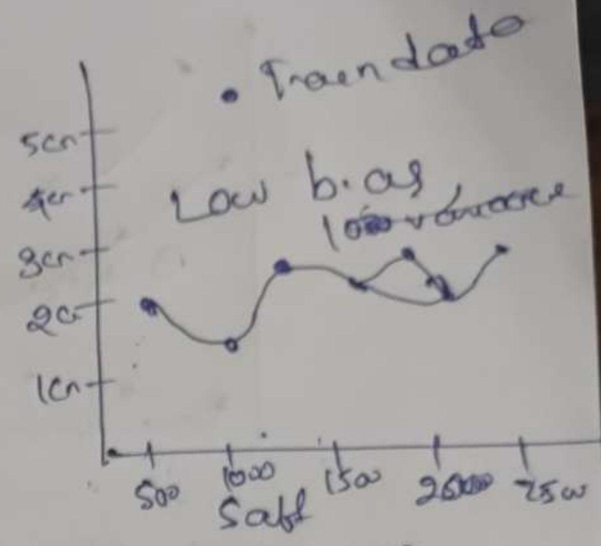
Note :- In this case, values not restricted to particular range.



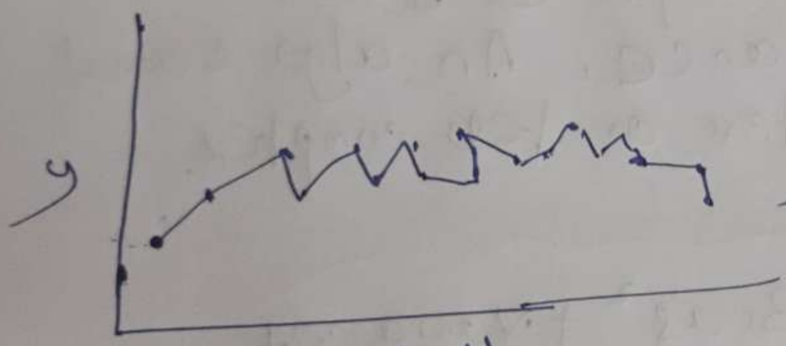
(Bias - Variance Tradeoff)



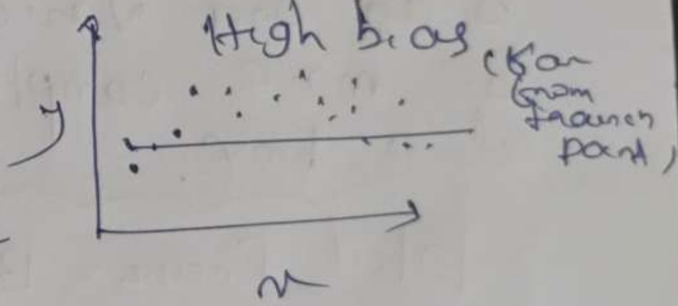
(1) underfitting

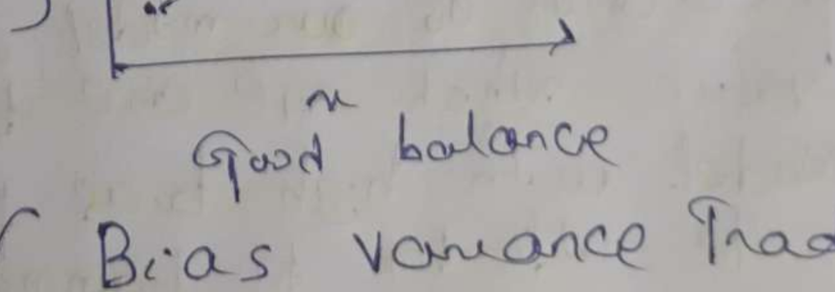


(2) Overfitting



overfitting

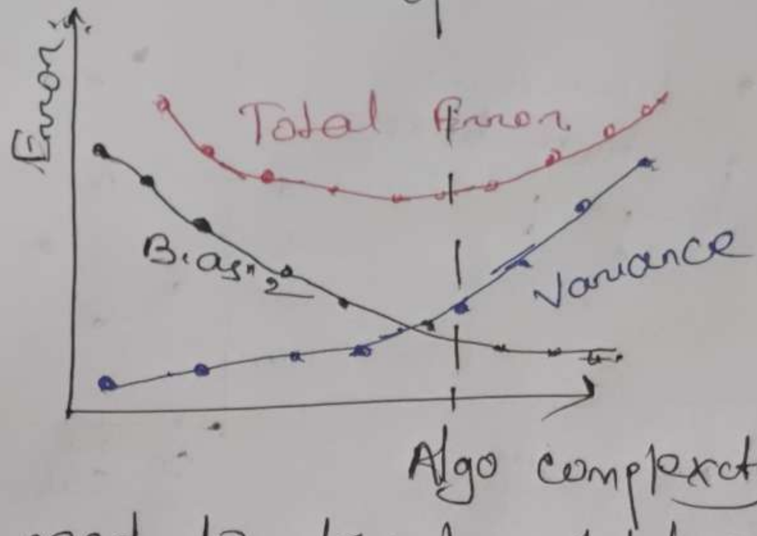




Good balance

(Bias variance Tradeoff)

Optimal Balance



We need to find right / good balance without overfitting and underfitting.

→ Why trade off complexity, betwn bias and variance. An algo cannot be more complex or less complex at a time.

Total Error = Bias ^{2} + Variance + Irreducible Error

→ Bias is difference betwn the avg prediction of our model and the correct value which we are trying to predict. Model with high bias pays very little attention to training data and oversimplifies model, always leads to high error on training and test data.

* Variance is the variability of model prediction for given data point or value which tells us spread of our data. Model with high variance pays a lot of attention to training data and does not generalize on the data which hasnot seen before. Such model perform very well on training data but has high error rates on test data.

Note

Normalization use → does not follow Gaussian distribution. (like k-Nearest Neural Net)

Standardization → on Gaussian distribution