Machine Learning Algorithms
Linear and Logistic Regression

Raghu Prasad K S

Introduction

- Raghu Prasad – BE, MS
- Total of 29 years of experience
- 7 years as a lecturer in an Engineering College
- 22 Years into IT
- Worked with companies like CISCO,CSC,ICICI,First Apex – NTT Data
- Currently into Corporate training and consultancy
- Worked with corporates and public sector
- **Technologies** – Java,Python,Web technologies,Java Script technologies (MEAN stack),IOT,Test Automation,Machine Learning,Artificial Intelligence,ERP,.NET

Topics

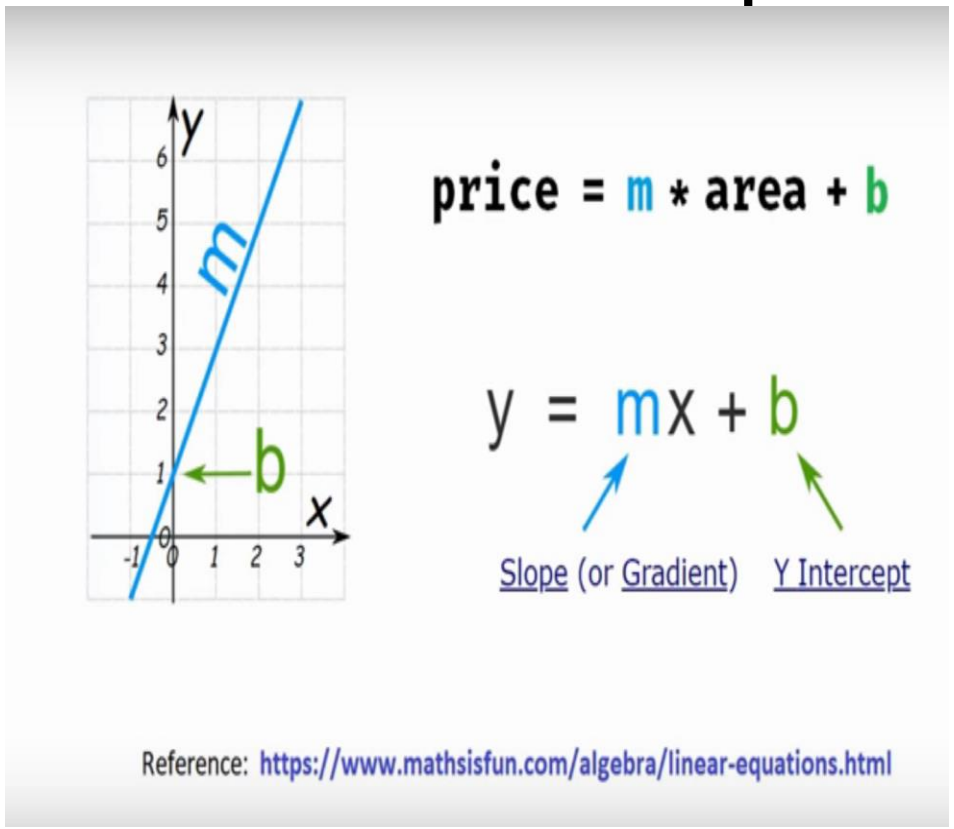
- Supervised machine learning
  - Linear Regression
  - Logistic Regression

Supervised Machine Learning

- **Supervised Learning**
- **How it works:** This algorithm consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data.
- Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.

Linear Regression – Single Variable

- Prediction of home price



price = m * area + b

Dependent variable

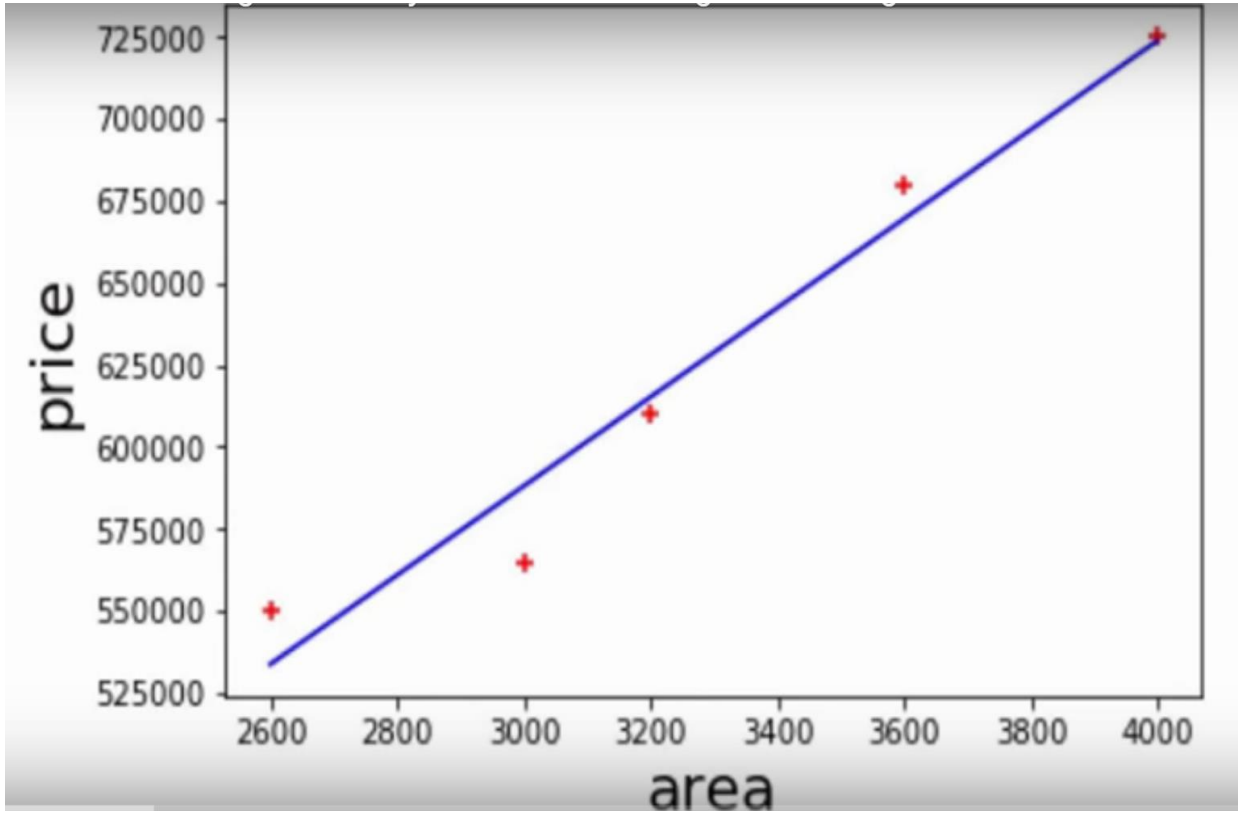
Independent variable

Linear Regression – Single Variable

- Prediction of home price

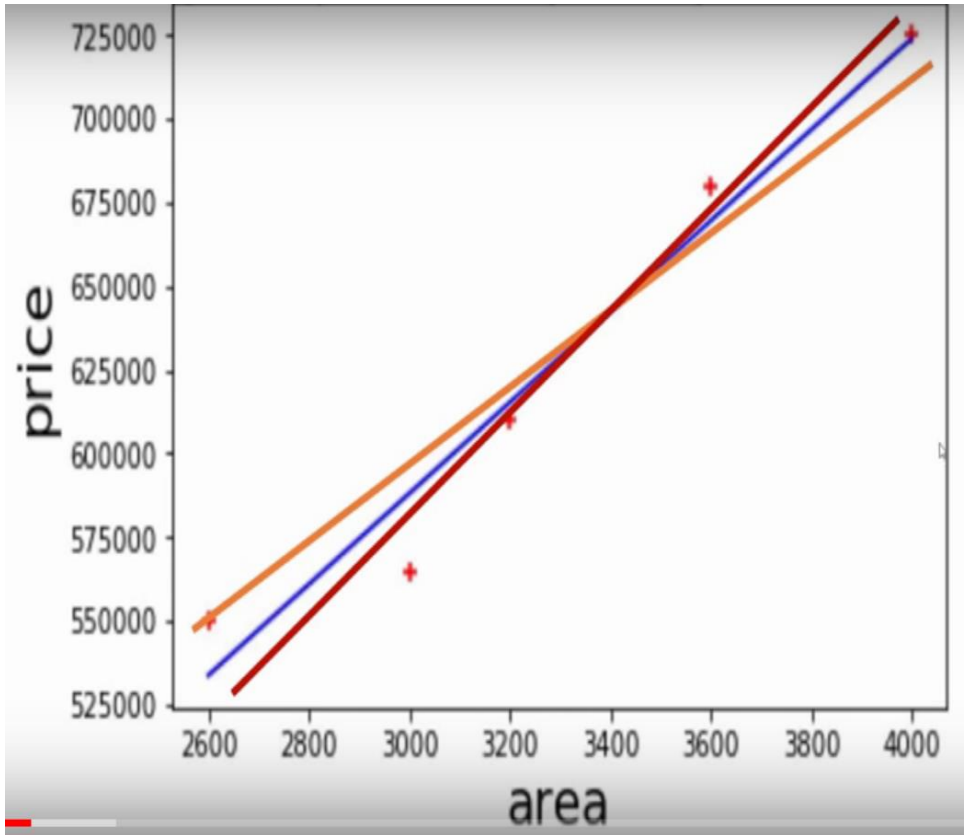
Home prices in Monroe Twp, NJ (USA)

