D Y PATIL DEEMED TO BE UNIVERSITY
RAMRAO ADIK INSTITUTE OF TECHNOLOGY
NAVI MUMBAI

Subject Name: MACHINE LEARNING

Unit 3: Learning with Regression

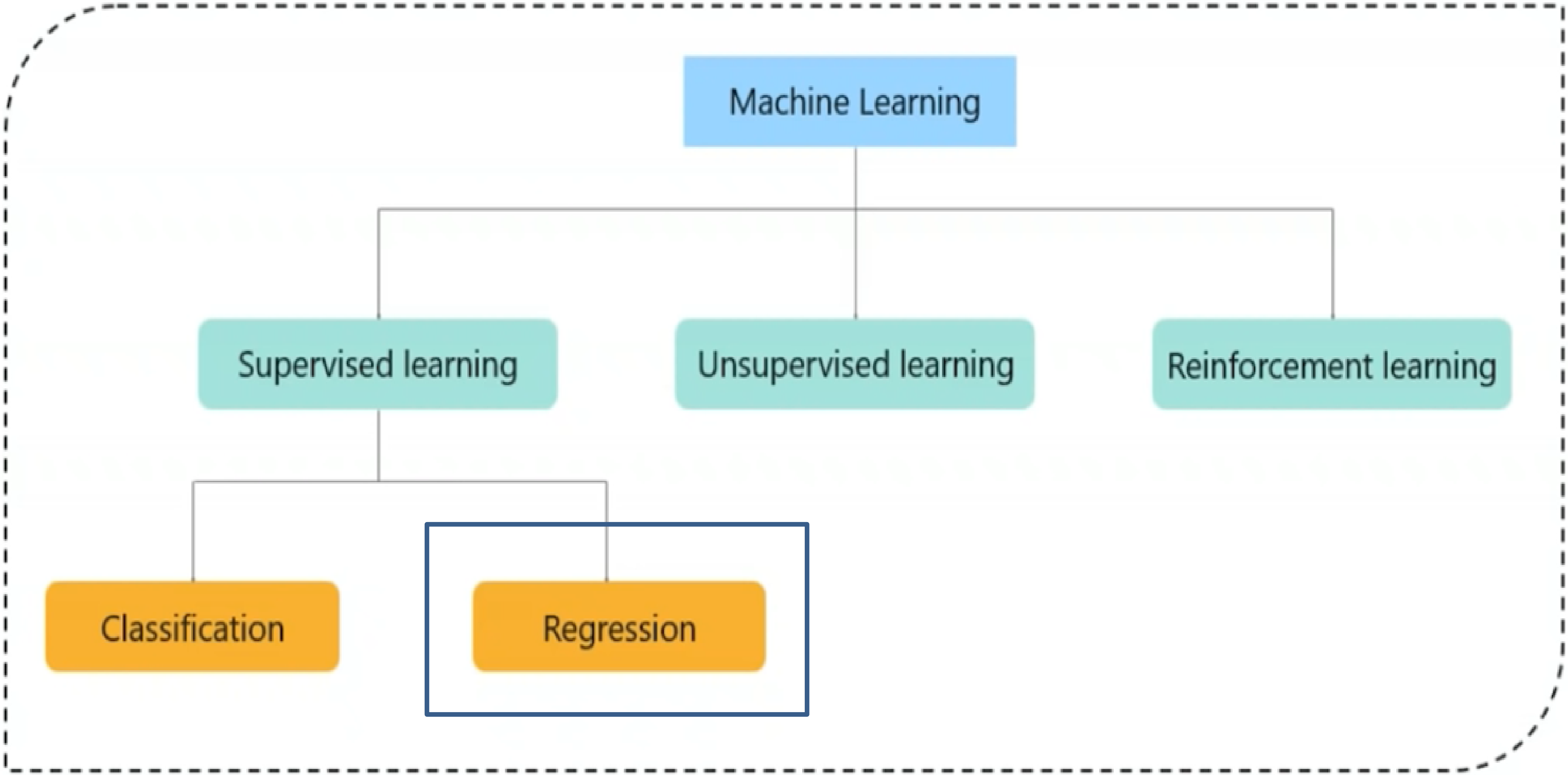
Faculty Name : Ms.Rajashree Shedge

Index

Unit No: 3 Unit Name : Learning with Regression

Lecture No: 13
Linear Regression

Supervised Learning Tasks



What is Regression?

- Regression analysis is a predictive modelling technique that analyzes the relation between the target or dependent variable and independent variable in a dataset.

- Function: a mathematical relationship enabling us to predict what values of one variable (Y) correspond to given values of another variable (X).

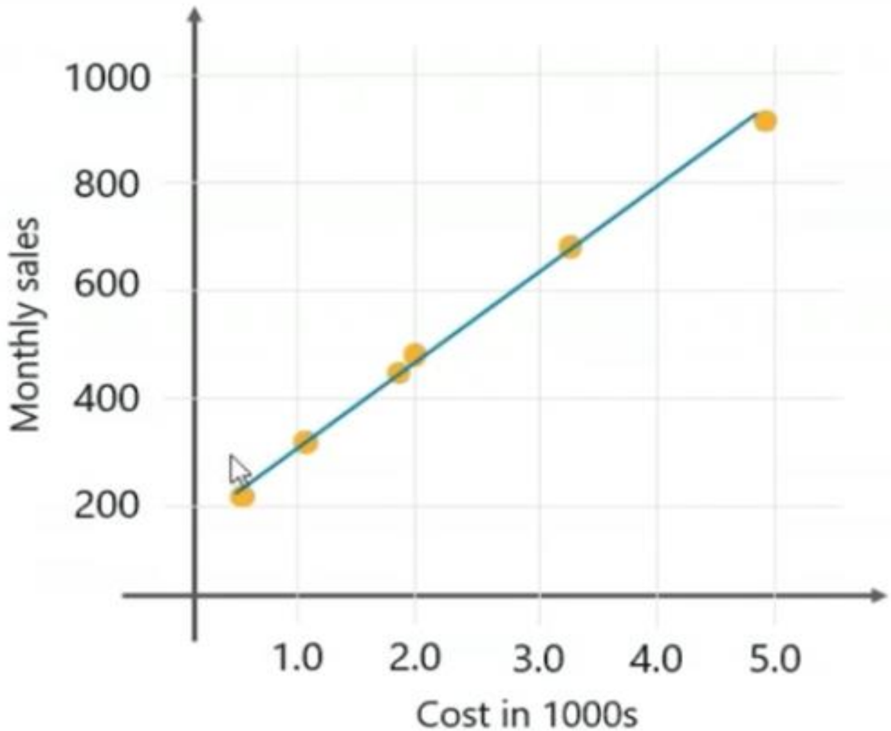
- Y: is referred to as the *dependent variable*, the *response variable* or the *predicted variable*.
- X: is referred to as the *independent variable*, the *explanatory variable* or the *predictor variable*.

- Used Mainly for **Prediction** & **Estimation**

Regression….

To forecast monthly sales by studying the relationship between the monthly e-commerce sales and the online advertising costs.

Monthly sales	Advertising cost In 1000s
200	0.5
900	5
450	1.9
680	3.2
490	2.0
300	1.0



Broad categories of Regression

Regression can be broadly classified into two major types.

- Linear Regression.
  - The simplest case of linear regression is to find a relationship using a linear model (i.e. line) between an input independent variable (input single feature) and an output dependent variable.
  - This is also called Bivariate Linear Regression.

- Multivariate Linear Regression:
  - When there is a linear model representing the relationship between a dependent output and multiple independent input variables is called Multivariate Linear Regression.

Broad categories of Regression….

- Logistic Regression:
  - It is used when the output is categorical. It is more like a classification problem. The output can be Success / Failure, Yes / No, True/ False or 0/1. There is no need for a linear relationship between the dependent output variable and independent input variables.

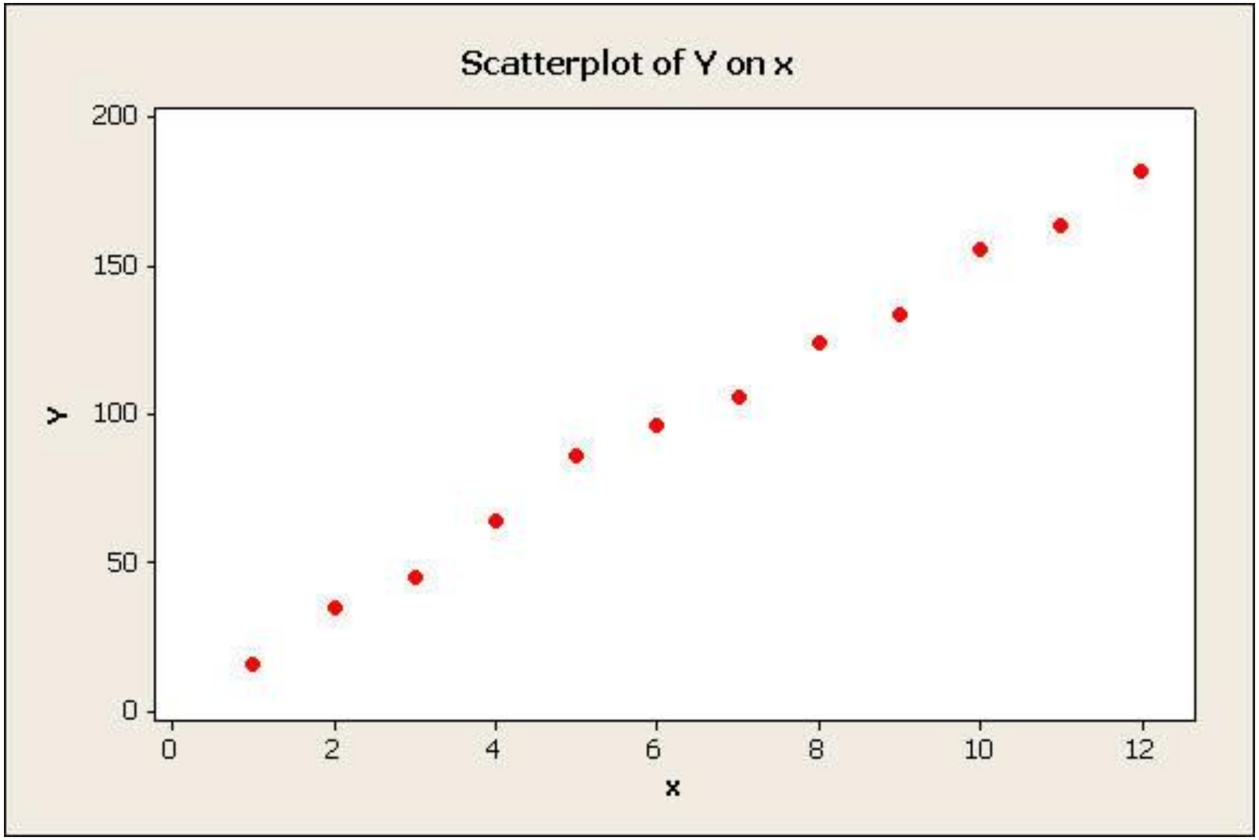
Scatter Plot

- A scatter plot can be a helpful tool in determining the strength of the relationship between two variables.

