

Raw Data

RAW DATA					
NAME	ITEM	PRICE	MOBIL	GENDER	ADDRES
ABC	FOOD	6000			
MAX	SPORT	10K	678900	M	
DEF		3K		F	MUMBAI

Clean Data

CLEAN DATA					
NAME	ITEM	PRICE	MOBIL	GENDER	ADDRES
ABC	FOOD	6000	123	M	HYD
MAX	SPORT	10K	678900	M	BANGLORE
DEF	PLAYING	3K	568	F	MUMBAI

Regression Algorithm

House for your family									
x1	x2	x3	x4	x5	x6	x7	Y	x8	
sft	bhk room	school	college	office	hospital	super market	view	price	mobl
100	3	y	n	y	y	green	1cr	123	
200		n	y				2cr	345	
300	2	y	y	y	n		3cr	567	
							4cr		
							5cr		

REGRESSION

Classification Algorithm

x1	x2	x3	x4	x5	x6	x7	x8	x9	Y	
sft	bhk room	school	college	office	hospital	super market	view	price	mobl	purchase
100	3	y	n	y	y	green	1cr	123	Y	
200		n	y				2cr	345	n	
300	2	y	y	y	n		3cr	567	y	
							4cr		n	
							5cr		y	

Classification

Irrelevant Attributes for buying a house-Dataset2

A-DATASET														
x1	x2	x3	x4	x5	x6	x7	x8	x9	Y	+ 10 Attribute				
sft	bhk room	school	college	office	hospital	super market	view	price	mobl		purchase			
100	3	y	n	y	y	green	1cr	123	Y					
200		n	y				2cr	345	n					
300	2	y	y	y	n		3cr	567	y		fication algorithm			
							4cr		n					
							5cr		y					
B-DATASET														
13 attribute	x1	x2	x3	x4	x5	x6	x7	x8	x9	Y	x10	x11	x12	
	sft	bhk room	school	college	office	hospital	super market	view	price	mobl	purchase	gender	skin	weight
	100	3	y	n	y	y	green	1cr	123	Y	m	w	50	
	200		n	y				2cr	345	n	f		60	
	300	2	y	y	y	n		3cr	567	y	m	b	70	

Handling missing dataset

NUMBER	SEASON	temp					sumer	35-50
10	SUMMER	45					winter	5 to 15
20	WINTER	7					rainy	20-30
30.25	rainy	23						
34	RAINY	30						
30.25	SUMMER	50						
57	WINTER	10						
34	RAINY	22						
67								
34								
34								

Handling missing dataset - Dummy Variable

NUMBER	SEASON	temp	sumer	35-50				
10	SUMMER	45	winter	5 to 15				
20	WINTER	7	rainy	20-30				
30.25	rainy	23						
34	RAINY	30						
30.25	SUMMER	50						
57	WINTER	10						
34	RAINY	22						
67								
34								
34								
dummy variable								
SEASON	SUMMER	WINTER	RAINY					
SUMMER	1	0	0					
WINTER	0	1	0					
rainy	0	0	1					
RAINY	0	0	1					
SUMMER	1	0	0					
WINTER	0	1	0					
RAINY	0	0	1					