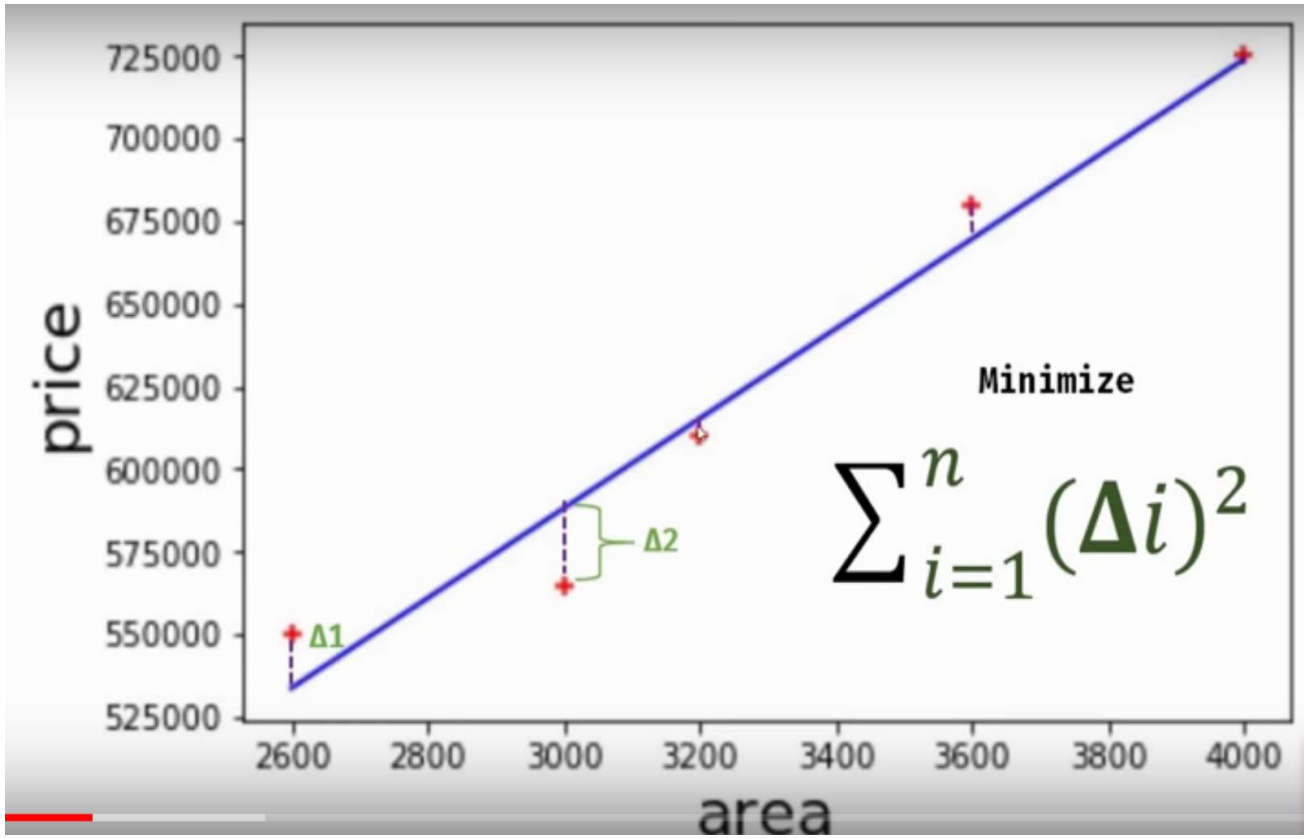
area	price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000