- A scatter plot is often employed to identify potential associations between two variables, where one may be considered to be an explanatory variable (such as years of education) and another may be considered a response variable (such as annual income).

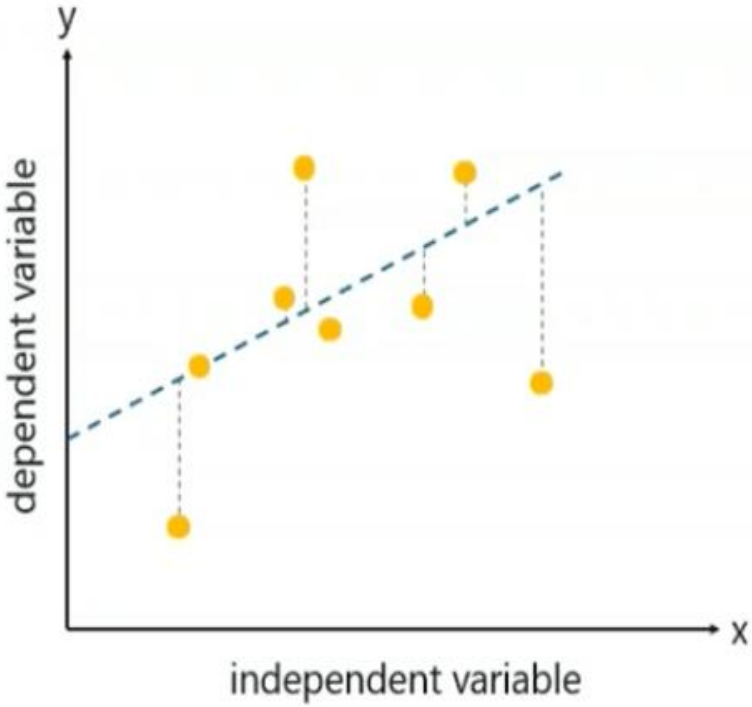
Scatter Plot : Example

x	1	2	3	4	5	6	7	8	9	10	11	12
y	16	35	45	64	86	96	106	124	134	156	164	182



Linear Regression

Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (X). It can be used for the cases where we want to predict some continuous quantity.



- *Dependent variable (Y):*
The response variable who's value needs to be predicted.

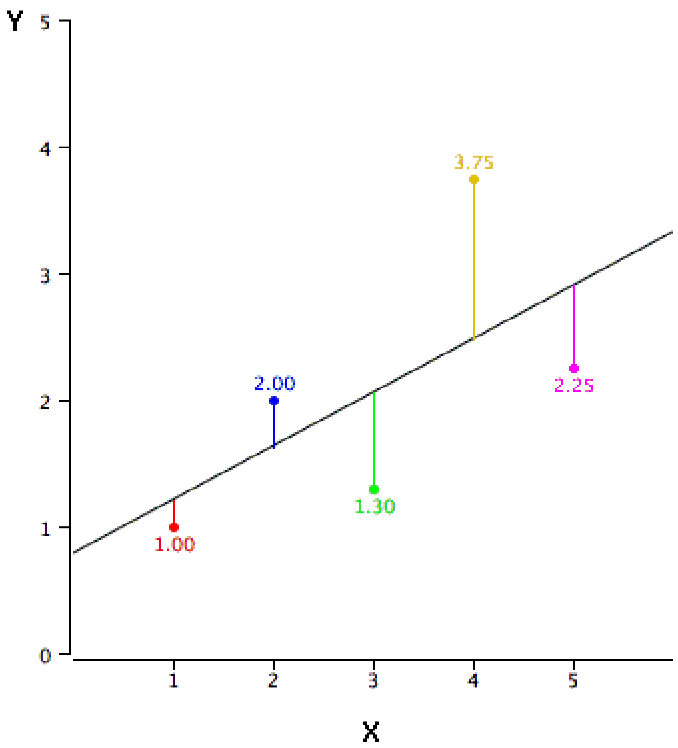
- *Independent variable (X):*
The predictor variable used to predict the response variable.

The following equation is used to represent a linear regression model:

$$Y = b_0 + b_1 x + e$$

Regression Line

- Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a *regression line*.



Error in prediction

- The black diagonal line in Figure is the regression line and consists of the predicted score on Y for each possible value of X. The vertical lines from the points to the regression line represent the **errors of prediction**.

- The error of prediction for a point is the value of the point minus the predicted value.

