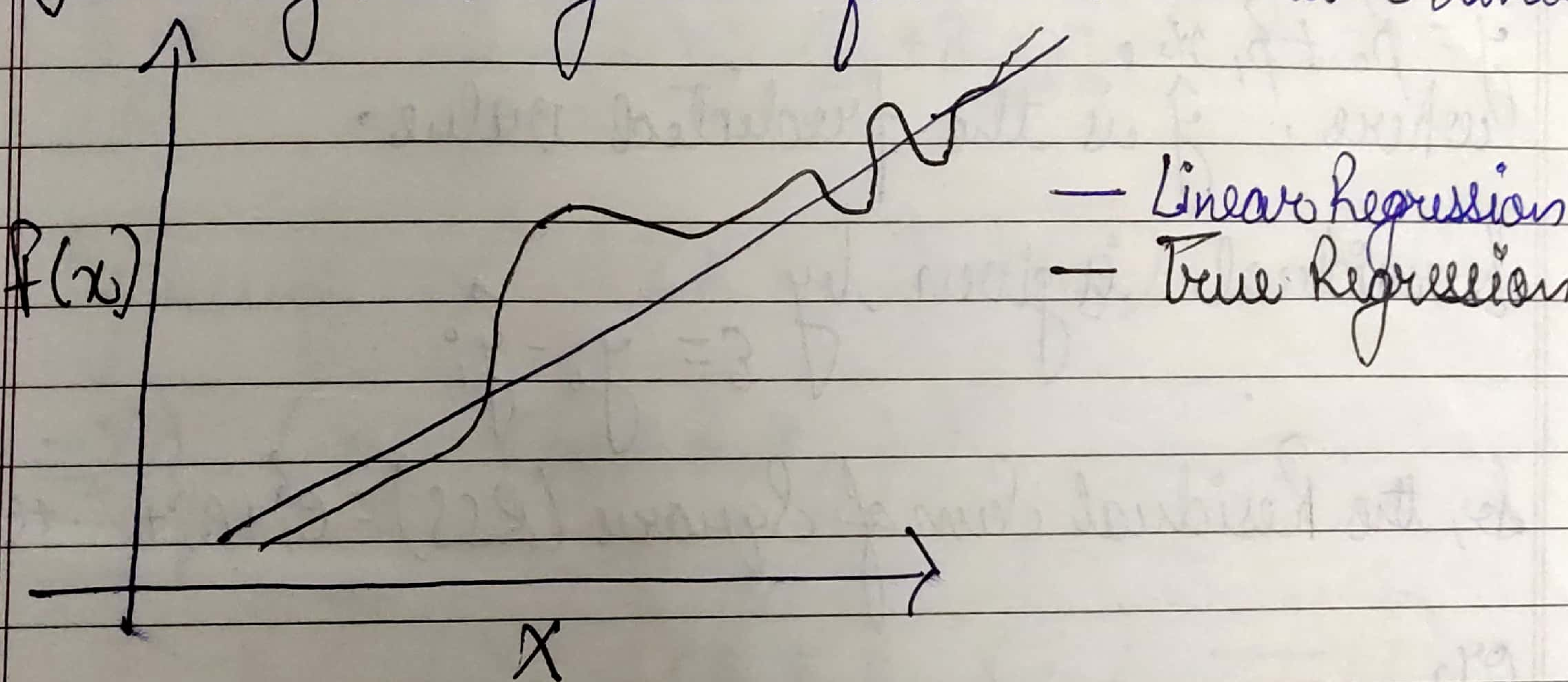
Date :
Page :

Ch-3. Linear Regression

Linear regression is a simple supervised learning approach that work on an assumption that "Y" has a linear dependence on the predictors, $x_1; x_2, x_3 \cdots x_p$.

In reality true regression functions are never linear.



Although, linear regression may sound overly simplistic, it is practically very useful.
And as my opinion "Simple is good"

We will consider the example of Linear Regression on Advertising Data. (Photo Below)

After looking at the photo, one might ask:
→ Is there a relation between budget & sales.
→ How strong is that relation
→ Which media contribute to sales
→ How can we predict future Sales
→ And what can we say about the Synergy in Advertising media?

