Lecture 4

Terminology of Linear Regression.

Outliers:- It is an extreme value. It is a problem bcz many times it hampers the result we get.

Multicollinearity:- Independent variable are highly correlated to each other then variable are said to be multicollinear. Many regression tc assume multicollinearity should not be present in dataset. Bcz it causes problem in ranking variable based on importance or it makes job difficult in selecting most imp. independent variable.

Correlation Coefficient

→ Correlation show strength of relationship between 2 variable and is expressed numerically by correlation coefficient.

→ Degree of association is measured by correlation coefficient denoted by r, sometimes called Pearson's correlation coefficient

→ Correlation coef. is measured on scale varies from +1 through 0 to -1.

→ Complete correlation b/w two variables is expressed by either +1 or -1

→ When one variable increases as the other increases the correlation is +ve, when one decreases as other increases it is -ve.

→ Complete absence of correlation is represented by 0.

Pearson's correlation Coefficient (r)

① $$\gamma = \frac{n \times \Sigma x \Sigma y - \Sigma x \Sigma y}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \times \sqrt{n \Sigma y^2 - (\Sigma y)^2}}$$

n = no. of values / data

ex

x	y	x^2	y^2
2	5	4	25
3	4	9	16
4	6	16	36
		29	77

$\sum x = 9$ $\quad$ $n = 4$

$\sum y = 15$

$\sum x \sum y = 135$

$$r = \frac{3 \times 9 \times 15 - 9 \times 15}{\sqrt{3 \times 29 \times 81} \times \sqrt{3 \times 77 \times 225}} = \frac{405 - 135}{83.9 \times 227.9} \Rightarrow$$

② $$r = \frac{Cov(x, y)}{\sqrt{Var(x) \times Var(y)}}$$

Cov = Covariance
Var = Variance

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 (y_i - \bar{y})^2$$

$$Var(x) = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$Var(y) = \frac{1}{n} \sum (y_i - \bar{y})^2$$

$$\boxed{② \; r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}}$$

Page No.:
Date:
YOUVA

d: ⓐ and ⓓ ans.
ⓐ Strong +ve correlation b/w X and Y
ⓓ Weake -ve ——— X and Z.

Regularization

It is technique used to overcome the problem of overfitting and feature selection.
↳ underfitting

Type of Regularization:-
① Lasso Regularization (L_1 Regularization)
② Ridge " (L_2 ")
③ Elastic Net " (L_1 and L_2 ")

① Lasso Regularization (L1 Regularization)
LASSO → Least Absolute Shrinkage and Selection operator.

$$\text{Cost function} = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}-y_i)^2 + \lambda\sum_{i=1}^{m}|\text{slope}_i|$$
a_i

L1 used for feature selection
n = no. of test dat
m = ——— feature (Independent variable)
$\hat{y}$ = Predicted target value, y_i = Actual target value
λ = hyperparameter.

② Ridge Regularization (L2 Regularization)
In this we add squared magnitude of the coefficient in the cost function
Reduce overfitting

Lecture 4

Terminology of Linear Regression.

Outliers :- It is an extreme value. It is a problem bcz Many times it hampers the result we get.

Multicollinearity :- Independent variable are highly correlated to each other then variable are said to be multicollinear. Many regression tec assume multicollinearity should not be present in dataset. Bcz it causes problem in ranking variable based on imp-ortance or it makes job difficult in selecting most imp. independent variable.

Correlation Coefficient

→ Correlation show strength of relationship between 2 variable and is expressed numerically by correlation coefficient.

→ Degree of association is measured by correlation coefficient denoted by r, sometimes called Pearson's correlation coefficient

→ Correlation coef. is measured on scale varies from +1 through 0 to -1.

→ Complete correlation b/w two variables is expressed by either +1 or -1

→ when one variable increases as the other increases the correlation is +ve (when one decreases as other increases it is -ve.

→ Complete absence of correlation is represented by 0.

Pearson's correlation Coefficient (r)

① $$r = \frac{n \times \Sigma x \Sigma y - \Sigma x \Sigma y}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \times \sqrt{n \Sigma y^2 - (\Sigma y)^2}}$$

n = no. of values / data

Date:

ex

x	y	x^2	y^2
2	5	4	25
3	4	9	16
4	6	16	36
		29	77

$\Sigma x = 9$

$\Sigma y = 15$

$\Sigma x \Sigma y = 135$

$n = 4$

$$r = \frac{3 \times 9 \times 15 - 9 \times 15}{\sqrt{3 \times 29 \times 81} \times \sqrt{3 \times \cancel{77} \times 225}} = \frac{405 - 135}{83.9 \times 22.79} \Rightarrow$$

② $r = \dfrac{Cov(x,y)}{\sqrt{Var(x) \times Var(y)}}$ Cov = Covariance
Var = Variance

$$Cov(x,y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 (y_i - \bar{y})^2$$

$$Var(x) = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

$$var(y) = \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2}$$

$$\boxed{② \; r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}}$$

d: (a) and (d) ans.
(a) strong +ve correlation b/w X and Y
(d) weake -ve ——— X and Z.

Regularization

It is technique used to overcome the problem of overfitting and feature selection.
↳ underfiting

Type of Regularization:-
1. Lasso Regularization (L_1 Regularization)
2. Ridge " (L_2 ")
3. Elastic Net " (L1 and L2 ")

(1) Lasso Regularization (L1 Regularization)
LASSO → Least Absolute Shrinkage and Selection operator.

$$\text{Cost function} = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}-y_i)^2 + \lambda\sum_{i=1}^{m}|\text{slope}_i|$$

L1 used for feature selection
n = no. of test dat
m = ——— feature (Independent variable)
$\hat{y}$ = Predicted target value, y_i = Actual target value
λ = hyperparameter.

(2) Ridge ~~Regression~~ Regularization (L2 Regularization)
In this we add squared magnitude of the cofficient in the cost function.
Reduce overfitting

Date:

Cost function $= \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}-y_i)^2 + \lambda\sum_{i=1}^{m}(a_i)^2$

③ Elastic Net Regularization (L1 and L2 regularisation)

Cost function $= \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}-y_i)^2 + \underbrace{\lambda\sum|a_i|}_{\text{feature selection.}} + \underbrace{\lambda\sum|a_i|^2}_{\text{overfitting}}$

Logistic Regression (Classification Algo)

→ It is classification alg. used to assign observation to a discrete set of classes.

→ Predictive analysis algorithm and based on concept of probability.

→ transforms its output using the logistic sigmoid function to return a probability value.

Types:-

① Binary ② Multilinear function

Hypothesis of this tends it to limit the cost function b/w 0 and 1.

⟶ Sigmoid function:-

- In order to map predicted values to probabilities, we use sigmoid function. function help map any real value into another value b/w 0 & 1

$$f(x) = \frac{1}{1+e^{-(x)}}$$

Page No.:

Date:

YOUVA

$$h_\theta(x) = \theta_0 + \theta_1(x)$$

$$= \frac{1}{1+e^{-h_\theta(x)}}$$

y = dependent Variable

x = Independepent Varilable

θ_0 = Intercept

θ_1 = Slope.

$$\boxed{y = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$$