Error in prediction

Objective: Minimize the difference between the observation and its prediction according to the line.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$\text{for } i = 1, 2, \ldots, n$$

$$\varepsilon_i = y_i - \hat{y}_i$$
$$= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$\hat{y}_i = \text{predicted y value when x} = x_i$$

Method of Least Squares

We want the line which is best for all points. This is done by finding the values of b0 and b1 which minimizes some sum of errors. There are a number of ways of doing this. Consider these two

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} |\varepsilon_i|$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \varepsilon_i^{\,2}$$

$\hat{\beta}_0 \quad \hat{\beta}_1$ referred to as least squares estimates

The method of least squares produces estimates with statistical properties (e.g. sampling distributions) which are easier to determine

Method of Least Squares

'Best Fit' Means Difference Between Actual Y Values & Predicted Y Values Are a Minimum. But Positive Differences Off-Set Negative. So square errors

$$E(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^{\ 2} = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

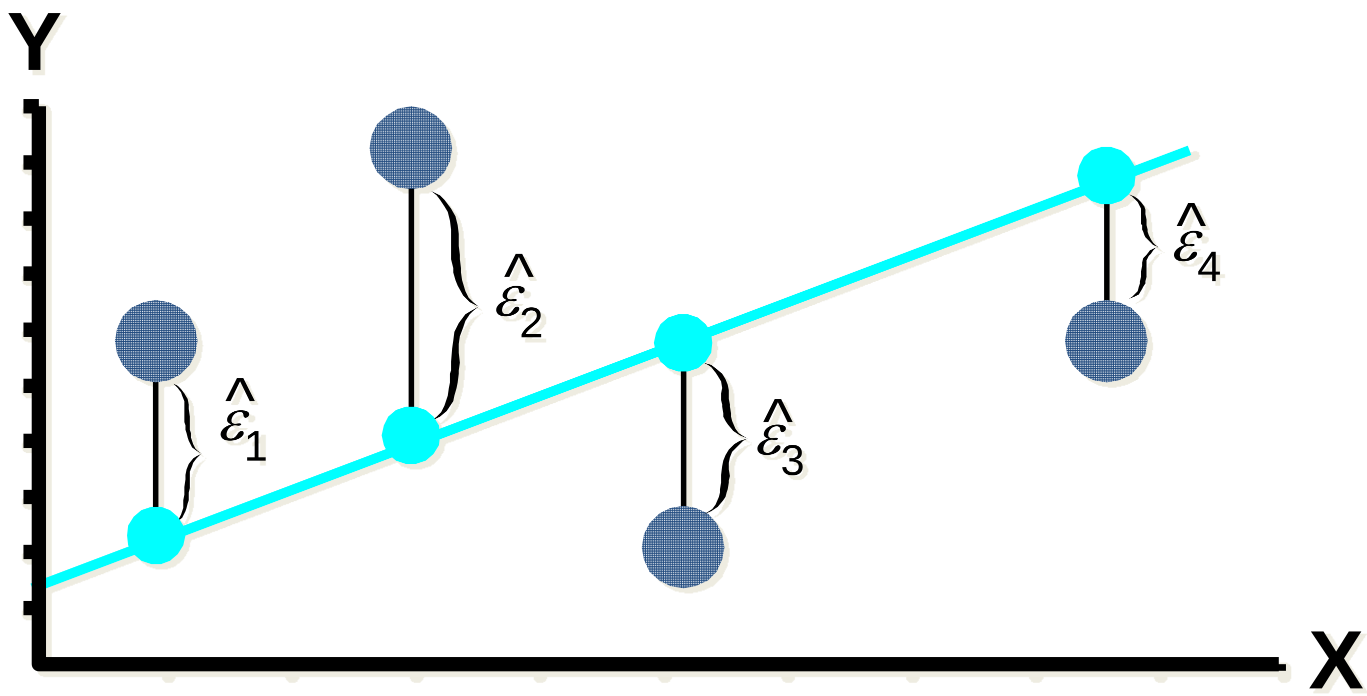
LS Minimizes the Sum of the Squared Differences (errors) (SSE)

$$\frac{\partial E}{\partial \beta_0} = 0$$

$$\frac{\partial E}{\partial \beta_1} = 0$$

Least Square Graphically

$$\textbf{LS minimizes} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$$



Derivation of Parameters

Least Squares (L-S): Minimize squared error

$$
\begin{aligned}
0 &= \frac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = \frac{\partial \sum \left(y_i - \beta_0 - \beta_1 x_i \right)^2}{\partial \beta_1} \\
&= -2 \sum x_i \left(y_i - \beta_0 - \beta_1 x_i \right) \\
&= -2 \sum x_i \left(y_i - \bar{y} + \beta_1 \bar{x} - \beta_1 x_i \right)
\end{aligned}
$$

$$
\begin{aligned}
&\beta_1 \sum x_i \left(x_i - \bar{x} \right) = \sum x_i \left(y_i - \bar{y} \right) \\
&\beta_1 \sum \left(x_i - \bar{x} \right)\left(x_i - \bar{x} \right) = \sum \left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right) \\
&\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}
\end{aligned}
$$

Derivation of Parameters

$$
\begin{aligned}
S_{xx} &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2 \\
&= \sum_{i=1}^{n}(x_i^{\ 2}) - \frac{1}{n}\left(\sum_{i-1}^{n} x_i\right)^2
\end{aligned}
$$

Sums of squares of x.

$$
\begin{aligned}
S_{yy} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\
&= (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2 \\
&= \sum_{i=1}^{n}(y_i^{\ 2}) - \frac{1}{n}\left(\sum_{i-1}^{n} y_i\right)^2
\end{aligned}
$$

Sums of squares of y.

Derivation of Parameters

$$
\begin{aligned}
S_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&= (x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^{n}(x_i y_i) - \frac{1}{n}\left(\sum_{i-1}^{n} x_i\right)\left(\sum_{i-1}^{n} y_i\right)
\end{aligned}
$$

Sums of cross products of x and y.