Simple Linear Regression: using a single predictor

The Model is given by:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where β_0 & β_1 are Intercept & Slope respectively & ε is "irreducible error"

→ We can predict the future sales using

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

where, $\hat{y}$ is the predicted value.

& residual is given by

$$\varepsilon = y_i - \hat{y}_i$$

So, the Residual Sum of Squares (RSS) $= e_1^2 + e_2^2 + \cdots + e_n^2$

or

$$RSS = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \cdots$$

→ The Least RSS approach ~~chooses~~ chooses $\hat{\beta}_0$ & $\hat{\beta}_1$ to minimize RSS

This can be shown by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $\bar{x}$ & $\bar{y}$ are Sample Means.

** OPTIONAL

Derivation for $\hat{\beta}_0$ & $\hat{\beta}_1$

$$RSS = \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$$

To minimize RSS we calculate partial derivatives w.r.t β_0 & β_1.

Part 1.

$$\frac{\partial}{\partial \beta_0} RSS = \frac{\partial}{\partial \beta_0} \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$$

By using $\rightarrow \partial \sum X = \sum \partial (x)$

$$= \sum \frac{\partial}{\partial \beta_0} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$$

$$= -2 \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) \quad \text{— (1)}$$

Similarily,

$$\frac{\partial}{\beta_1} RSS = -2 \sum (X_i (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i)) \quad \text{— (2)}$$

Set, (1) & (2) $= 0$

this gives $-2 \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) = 0$ → Solve this for $\hat{\beta}_0$

$$-2 \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) X_i = 0$$

↑ Substitute $\hat{\beta}_0$ here & Solve for $\hat{\beta}_1$

Solving for $\hat{\beta}_0$

$$\sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) = 0$$

$$\sum Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum X_i = 0$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_{01} \bar{X}$$

Solvarving for $\hat{\beta}_1$

$$\sum X_i (Y_i - (\hat{\beta}_{0} + \hat{\beta}_1 X_i)) = 0$$

Substituting $\hat{\beta}_0$

$$\sum X_i (Y_i - Y - \hat{\beta}_1 (X_i - \bar{X})) = 0$$

So, $\hat{\beta}_1 = \dfrac{\sum X_i (Y_i - \bar{Y})}{\sum X_i (X_i - \bar{X})}$

this can be written as

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Proof $\Rightarrow \sum (X_i - \bar{X})(Y_i - \bar{Y}) =$

$$= \sum X_i (Y_i - \bar{Y}) + \bar{X} \sum Y_i - \bar{Y})$$

$$\sum Y_i - \bar{Y} = 0$$

as $\sum Y_i - n\bar{Y} = 0$

So, $\hat{\beta}_1 = \dfrac{\sum (X_i - \bar{X})(Y_j - \bar{Y})}{\sum (X_i - \bar{X})^2}$

Date :
Page :

To test if ESTIMATED Coefficient Estimates works for our Data

→ The Standard Error of an estimator reflects how it varies under Repeated Sampling.

$$SE(\hat{\beta}_1)^2 = \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}$$

$$SE(\hat{\beta}_0)^2 = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2}\right]$$

$\sigma^2 = Var(\epsilon)$

Derivation:

Properties: $E(a+bX) = a + b\,E(X)$

$Var(a+bX) = b^2\,Var(X)$

We know,

$$\hat{\beta}_1 = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = \Sigma(x_i-\bar{x})y_i$$

$= \Sigma(x_i)(x_i-\bar{x})$

Lets start by calculating mean of $\hat{\beta}_1$.

$$E(\hat{\beta}_1) = E\left(\frac{\Sigma(x_i-\bar{x})y_i}{\Sigma(x_i-\bar{x})^2}\right)$$

$$= \frac{1}{\Sigma(x_i-\bar{x})^2} E(\Sigma(x_i-\bar{x})y_i) = \frac{1}{\Sigma(x_i-\bar{x})^2}\Sigma(x_i-\bar{x})E(y_i)$$

