20/3/24 42.

linear, Lasso & Ridge Regression

09 MAY'23 TUESDAY
19th Week • 129-236

Model Building → linear regression → $y = mx + c$

1. calculation
2. loss func / cost func
3. optimization (Gradient descent)

model evaluation → R^2 / adjusted R^2.

Loss func ⟹
1. MSE $= \frac{1}{N} \sum_{i=1}^{N} (y - \hat{y})^2$
2. MAE $= \frac{1}{N} \sum_{i=1}^{N} (y - \hat{y})$
3. RMSE $= \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y - \hat{y})^2}$

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$

SS_{res} = sum of square of residual or error

$$= \frac{SS_{total} - SS_{res}}{SS_{total}}$$

$$\boxed{R^2 = \frac{\sum_{i=1}^{N} (y - \bar{y})^2 - \sum_{i=1}^{N} (y - \hat{y})^2}{\sum_{i=1}^{N} (y - \bar{y})^2}}$$

$\bar{y}$ = average value

$\hat{y}$ = predicted value

$$\boxed{\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(N-1)}{N-P-1}}$$

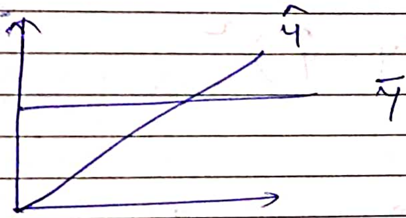
N = no. of data point

p = no. of feature

Variant of linear Reg

1. Ridge reg (L_2)
2. lasso reg (L_1)
3. elastic reg ($L_1 + L_2$)

If R^2 comes betn (0 – 1) or
0 → very bad model. 1 → very good model
0.5 – average model



If $\bar{y}$ intersects $\hat{y}$ we will have a good model.

If $\bar{y} = \hat{y}$ = parallel then worst model. $R^2 = 0$

Assumptions

1. There should be linearity in data. (relationship betn independent and dependent features in the data should be linear.
2. Independence (no dependency within features)
3. normality (Features should be normally distributed)
4. no multi colinearity (it occurs when independent features are highly correlated with each other)

(5) no autocorrelation

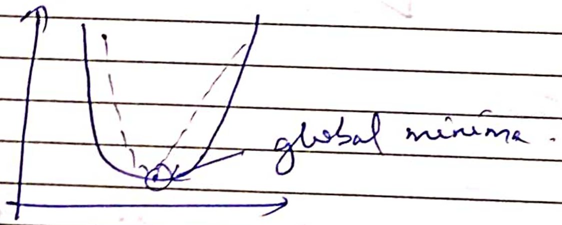
(6) homoscedasticity (the variance of the residual is constant across all levels of the independent variable).

→ If your data is going to satisfy this particular assumption then in that case we can use linear regression.

MSE:

$$\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$$

Eg:- x^2



adv

(i) it is differentiable.

(ii) It has one local and global minima.

disadv

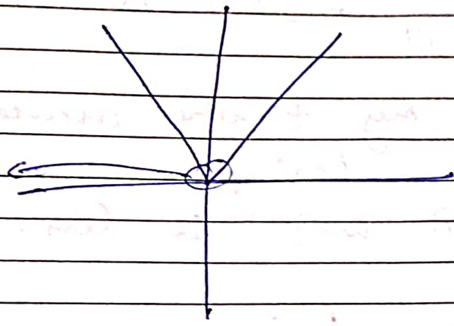
(i) no robust for outlier

(ii) mse change the unit also.

MAE

$$\frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}|$$

Eg:- $|x|$



global minima

$y = mx + c$

$\Rightarrow \frac{d}{dm}[y-(mx+c)]$

$\Rightarrow \frac{dy}{dm} - x\frac{dm}{dm} + \frac{dc}{dm}$

$\Rightarrow 0 - x \times 1 + 0$

$\Rightarrow -x$

we will always have a constant value.

adv

1. Robust to outliers.
2. it will be in same unit

dis

1. Time complexity will be high. we won't be able to reach global minima.

Ridge Reg or L_2 reg.

→ used for reducing the overfitting.

dataset

validation

training data — train our model

test data — test our model — performance of the model.

09.00

If training accuracy = 95%

10.00

& testing accuracy = 55%.

11.00

Then my model is overfit.

12.00

→ Overfitting :- my training accuracy will be high and my testing accuracy won't be high or it will be less.

13.00

14.00

15.00

→ Underfitting :- training accuracy is low & testing accuracy is also low.

16.00

17.00

→ for reducing this overfitting ⇒ we use Regularization

18.00

$$\boxed{\text{cost func} \Rightarrow \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 + \lambda \sum_{i=1}^{n}(m^2)}$$

ridge regularization

ATTO

const value (hyperparameter)

Slope

$+ 1 \times [m_1^2 + m_2^2 + m_3^2] \Rightarrow$ +ve value.

m value will be near to 0.

Lasso regression (L_1 regularizatn)

cost fun / loss = Feature selectn also.

$$\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 + \lambda\sum_{i=1}^{n}|slope|$$

Here slope value will be 0.

Elasticnet

cost fun + L_1 + L_2

$$\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 + \lambda_1\sum_{i=1}^{n}(m)^2 + \lambda_2\sum_{i=1}^{n}|m|$$

→ Reduce overfitting and feature selection.