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Linear Regression : In Simpler Form

- The simple linear model is expressed using the following equation:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$\text{for } i = 1, 2, \ldots, n$$

Where,

y – variable that is dependent
x – Independent (explanatory) variable
– Intercept
β_0 – Slope
β_1 - Residual (error)

- For simplicity in calculations, we assume the error to be 0.

- The simple form of the regression model is a line equation:

$$\boxed{y_i = \beta_0 + \beta_1 x_i}$$

Linear Regression : In Simpler Form

- Regression model : $\boxed{y_i = \beta_0 + \beta_1 x_i}$
- Here β_0 & β_1 are called the regression coefficients.
- To calculate β_0 , u β_1 ollowing formula:

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

- Where, $\bar{x}$ is the mean of x and $\bar{y}$ is the mean of y

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Linear Regression : Example

1. The following data pertain to number of computer jobs per day and the central processing unit (CPU) time required.

Number of jobs x	CPU time y
1	2
2	5
3	4
4	9
5	10

Linear Regression : Example

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
	1	2	-2	-4	8	4	16
	2	5	-1	-1	1	1	1
	3	4	0	-2	0	0	4
	4	9	1	3	3	1	9
	5	10	2	4	8	4	16
Total	15	30	0	0	20	10	46

So $S_{xx} = 10$, $S_{yy} = 46$ and $S_{xy} = 20$. Also, $\bar{x} = 3$ and $\bar{y} = 6$.

$$y = 2x$$

D Y PATIL
DEEMED TO BE
UNIVERSITY
RAMRAO ADIK
INSTITUTE OF TECHNOLOGY
NAVI MUMBAI

Exercise

2. The following table shows the midterm and final exam grades obtained for students in a database course. Use the method of Least squares using regression to predict the final exam grade of a student who received 80 on the midterm exam.

Midterm Exam (X)	Final Exam (Y)
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

Exercise

3. A clinical trail gave the following data about the BMI and Cholesterol level of 10 patients. Predict the likely value of Cholesterol level for a patient who has BMI of 27.

BMI	Cholesterol
17	140
21	189
24	210
28	240
14	130
16	100
19	135
22	166
15	130
18	170

Exercise

4. Find the regression coefficients for the following data:

Age	Glucose Level
43	99
21	65
25	79
42	75
57	87
59	81

Let’s revise through a small video

Introduction to Linear Regression:

https://www.youtube.com/watch?v=zPG4NjIkCjc

Multivariable/Multivariate Regression

- Multivariate regression is a technique used to measure the degree to which the **various independent variable and various dependent variables are linearly related to each other**.
- The relation is said to be linear due to the correlation between the variables.
- E.g.
    - **An agriculture expert decides to study the crops that were ruined in a certain region. He collects the data about recent climatic changes, water supply, irrigation methods, pesticide usage, etc. To understand why the crops are turning black, do not yield any fruits and dry out soon.**

Steps for Multivariate Regression

1. Select the features

- Features that are highly responsible for the change in your dependent variable.

2. Normalize the feature

- Scale them in a certain range (preferably 0-1) so that analysing them gets a bit easy.

3. Select Loss function and Hypothesis

- A formulated hypothesis is nothing but a predicted value of the response variable and is denoted by h(x).
- A loss function is a calculated loss when the hypothesis predicts a wrong value.

Steps for Multivariate Regression

4. Minimize the loss function

- Loss M\minimization algorithms can be run over the datasets. These algorithms then adjust the parameters of the hypothesis.
- One of the minimization algorithms that can be used is the gradient descent algorithm.

5. Test the hypothesis

- The formulated hypothesis is then tested with a test set to check its accuracy and correctness.

Multivariable regression

Solved Problem:

https://www.statology.org/multiple-linear-regression-by-hand/

Applications of Regression

- **Forecasting continuous outcomes** like house prices, stock prices, or sales.
- **Predicting the success of future retail sales** or marketing campaigns to ensure resources are used effectively.
- **Predicting customer or user trends**, such as on streaming services or ecommerce websites.
- Analysing datasets to establish the relationships between variables and an output.
- **Predicting interest rates** or stock prices from a variety of factors.
- Creating **time series visualisations.**

Real Life Examples on Linear Regression

1. Businesses often use linear regression to understand the relationship between advertising spending and revenue.
   - The regression model would take the following form:

     revenue = β_0 + β_1(ad spending)
   - The coefficient **β_0** would represent total expected revenue when ad spending is zero.
   - The coefficient **β_1** would represent the average change in total revenue when ad spending is increased by one unit (e.g. one dollar).
   - If β_1 is negative, it would mean that more ad spending is associated with less revenue.
   - If β_1 is close to zero, it would mean that ad spending has little effect on revenue.
   - And if β_1 is positive, it would mean more ad spending is associated with more revenue.
   - Depending on the value of β_1, a company may decide to either decrease or increase their ad spending.

Real Life Examples on Linear Regression

2. Medical researchers often use linear regression to understand the relationship between drug dosage and blood pressure of patients.
    - The regression model would take the following form:

      blood pressure = $\beta_0 + \beta_1$(dosage)
    - The coefficient $\boldsymbol{\beta_0}$ would represent the expected blood pressure when dosage is zero.
    - The coefficient $\boldsymbol{\beta_1}$ would represent the average change in blood pressure when dosage is increased by one unit.
    - If β_1 is negative, it would mean that an increase in dosage is associated with a decrease in blood pressure.
    - If β_1 is close to zero, it would mean that an increase in dosage is associated with no change in blood pressure.
    - If β_1 is positive, it would mean that an increase in dosage is associated with an increase in blood pressure.
    - Depending on the value of β_1, researchers may decide to change the dosage given to a patient.