Linear Regression – Single Variable

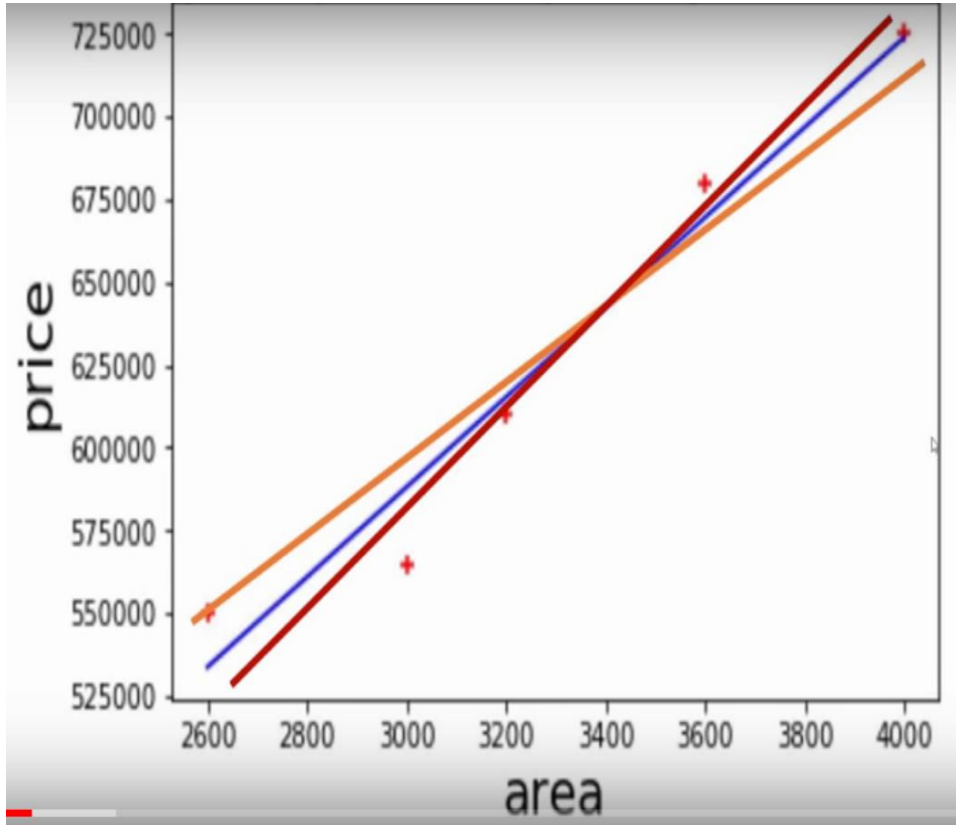
- Bestfit line

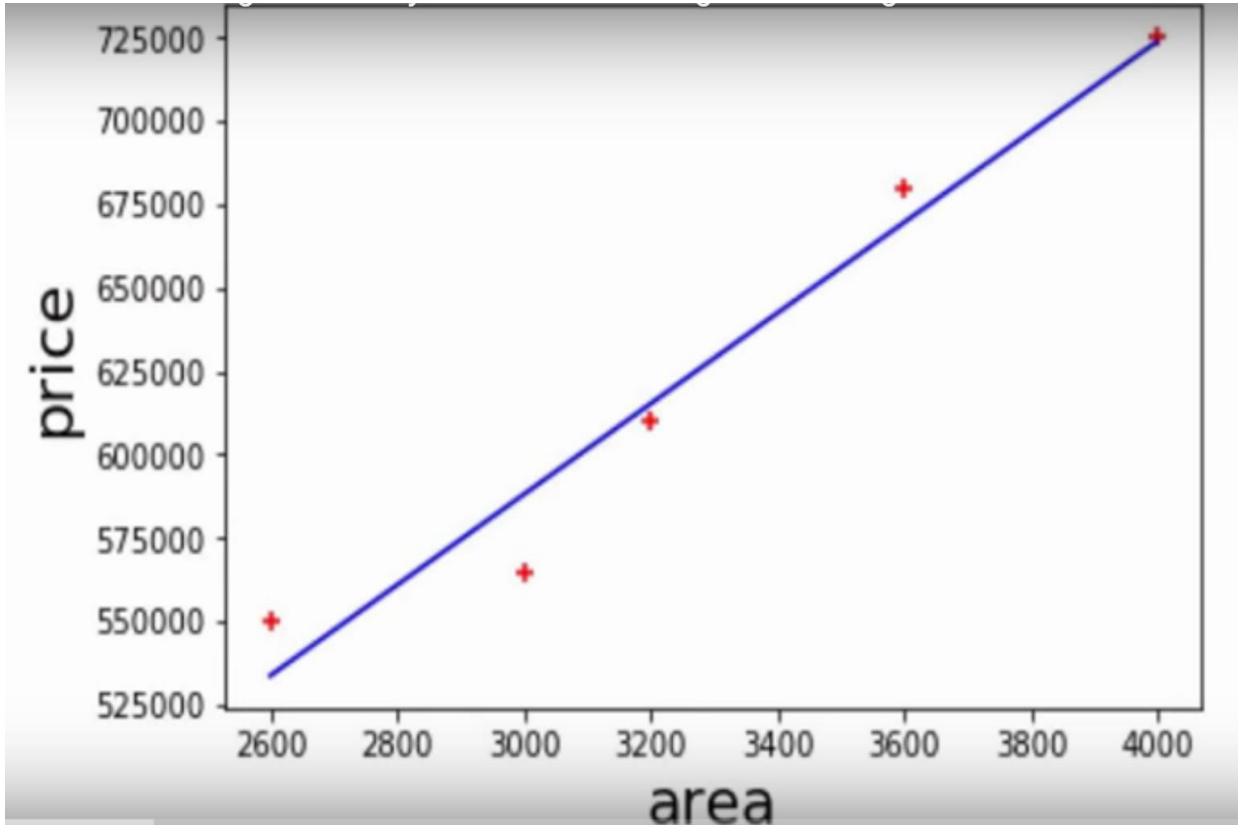




Linear Regression – Single Variable

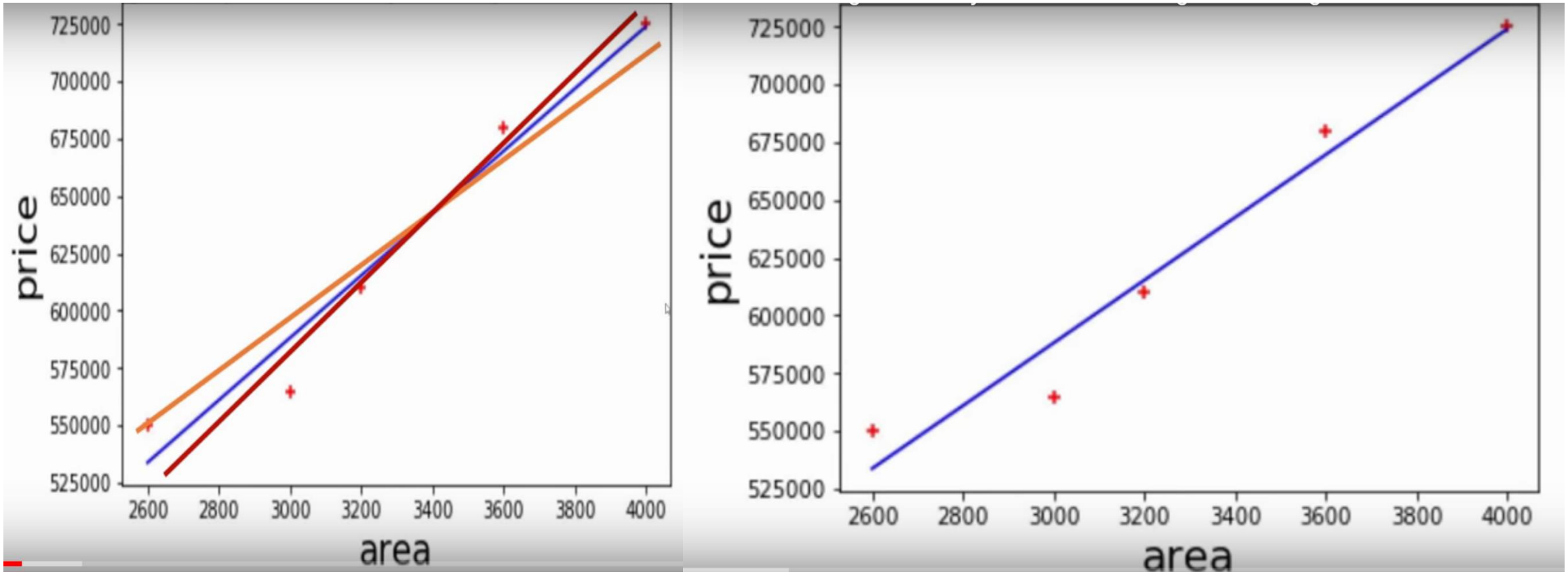
- Bestfit line





Linear Regression – Single Variable

- Bestfit line

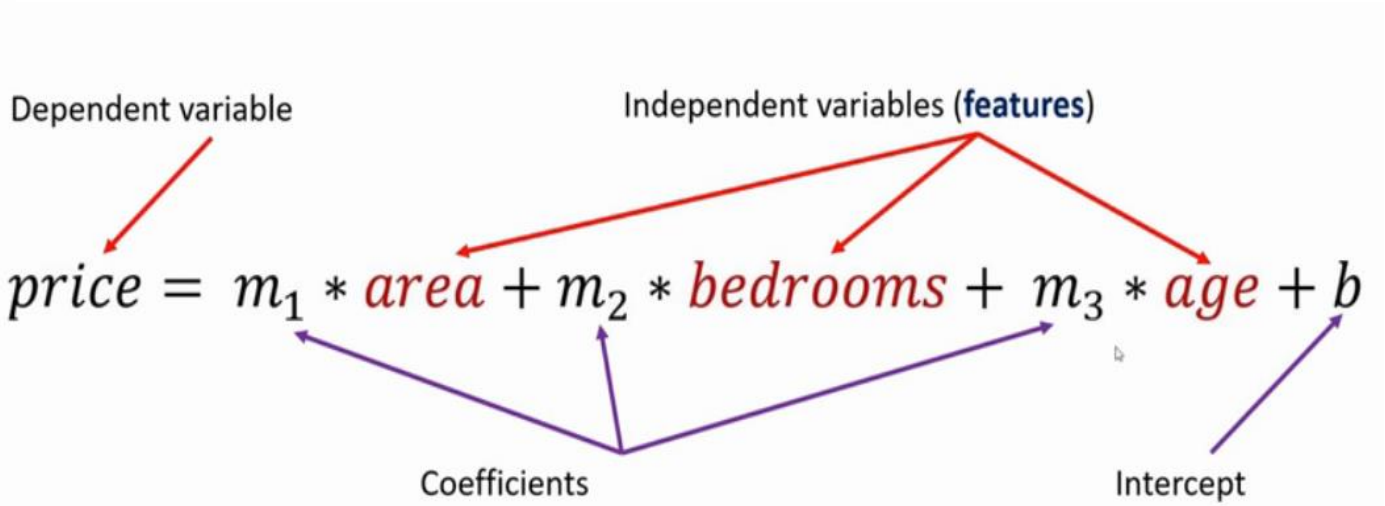


Linear Regression – Multiple Variable

- Home price

Home prices in Monroe Township, NJ (USA)

area	bedrooms	age	price
2600	3	20	550000
3000	4	15	565000
3200		18	610000
3600	3	30	595000
4000	5	8	760000



Dummy Variables – One Hot Encoding

- Using Pandas – get_dummies
- Using sklearn – OneHotEncoding

How should we handle text data in numeric model?

town	area	price
monroe township	2600	550000
monroe township	3000	565000
monroe township	3200	610000
monroe township	3600	680000
monroe township	4000	725000
west windsor	2600	585000
west windsor	2800	615000
west windsor	3300	650000
west windsor	3600	710000
robbinsville	2600	575000
robbinsville	2900	600000
robbinsville	3100	620000

town	area	price
1	2600	550000
1	3000	565000
1	3200	610000
1	3600	680000
1	4000	725000
2	2600	585000
2	2800	615000
2	3300	650000
2	3600	710000
3	2600	575000

Dummy Variables – One Hot Encoding

- Using Pandas – get_dummies
- Using sklearn – OneHotEncoding

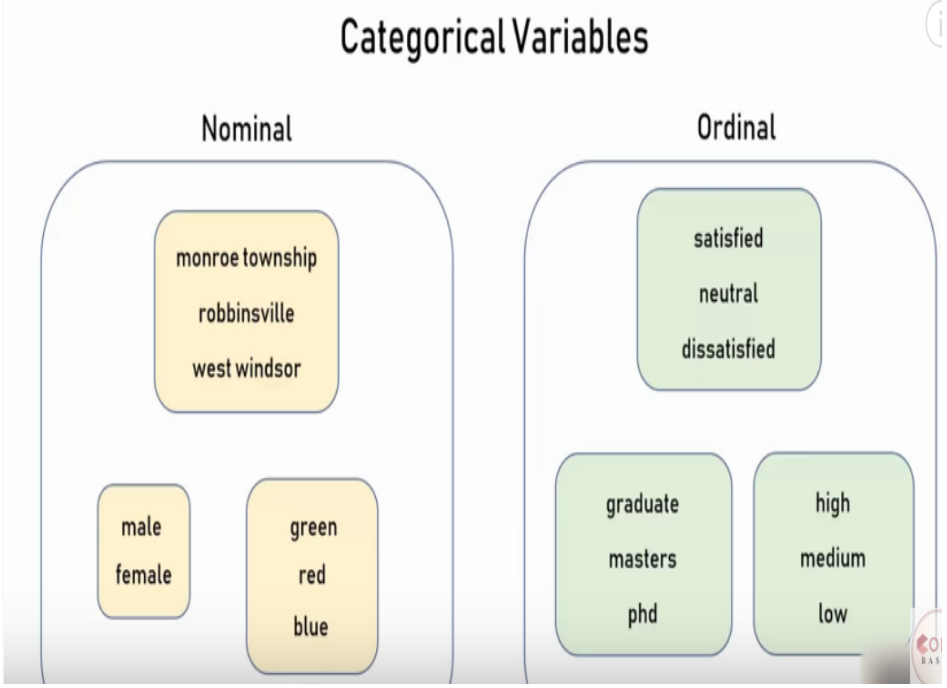
How should we handle text data in numeric model?