Substitute $Y_i = \beta_0 + \beta_1 x_i + \varepsilon$ $\underbrace{\text{This will be fixed}}$

$$E(\hat{\beta}_1) = \frac{1}{\Sigma(x_i - \bar{x})^2} \Sigma(x_i - \bar{x}) E(\beta_0 + \beta_1 x_i + \varepsilon)$$

$\& \; E(\varepsilon) = 0$

So,

$$E(\hat{\beta}_1) = \frac{1}{\Sigma(x_i - \bar{x})^2} \Sigma(x_i - \bar{x}) (\beta_0 + \beta_1 x_i)$$

$$= \frac{1}{\Sigma(x_i - \bar{x})^2} \beta_0 \underbrace{\Sigma(x_i - \bar{x})}_{=0} + \beta_1 \Sigma (x_i - \bar{x}) x_i$$

$$= \frac{1}{\cancel{\Sigma(x_i - \bar{x})^2}} \times \beta_1 \Sigma \cancel{(x_i - \bar{x}) x_i}$$

$= \beta_1$ that means β_1 is the best point estimate.

$$Var(\hat{\beta}_1) = Var\left(\frac{\Sigma(x_i - \bar{x}) Y_i}{(x_i - \bar{x})^2}\right)$$

$$= \frac{1}{\Sigma(x_i - \bar{x})^4} Var(\Sigma(x_i - \bar{x}) Y_i)$$

$$= \frac{1}{\Sigma(x_i - \bar{x})^4} Var(\Sigma(x_i - \bar{x})(\beta_0 + \beta_1 X_i + \varepsilon))$$

$$= \frac{1}{\Sigma(x_i - \bar{x})^4} Var(\Sigma(x_i - \bar{x}) \varepsilon)$$

$$= \frac{1}{\sum (x_i - \bar{x})^{4th}} \sum (x_i - \bar{x})^2 \, Var(\varepsilon)$$

$$= \frac{\sigma}{\sum (x_i - \bar{x})^2}$$ ~~vip~~

So now we can say, that there is 95% chances that β_1 will be in the range of

$\hat{\beta}_1 \pm 2 \cdot SE(\hat{\beta}_1)$

HYPOTHESIS TESTING:

Now after calculation standard error we can check if Y & X have any relation or not.

To do so, we will create a Null Hypothesis & Alternate Hypothesis

H_0: There is no relation b/w X & Y ; $\beta_1 = 0$

H_A: There is relation between X & Y ; $\beta_1 \neq 0$

To test hypothesis we will use t-statistic

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$$ where degree of freedom $= n-2$

& now we can calculate p-value for it

R-Squared Statistic

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

This explain what fraction of Y can be explained by X.

$$TSS = \Sigma (y_i - \bar{y})^2$$

It can be shown in Simple Linear Regression that $R^2 = r^2$, where r is the correlation between X & Y.

$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

Multiple Linear Regression:

In Simple Regression we studied our sales wrt TV.

but

In Multiple Regression we study the effect of synergy of TV, Radio & Newspaper on Sales.

So,

$$Sales = \beta_0 + \beta_1 \times TV + \beta_2 \times Radio + \beta_3 \times Newspaper + \epsilon$$

Date :
Page :

In Simple Regression we had a line
in
Multiple Regression we have a hyperplane

Interpretting Regression Coefficients:

→ The ideal scenario is when the predictors are uncorrelated.

→ Correlation amongst predictors cause problems.
- The variance of all the coefficients tends to increase
- Interpretation becomes hazardous - when X_j changes everything else changes.

→ Claims of Causality should be avoided for ~~obs~~ objectionable data.

OVERVIEW OF MULTIPLE REGRESSION:

A regression coefficient β_j estimates the $E(Y|X_j)$ with all other predictors held fixed,

But that's not true. The predictors usually change together.