Real Life Examples on Linear Regression

3. Agricultural scientists often use linear regression to measure the effect of fertilizer and water on crop yields.
    - The regression model would take the following form:

      crop yield = β_0 + β_1(amount of fertilizer) + β_2(amount of water)
    - The coefficient $\boldsymbol{\beta_0}$ would represent the expected crop yield with no fertilizer or water.
    - The coefficient $\boldsymbol{\beta_1}$ would represent the average change in crop yield when fertilizer is increased by one unit, *assuming the amount of water remains unchanged.*
    - The coefficient $\boldsymbol{\beta_2}$ would represent the average change in crop yield when water is increased by one unit, *assuming the amount of fertilizer remains unchanged.*
    - Depending on the values of β_1 and β_2, the scientists may change the amount of fertilizer and water used to maximize the crop yield.

Real Life Examples on Linear Regression

4. Data scientists for professional sports teams often use linear regression to measure the effect that different training regimens have on player performance.
   - For example, data scientists in the NBA might analyze how different amounts of weekly yoga sessions and weightlifting sessions affect the number of points a player scores.
   - The regression model would take the following form:

     points scored = β_0 + β_1(yoga sessions) + β_2(weightlifting sessions)
   - The coefficient β0 would represent the expected points scored for a player who participates in zero yoga sessions and zero weightlifting sessions.

Unit No: 3 Unit Name :Learning with regression

Lecture No: 14
Logistic Regression

Logistic regression introduction

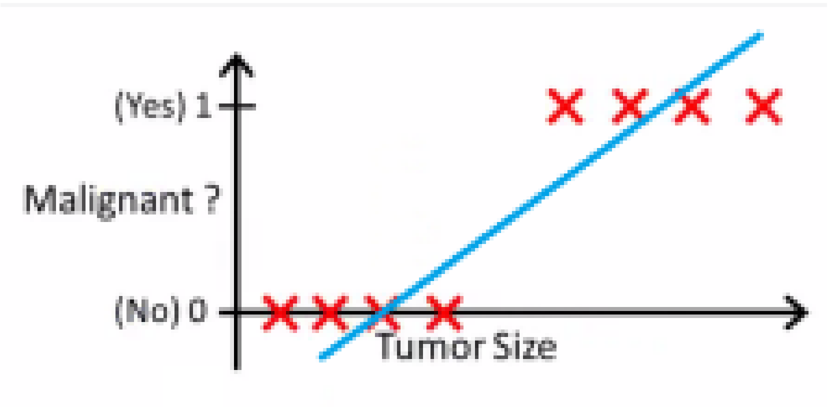
- Logistic regression models a relationship between predictor variables and a **categorical response variable**.

- Logistic regression helps us estimate a probability of falling into a certain level of the categorical response given a set of predictors

- We can choose from three types of logistic regression, depending on the nature of the categorical response variable:
  - Binary logistic regression
  - Nominal logistic regression
  - Ordinal logistic regression

Why not Linear Regression?

Suppose we have data of tumor size vs its malignancy. As it is a classification problem, if we plot, we can see, all the values will lie on 0 and 1. And if we fit the best-found regression line, by assuming the threshold at 0.5, we can do line pretty reasonable job.