town	area	price
monroe township	2600	550000
monroe township	3000	565000
monroe township	3200	610000
monroe township	3600	680000
monroe township	4000	725000
west windsor	2600	585000
west windsor	2800	615000
west windsor	3300	650000
west windsor	3600	710000
robbinsville	2600	575000
robbinsville	2900	600000
robbinsville	3100	620000

town	area	price
1	2600	550000
1	3000	565000
1	3200	610000
1	3600	680000
1	4000	725000
2	2600	585000
2	2800	615000
2	3300	650000
2	3600	710000
3	2600	575000

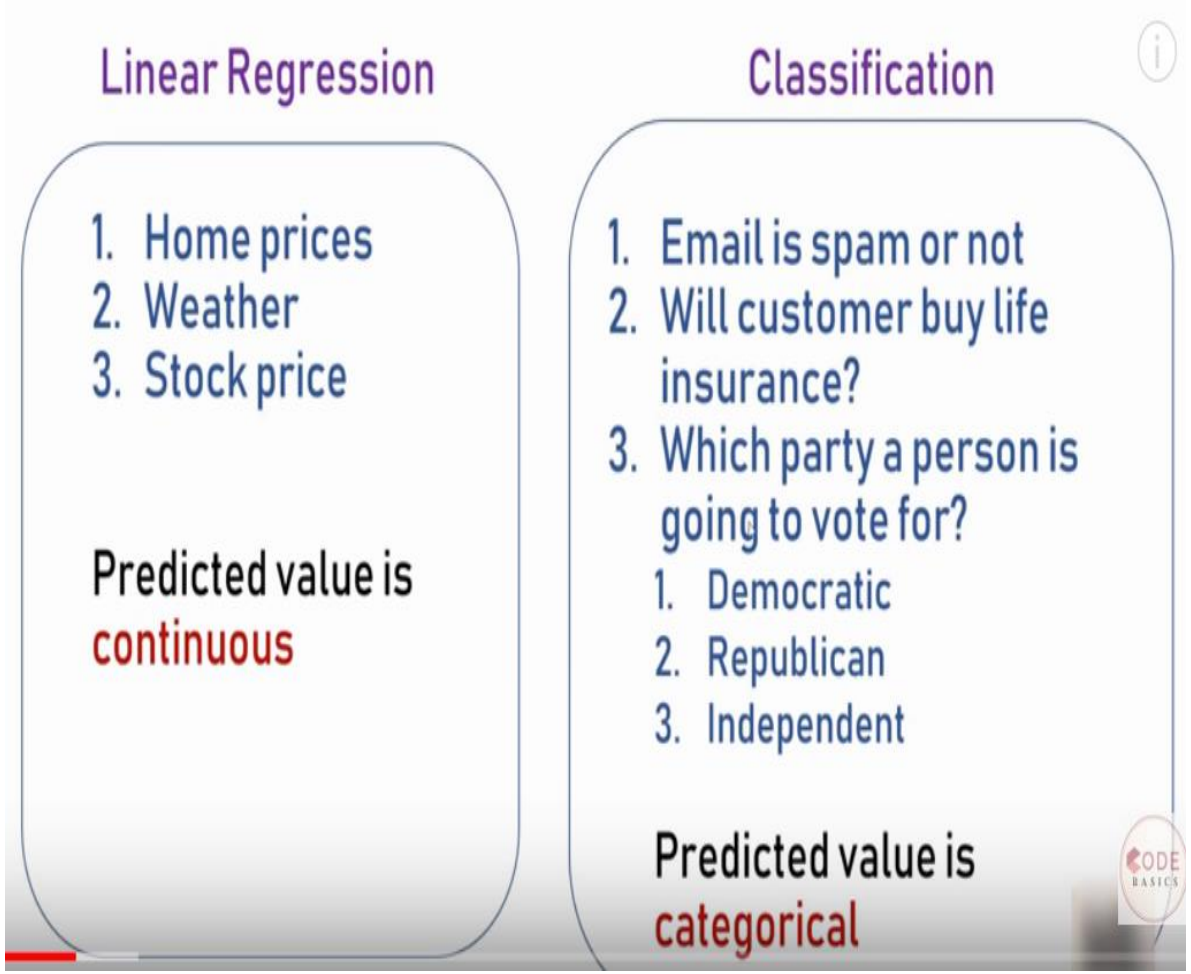
Dummy Variables – One Hot Encoding

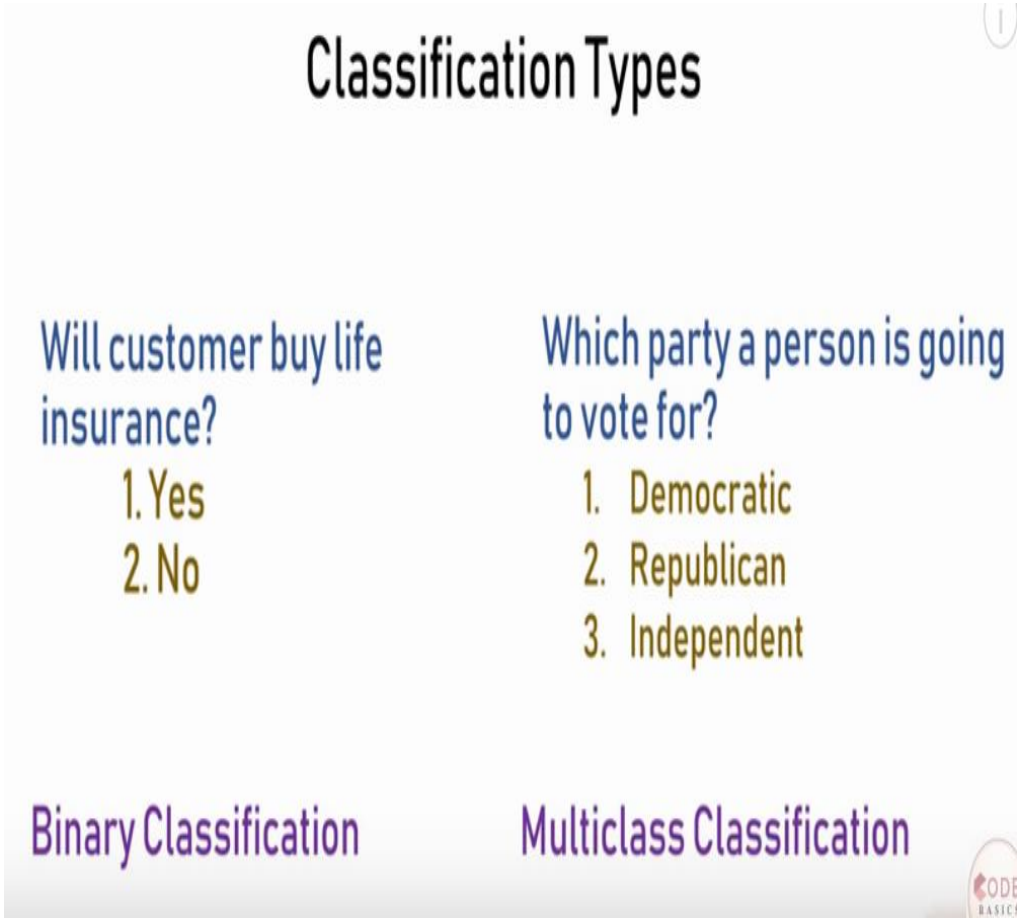
- Using Pandas – get_dummies
- Using sklearn – OneHotEncoding