If you couldn't understand the last statement, here is an example:

Let, Y = total amount of change in your pockets
Let, X_1 = 5₹ coins
& X_2 = 10₹ Coins

& you have $Y = 30$

So, $\beta_1 X_1 + \beta_2 X_2 = 30$

$\beta_1 \times 5 + \beta_2 \times 10 = 30$

Let, $\beta_1 = 2$ & $\beta_2 = 2$

Now, let $Y = 35$

& try keeping β_1 constant

So, $2 \times 5 + \beta_2 \times 10 = 35$

So, $\beta_2 = 2.5$ which is not possible.

Now let Y = # of tackles by a football player in a season based on

X_1 = Weight & X_2 = Height

Fitted Model is $\Rightarrow \hat{Y} = b_0 + 0.5 X_1 - 0.1 X_2$

How do we Interpret $\beta_2 < 0$

66 "Essentially all Models are wrong, but some are Useful"

~ George Box

So, it doesn't matter if $\beta_2 < 0$, it doesn't matter what it means until & unless Model works.

ESTIMATION & PREDICTION FOR MULTIPLE REGRESSION.

→ here, $\hat{y}$ is given by: $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \dots + \hat{\beta}_p x_p$

→ we estimate $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ as values that minimize Residual Sum of Squares (RSS)

$$R_S = \sum_{i=1}^{n} (y_i - \hat{y})^2 \quad \text{where } \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p$$

Now let's answer the Questions we ~~asked~~ asked in the starting of this chapter:

Q Is atleast one predictor useful?

To answer this question we can calculate F-statistic

$$F = \left[(SSE_1 - SSE_2)/m\right] / SSE_2/n-K$$

where, K is # of independent variables.

In our case, $F = \dfrac{(TSS - RSS)/p}{RSS/(n-p-1)}$ → This answered the first 3 questions here.

Quantity	Value
RSE	1.69
R^2	0.897 → ✓
F-statistic	570 → p value < 0.001 ✓

Till now we've dealt with only Quantitative Variables. What would we do if we get Qualitative Variables?

If we get Qualitative Variable such as Gender

Gender	Male Female

we process it into

Gender	1 or 0 0 1

So, $\hat{y} = \beta_0 + \beta_1 x + \epsilon$, when $x = 1$

$= \beta_0 + \epsilon$, when $x = 0$

SUPPOSE that spending money on radio advertising actually increases the effectiveness of TV advertising.

In this situation spending half on radio & half on TV does better than allocating all budget to TV or Radio Entirely.

This is called SYNERGY EFFECT.

To deal with such situation we add a Synergy Term.

Sales $= \beta_0 + \beta_1 \times TV + \beta_2 \times Radio + \beta_3 (TV \times Radio) + \epsilon$

Resulting in

Date : Page :

	P-value
Intercept	<0.0001
TV	<0.0001
Radio	0.0014
TV×Radio	<0.0001

This works out pretty well

it can also be written as

$$Sales = \beta_0 + \underbrace{(\beta_1 + \beta_3 \times radio) \times TV}_{} + \beta_2 \times radio + \varepsilon$$

This part here shows that Radio is now a function of TV.

Interpretations from the output:-

(1) Here, p value of TV×Radio Synergy is very low hence indicating strong evidence for

$\boxed{H_A : \beta_3 \neq 0}$

(2) Secondly, R^2 for the model increased from 89.7% to 96.8% because of Synergy Term.

Interaction between Qualitative & Quantitative Variables:

Lets say we want to predict balance using income (quantitative) & student (qualitative) it is quite easy.

$$balance = \beta_0 + \beta_1 \times income + \begin{cases} \beta_2 & \text{student} \\ 0 & \text{not student} \end{cases}$$

$$= \beta_1 \times income + \begin{cases} \beta_2 + \beta_0 \\ \beta_0 \end{cases}$$

with Interaction

$$\text{balance} = \beta_0 + \beta_1 \times \text{income} + \begin{cases} \beta_2 + \beta_3 \times \text{Income} & \text{student} \\ 0 & \text{not student} \end{cases}$$

$$= \begin{cases} (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \times \text{Income} & \text{student} \\ \beta_0 + \beta_1 \times \text{Income} \end{cases}$$