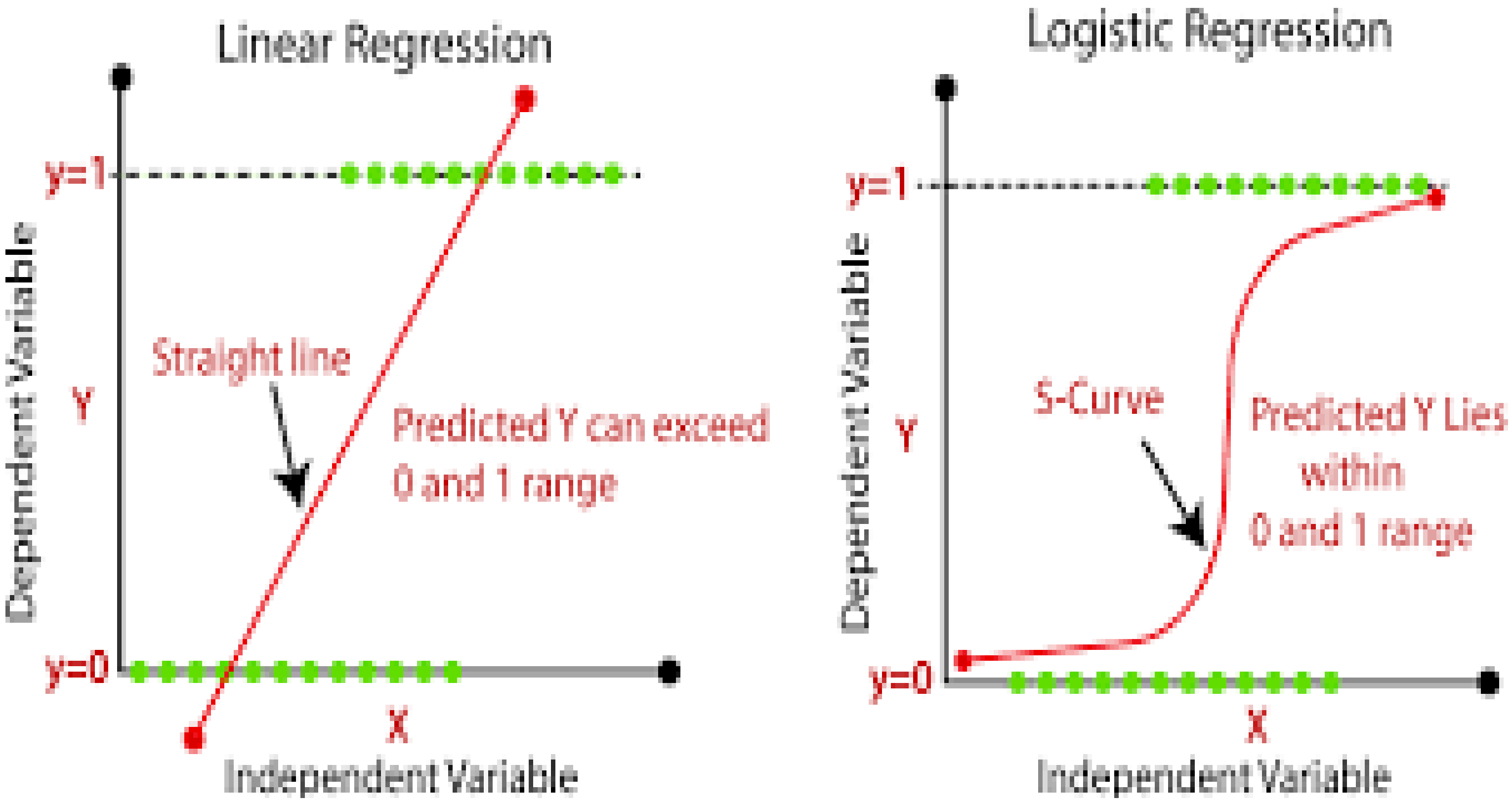
5

Why not Linear Regression?

1. We cannot use any of the well-established routines for statistical inference with least squares (e.g., confidence intervals, etc.), because these are based on a model in which the outcome is continuously distributed. At an even more basic level, it is hard to precisely interpret β

2. We cannot use this method when the number of classes exceeds 2. If we were to simply code the response as 1, . . . , K for a number of classes K > 2, then the ordering here would be arbitrary, but it actually matters

Why not Linear Regression?



Logistic regression

The y is usually a yes/no type of response.

This is usually interpreted as the probability of an event happening (y = 1) or not happening (y = 0). This can be deconstructed as:

- If y is an event (response, pass/fail, etc.),
- and p is the probability of the event happening (y = 1),
- then (1 - p) is the probability of the event not happening (y = 0),
- and p/(1 - p) are the odds of the event happening

$$P = \beta_0 + \beta_1 x$$

But there is an issue here, the value of (P) will exceed 1 or go below 0 and we know that range of Probability is (0-1). To overcome this issue we take ***"odds"*** of P:

$$\frac{P}{1-P} = \beta_0 + \beta_1 x$$

D Y PATIL DEEMED TO BE UNIVERSITY RAMRAO ADIK INSTITUTE OF TECHNOLOGY NAVI MUMBAI

Logistic regression

For a more general case, involving multiple independent variables, x, there is:

logit $= b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n$

The logit is the logarithm of the odds of the response, y, expressed as a function of independent or predictor variables, x, and a constant term.

D Y PATIL DEEMED TO BE UNIVERSITY RAMRAO ADIK INSTITUTE OF TECHNOLOGY NAVI MUMBAI

Logistic regression

The problem here is that the range is restricted and we don't want a restricted range because if we do so then our correlation will decrease.

It is difficult to model a variable that has a restricted range. To control this we take the ***log of odds*** which has a range from (-∞,+∞).

Logit function :

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$

This formulation is also useful for interpreting the model, since the logit can be interpreted as the log odds of a success

Logistic regression

we will multiply by ***exponent*** on both sides and then solve for P.

$$\exp[\log(\frac{p}{1-p})] = \exp(\beta_0 + \beta_1 x)$$

$$e^{\ln[\frac{p}{1-p}]} = e^{(\beta_0 + \beta_1 x)}$$

$$\frac{p}{1-p} = e^{\left(\beta_0 + \beta_1 x\right)}$$

$$p = e^{\left(\beta_0 + \beta_1 x\right)} - pe^{\left(\beta_0 + \beta_1 x\right)}$$

Logistic regression

$$p = p\left[\frac{e^{(\beta_0 + \beta_1 x)}}{p} - e^{(\beta_0 + \beta_1 x)}\right]$$

Euler's number, **e ≈ 2.71828**,

$$1 = \frac{e^{(\beta_0 + \beta_1 x)}}{p} - e^{(\beta_0 + \beta_1 x)}$$

$$p\left[1 + e^{(\beta_0 + \beta_1 x)}\right] = e^{(\beta_0 + \beta_1 x)}$$

$$p = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

$Now\ dividing\ by\ e^{(\beta_0 + \beta_1 x)},\ we\ will\ get$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad This\ is\ our\ sigmoid\ function.$$