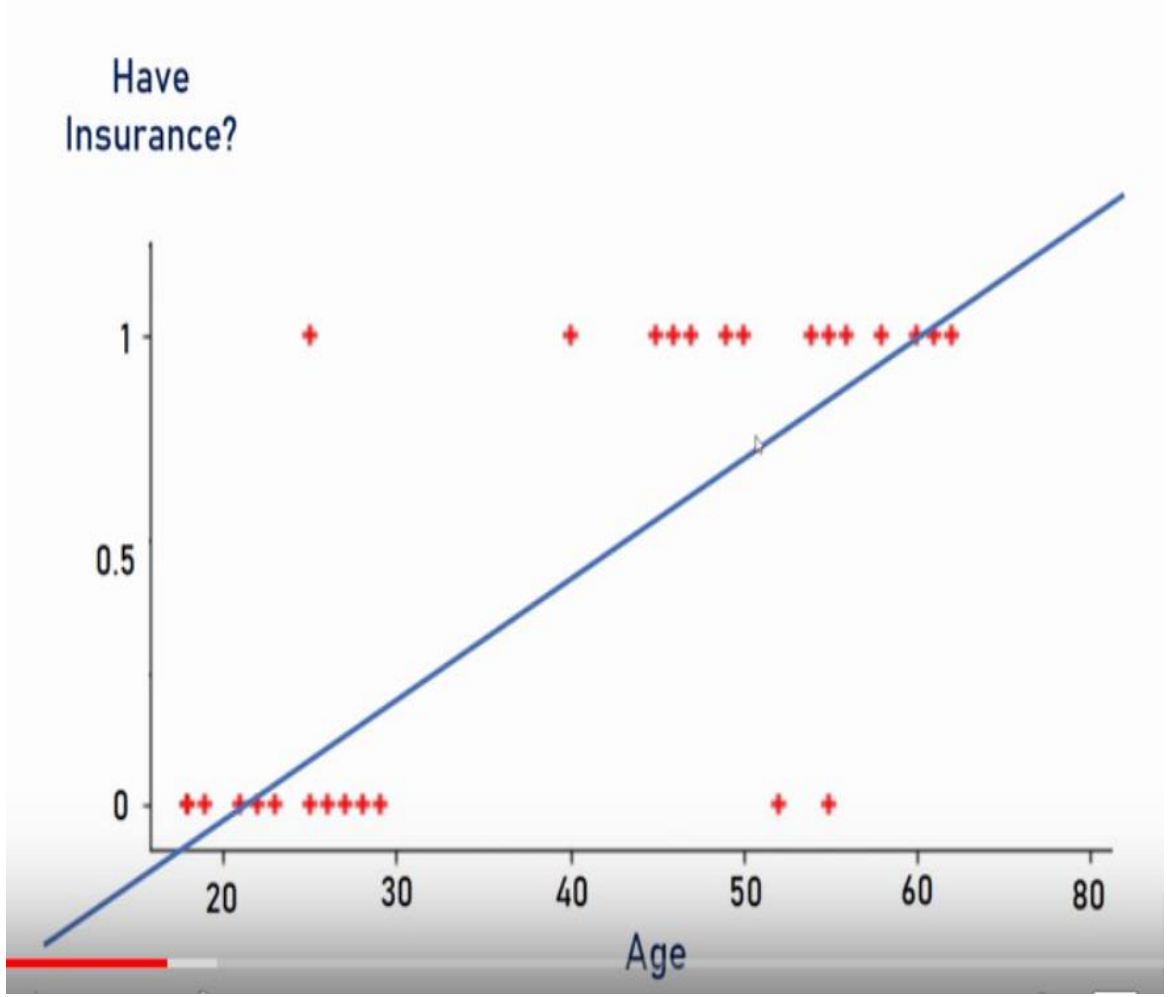
town	area	price	monroe township	west windsor	robbinsville
monroe township	2600	550000	1	0	0
monroe township	3000	565000	1	0	0
monroe township	3200	610000	1	0	0
monroe township	3600	680000	1	0	0
monroe township	4000	725000	1	0	0
west windsor	2600	585000	0	1	0
west windsor	2800	615000	0	1	0
west windsor	3300	650000	0	1	0
west windsor	3600	710000	0	1	0
robbinsville	2600	575000	0	0	1

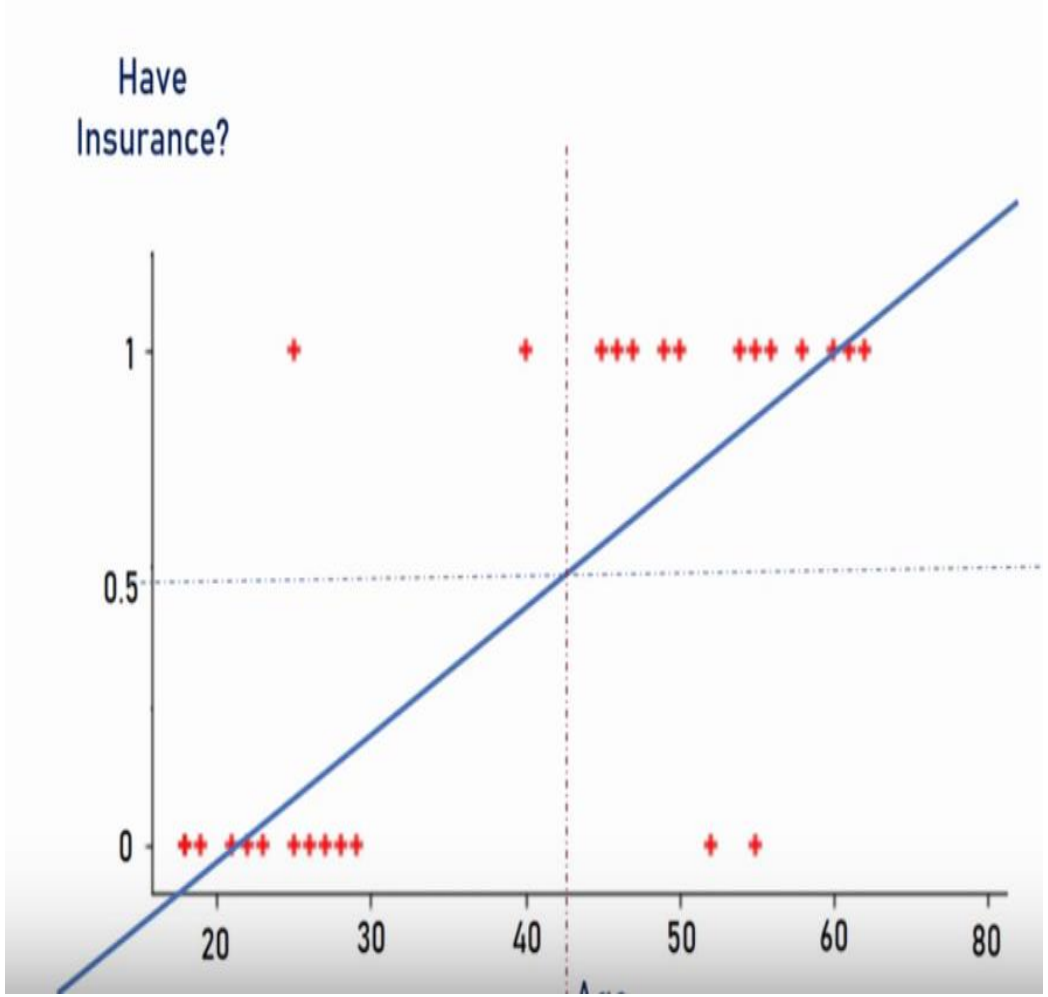
Logistic Regression – Binary Classification