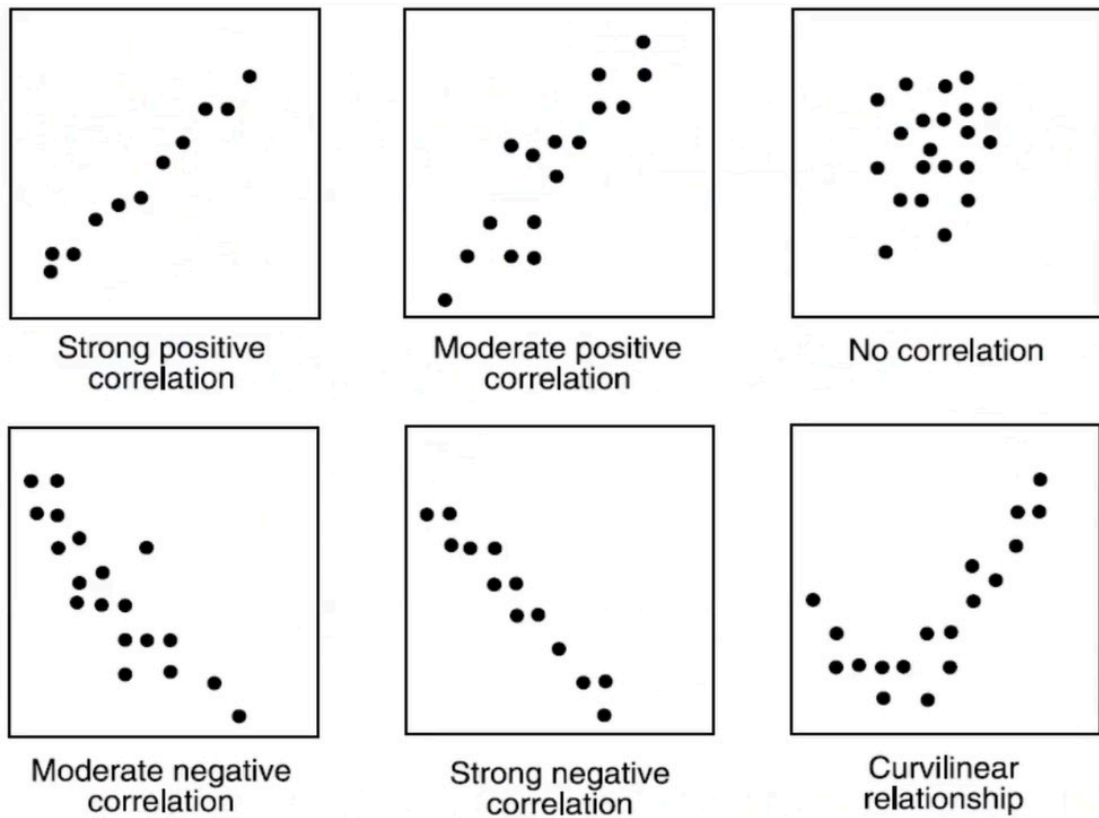
Label encoder

LABEL ENCODER			
SEASON			
SUMMER	0		
WINTER	1		
rainy	2		
RAINY	2		
SUMMER	0		
WINTER	1		
RAINY	2		
AUTOMN	3		

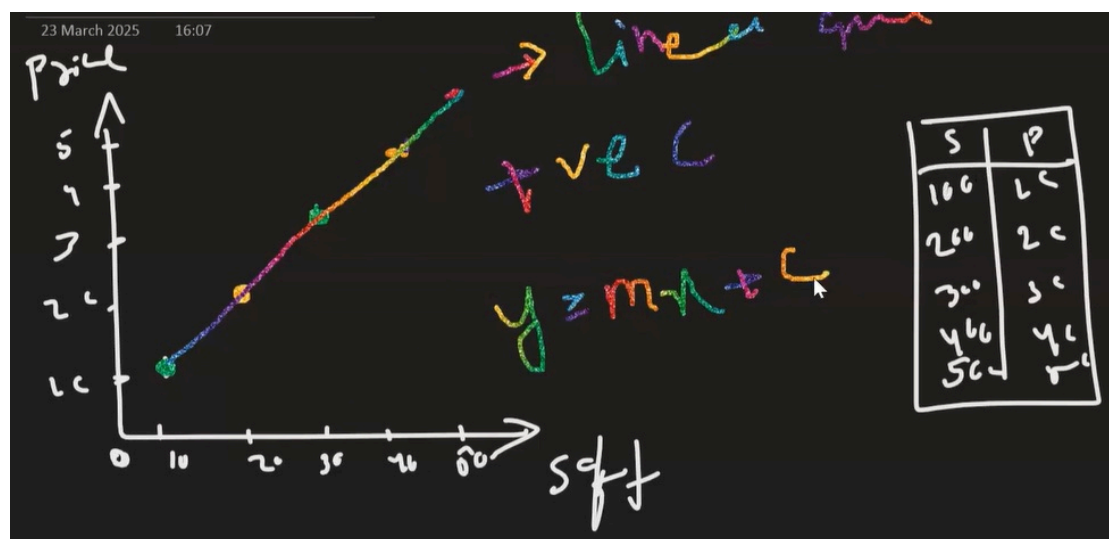
1 hot encoder

one hot encoder			
SEASON	SUMMER	WINTER	
SUMMER +	1	0	
WINTER	0	1	
rainy	0	0	
RAINY	0	0	
SUMMER	1	0	
WINTER	0	1	
RAINY	0	0	

Co-relation



Positive Co-relation



In []: