



A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering Data Science



Semester: VI Subject: CSC601 Data Analytics and Visualization Academic Year: 2023-2024
Multiple Linear Regression
Multiple Linear Regression (MLR) also known as simple multiple regression. It is a statistical technique that uses several explanatory variables to predict the outcomes of a response variable.
The aim of multiple linear regression is to model the linear relationship between the independent variables.
The multiple linear regression is the extension of Ordinary Least-Squares (OLS) regression because it involves more than one explanatory variable. MLR is used extensively in econometrics and financial inference.
Multiple Linear Regression Model
It is observed in agriculture that, the crop yield (Y) not only depends on the amount of rainfull (X1) but also on the amount of fertilizer (X2) applied, pesticides (X2) used and of seeds (X4) and little of soil (X2)

Subject Incharge :Prof. Anagha Aher

Page No._

1



A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
Data Science



Semester: VI Subject: CSC601 Data Analytics and Visualization Academic Year: 2023- 2024
Thus in multiple regression, the dependent variable y is a function of more than one independent variables, i.e.
Y= f(X1, X2,, Xn) In multiple pantinear regression, fts non-linear f is linear
Y = β ₀ + β ₁ X ₁ + β ₂ X ₂ + β _K X _K Suppose Y depends on two independent variables X ₁ and X ₂ .
$Y = \beta_0 + \beta_1 \times_1 + \beta_2 \times_2$
To estimate the coefficients Bo, B1, B2 we apply the least square method to minimise N & Yi - (bo + b1 X1 + b2 X2i) }
Subject Incharge :Prof. Anagha Aher Page No Department of CSE-Data Science APSIT





A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering Data Science



Semester: VI Subject: CSC601 Data Analytics and Visualization

Academic Year: 2023- 2024

This results in three normal equations
This regults in three normal equations given by N N N N N N N N N N N N N
$\frac{N}{\sum_{i=1}^{N} x_{ii} \cdot y_{i}^{i} = b_{0} \leq x_{ii} + b_{1} \leq x_{i1}^{2} + b_{2} \leq x_{ii} \cdot x_{2i}^{i}}$
Z X2; Y: = bo Z X2; + b1 Z X1; X2; + b2 Z X2;
Here bo, b1, b2 are the least squares estimates of \$1,\$2 Bo, \$1,\$2
Linear Multiple linear Regression in k-independent
The above analysis can be generalised to fit $N(k+1)$ tuples (XII, X2I,, XKI) (i=1+0N), to the tuples equation.
$y = \beta_0 + b_1 \times 1 + \beta_1 \times 2 + \dots + \beta_K \times K$

Subject Incharge : Prof. Anagha Aher

0

Page No._



A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering **Data Science**



Semester: VI Subject: CSC601 Data Analytics and Visualization Academic Year: 2023-2024 The (k+1) normal equations are:

Subject Incharge :Prof. Anagha Aher Page No.____



A.P. SHAH INSTITUTE OF TECHNOLOGY

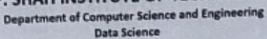
Department of Computer Science and Engineering
Data Science



Academic Year: 2023- 2024 Semester: VI Subject: CSC601 Data Analytics and Visualization Example 1 (multiple linear regression) [Dec 23] [lom] Fit a regression equation to estimate \$6,\$1.\$2 to the following data of a transport company on the weights of 6 shipments, the distances they were moved and the damage of the goods that was incurred. Estimate the damage when a shipment of 3700 kg is moved to a distance of 260km Weight X1 3.0 1.6 1.2 (1000 kg) Distance X2 0.8 1.6 Clookm Damage Y 160 90 Solution :-Let weight X1 and distance X2 be independent variables and the damage y be the dependent variable.

Subject Incharge : Prof. Anagha Aher

A.P. SHAH INSTITUTE OF TECHNOLOGY





Academic Year: 2023-2024

Semester: VI Subject: CSC601 Data Analytics and Visualization Academic Year: 2023-
Let the equation of regression be.
y= b0+b1 x1+b2 x2
where bo, b1, b2 are estimates of Bo, B1, Bo
The three normal equations become.
Ey: = nbo + b E x + b 2 E x 2 i
EXIIYi = bo EXII + b1 EXII + b2 EXII
5 X2i Yi = bo 2 X2i + b1 2 X1i X2i + b2 2)
lect Incharge : Prof. Anagha Aher Page No Department of CSE-Data Science APS

Page No.

Subject Incharge : Prof. Anagha Aher





A.P. SHAH INSTITUTE OF TECHNOLOGY



Semester: VI Subject: CSC601 Data Analytics and Visualization

Academic Year: 2023- 2024

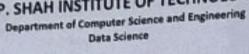
X1 Weight (lookg)	Distance (pokm)	J Damoge in Rs	×ı	×2°	×1.×2	XIY	X2.
4.0 3.0 1.6 1.2 3.4	1.5 2.2 1.0 2.0 0.8	160 112 69 90 123	16 09 2.56 1.44	2.25 4.84 1.0 4.0 0.64	6.6 1.6 2.4 2.72	640 336 110.4 108 418.2	241 246. 69 180 98:
18 2 XII	9.1 2.X2i	186	23.04 63.6 ZXii	AND DESCRIPTION OF	7.68 27 \(\infty\)	254.54	297

Now,	n=6, \(\Sigma \times 18, \Sigma \times 21 = 9.1, \(\frac{7}{24} \);
ALL BANK	5 Yi = 740
	$2 \times 10^{2} = 63.6$
	$\leq x_{2i}^2 = 15.29$
	≤ Xii· X2i= 27
Let Mary	Z X11. Yis = 250.54
	≤ X2i· Yi = 1131.4
	≤ X2i· Yi = 1131.4

Subject Incharge : Prof. Anagha Aher Page No._



A.P. SHAH INSTITUTE OF TECHNOLOGY





Semester: VI Subject: CSC601 Data Analytics and Visualization Academic Year: 2023- 2024
Semester, Vi Subject, Cacad Data Many 100
Normal equations become
740 = 6b0 + 18b1 + 9.1b2
250.54 = 1860 + 63.6 b1 + 2762
-1131.4 = 9.1 bo + 2751 + 15.29 b2
Solving these equations we get
50 = 14.56 $51 = 30.109$
b2 = 12.16
Thus the required regression equation is
$y = 14.56 + 30.109(x_1) + 12.16(x_2)$
For a weight of 3700 kg. (XI = 3.7) and for a distance of 260 km (X2 = 2.6) the damage incurred in rupees is
$y(x_1=3.7, x_2=2.6)=14.56+30.109(3.7)+12.16(2.6)$ = 714.58
= 715 R5. Subject Incharge : Prof. Anagha Aher Page No. Department of CSE-Data Science APSIT
Subject Incharge :Prof. Anagha Aher Page No. Department of CSE-Data Science APSIT