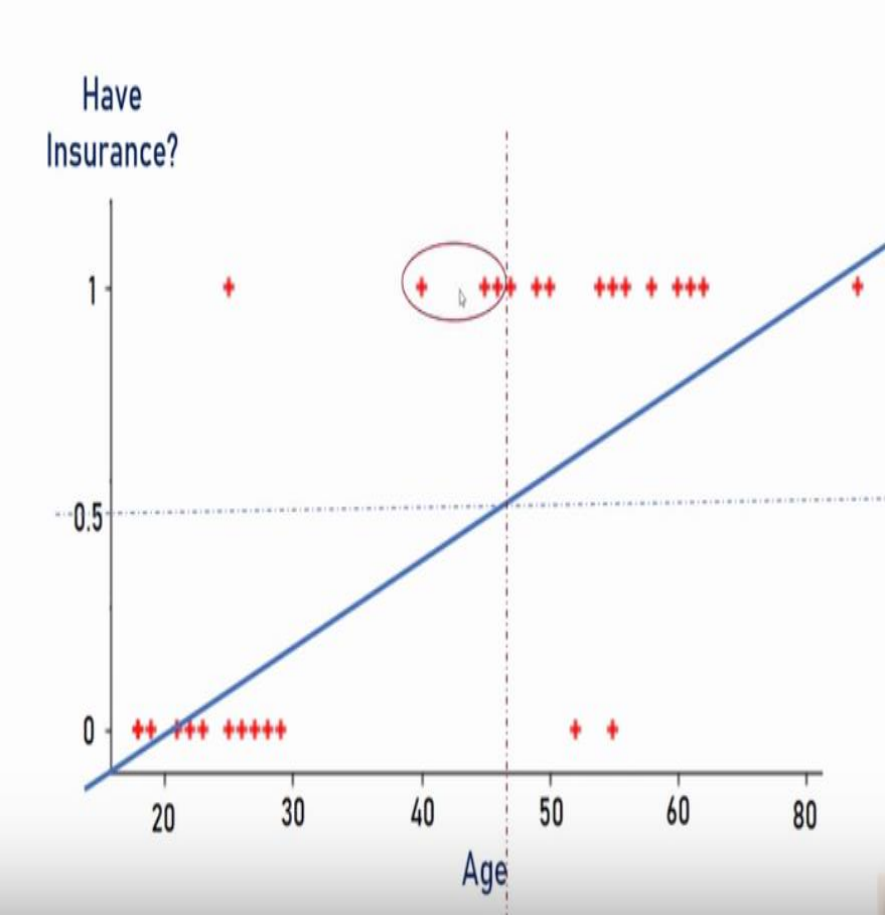


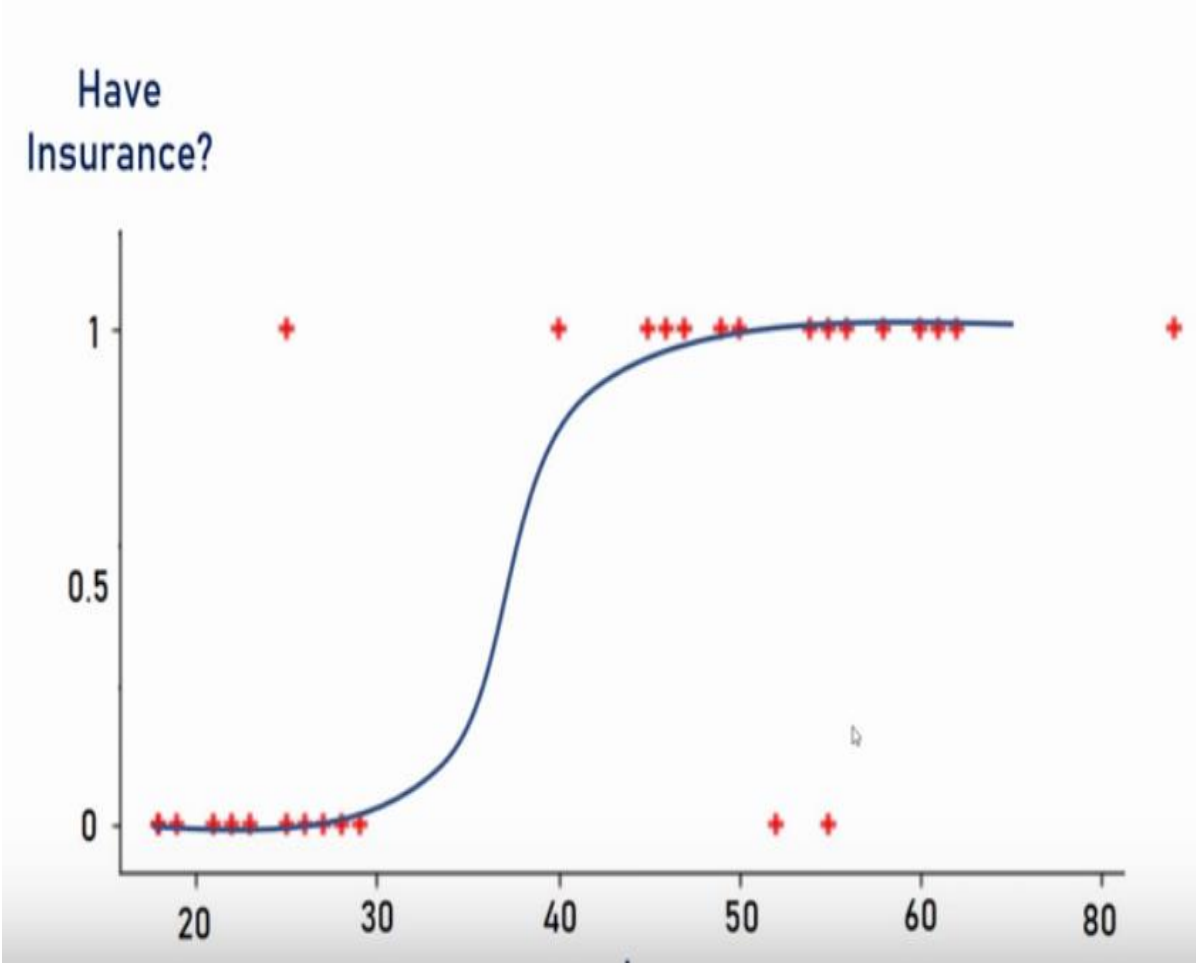
Logistic Regression – Binary Classification