Linear Vs Logistic regression

Linear regression	Logistic regression
Used to predict the continuous dependent variable using a given set of independent variables	Used to predict the categorical dependent variable using a given set of independent variables
Used for solving Regression problem Predict the value of continuous variables	Used for solving Classification problems Predict the values of categorical variables
Find the best fit line to predict the output Least square estimation method is used for estimation of accuracy Output must be a continuous value, such as price, age	Find the S-curve to classify the samples Maximum likelihood estimation method is used for estimation of accuracy Output must be a Categorical value such as 0 or 1, Yes or No
Required that relationship between dependent variable and independent variable must be linear There may be collinearity between the independent variables	Not required to have the linear relationship between the dependent and independent variable There should not be collinearity between the independent variable
No activation function is used No threshold value is needed	Activation function is used to convert a linear regression equation to the logistic regression equation A threshold value is added

Logistic regression- Example

The dataset of pass/fail in an exam for 5 students is given in the table below.
If we use **Logistic Regression** as the classifier and assume the model suggested by the optimizer will become the following for Odds of passing a course:

$$\log(Odds) = -64 + 2 \times hours$$

Hours Studies	Result (1=Pass, 0=Fail)
29	0
15	0
33	1
28	1
39	1

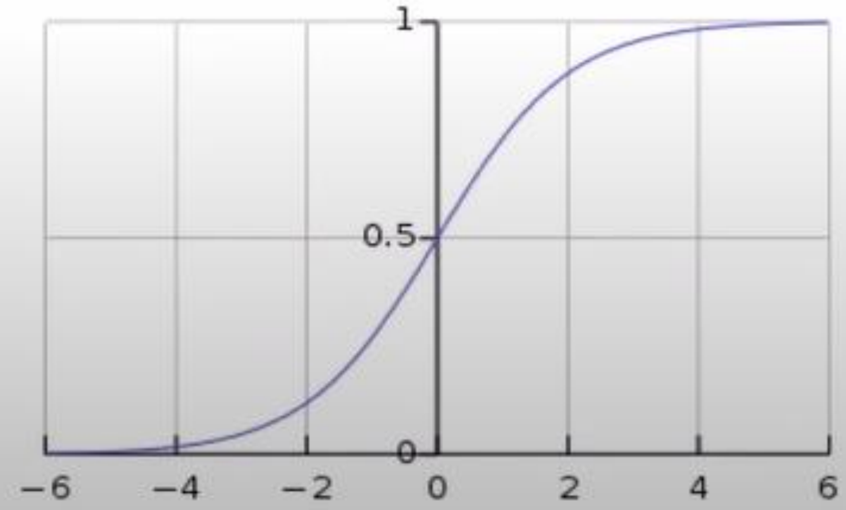
1) How to calculate the **probability of Pass** for the student who studied 33 hours?

D Y PATIL DEEMED TO BE UNIVERSITY
RAMRAO ADIK INSTITUTE OF TECHNOLOGY
NAVI MUMBAI

Logistic regression- Example

Sigmoid Function

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



$$P = \frac{1}{1 + e^{-z}}$$

$$z = -64 + 2 \cdot \text{Hours}$$

$\#\text{Hours} = ?$

Logistic regression- Example

$$Z = -64 + 2 \cdot \text{Hours}$$

$\#\text{Hours} = 33$

$$Z = -64 + 2 \cdot 33$$

$$-64 + 66$$

$$Z = 2$$

$$P = \frac{1}{1 + e^{-2}}$$

$$P = 0.88$$

$\therefore$ A Student who studies for 33 Hours has a 88% Chance of passing

Logistic regression-Excercise

The dataset of amount of saving and loan non defaulter is given in below table. Find the sigmoid function values for logistic regression

Log odd= - 4.0778+1.5046*amoumt of savings

Calculate the probability of loan non deafaulter for 2.5

X (Amount of savings)	Y (Loan Non Defaulter)
0.5	0
1.0	0
2.0	1
2.5	0
4.0	1

Unit No: 3 Unit Name : Learning with Regression

Lecture No: 15
Evaluation Metrics for Regression

Model Evaluation

- Model evaluation helps you to understand the performance of your model and makes it easy to present your model to others.
- There are 3 main metrics for model evaluation in regression:
  1. R Square/Adjusted R Square
  2. Mean Square Error(MSE)/Root Mean Square Error(RMSE)
  3. Mean Absolute Error(MAE)

R Square/Adjusted R Square

- R Square measures how much **variability in dependent variable** can be explained by the model.
- It is the **square of the Correlation Coefficient(R)** and that is why it is

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

- R Square is calculated by the sum of squared of prediction error divided by the total sum of the square which replaces the calculated prediction with mean.
- R Square value is between 0 to 1 and a bigger value indicates a better fit between prediction and actual value.
- R Square is a good measure to determine how well the model fits the dependent variables. However, it does not take into consideration of overfitting problem.
- Adjusted R Square is introduced because it will penalize additional independent variables added to the model and adjust

Mean Square Error(MSE)/Root Mean Square Error(RMSE)

- While R Square is a relative measure of how well the model fits dependent variables, Mean Square Error is an absolute measure of the goodness of the fit.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

- MSE is calculated by the sum of square of prediction error which is real output minus predicted output and then divide by the number of data points.
- It gives you an absolute number on how much your predicted results deviate from the actual number.
- Root Mean Square Error(RMSE) is the square root of MSE. MSE is calculated by the square of error, and thus square root brings it back to the same level of prediction error and makes it easier for interpretation.

Mean Absolute Error(MAE)

- Mean Absolute Error(MAE) is similar to Mean Square Error(MSE). However, instead of the sum of square of error in MSE, MAE is t[illegible]e of error.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

- Compared to MSE or RMSE, MAE is a more direct representation of sum of error terms.
- MSE gives larger penalization to big prediction error by square it while MAE treats all errors the same.

Thank You