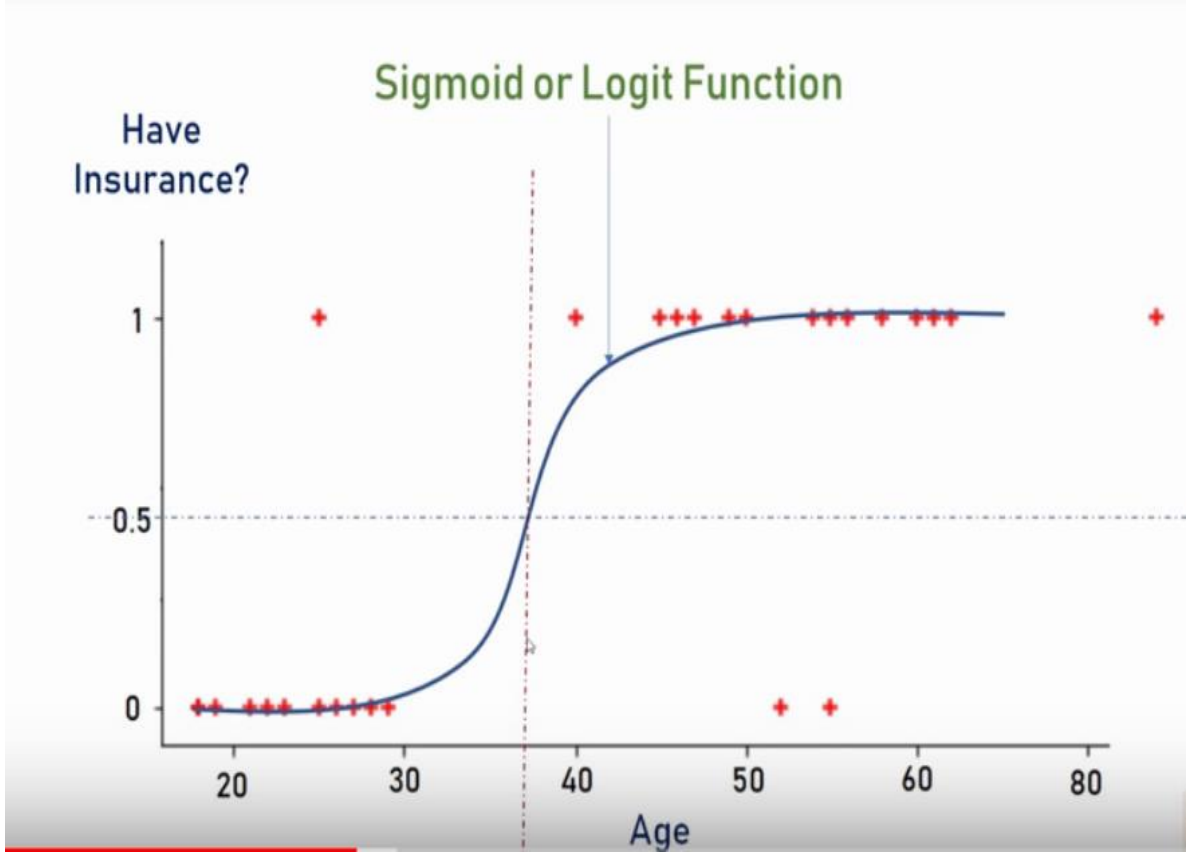


Logistic Regression – Binary Classification





Logistic Regression – Binary Classification



$$sigmoid(z) = \frac{1}{1 + e^{-z}}$$

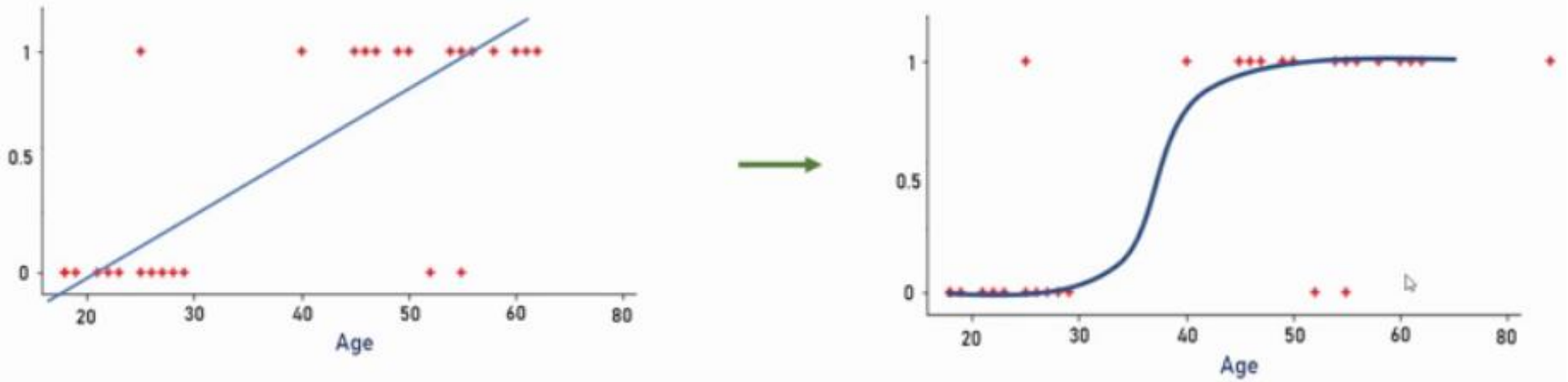
e = Euler's number ~ 2.71828

Sigmoid function converts input into range 0 to 1

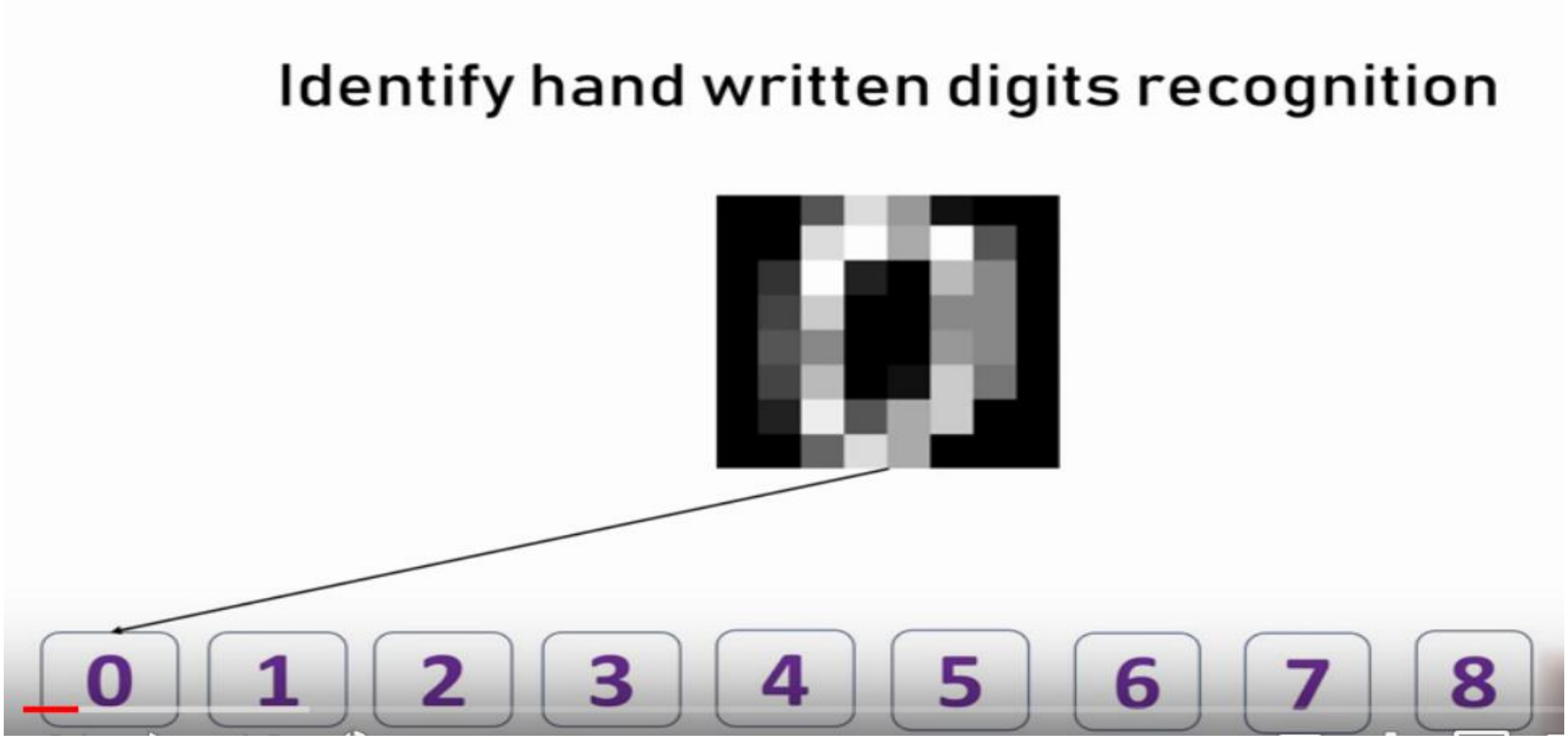
Logistic Regression – Binary Classification

$$y = m * x + b$$

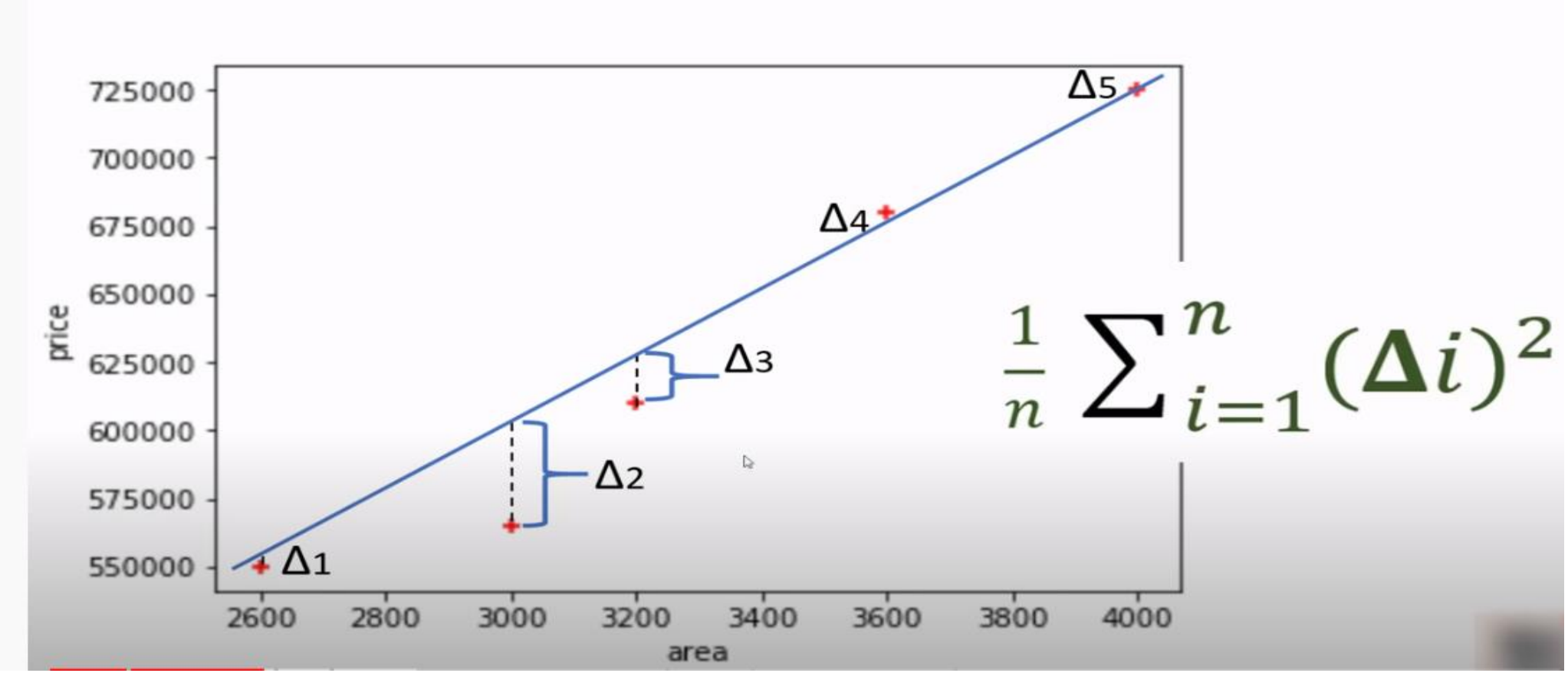
$$y = \frac{1}{1 + e^{-(m*x+b)}}$$



Logistic Regression – Multiclass Classification



Gradient Descent



Gradient Descent

Mean Squared Error

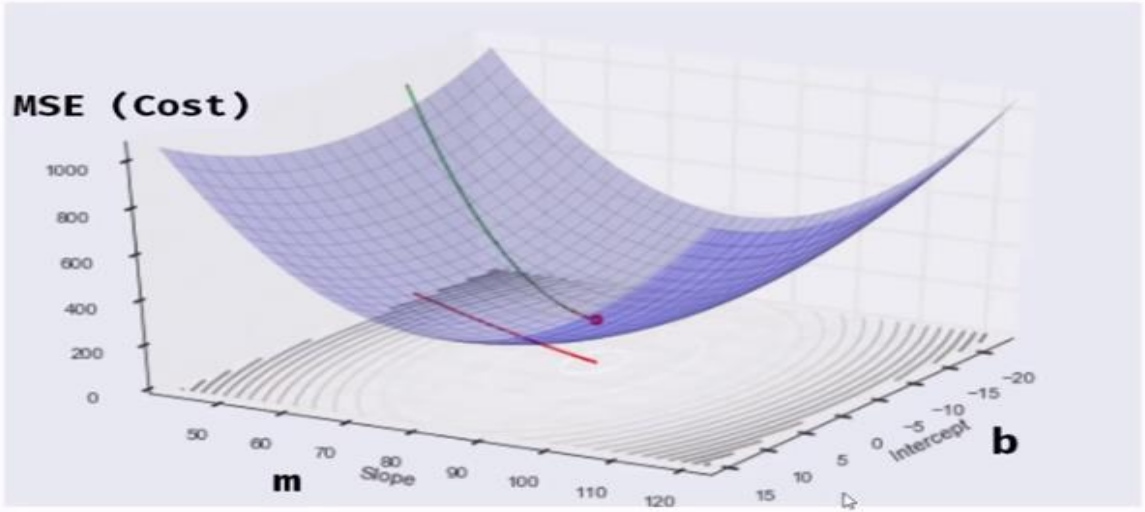
$$mse = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{predicted}\right)^2$$

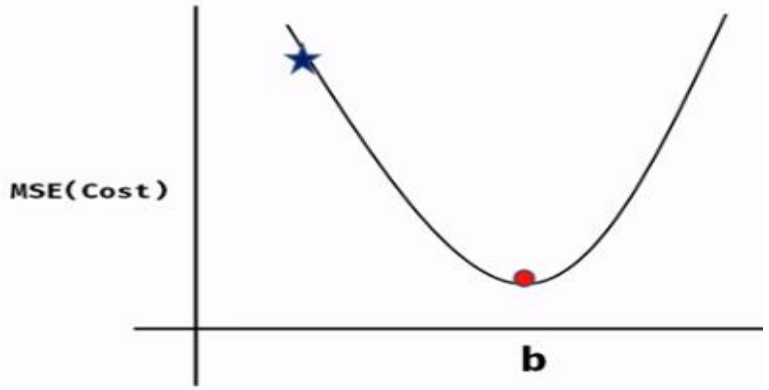
Mean Squared Error

$$mse = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2$$

Cost Function

Gradient Descent

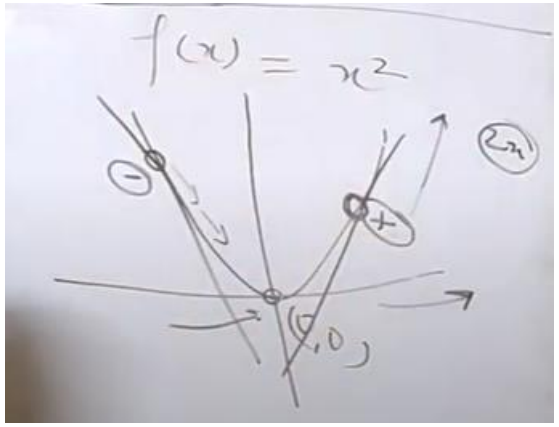


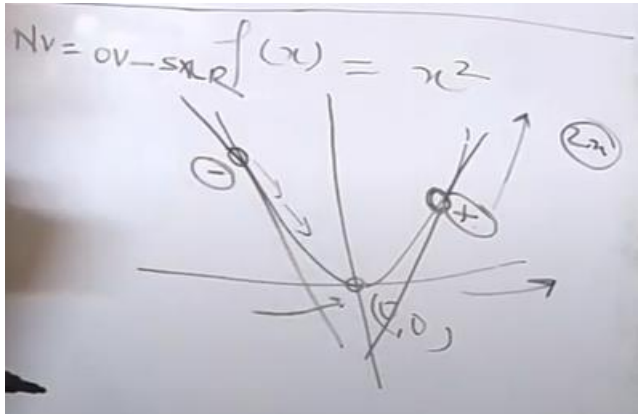


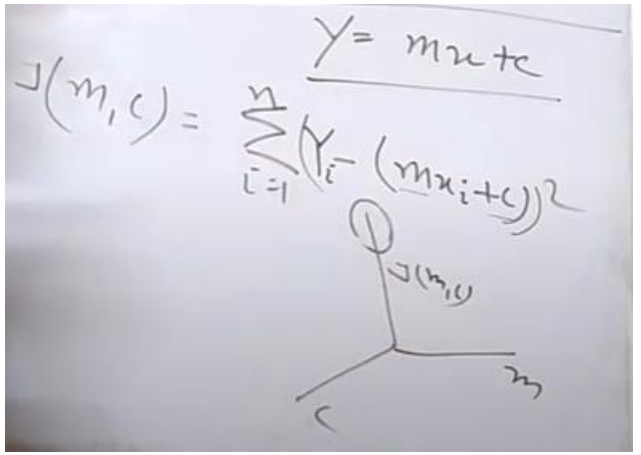
Gradient descent

- An algorithm to minimize a function by optimizing parameters
- Example – Maths test max marks 50
- Guess marks – 40. Its too much
- Guess marks – 30 .Its too less
- Guess marks – 35 .. Correct
- New Value = Old Value – step size
- Step size = Learing rate* slope

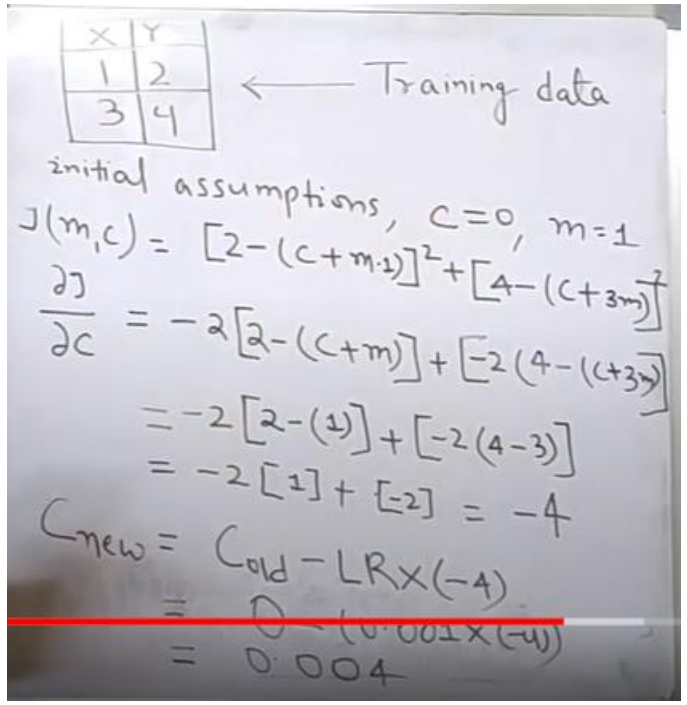
Gradient descent







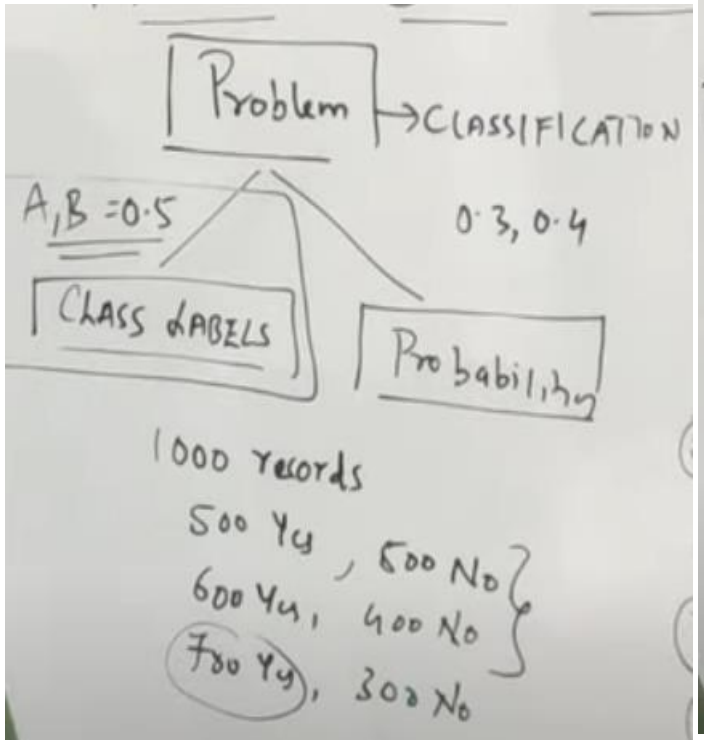
Gradient Descent

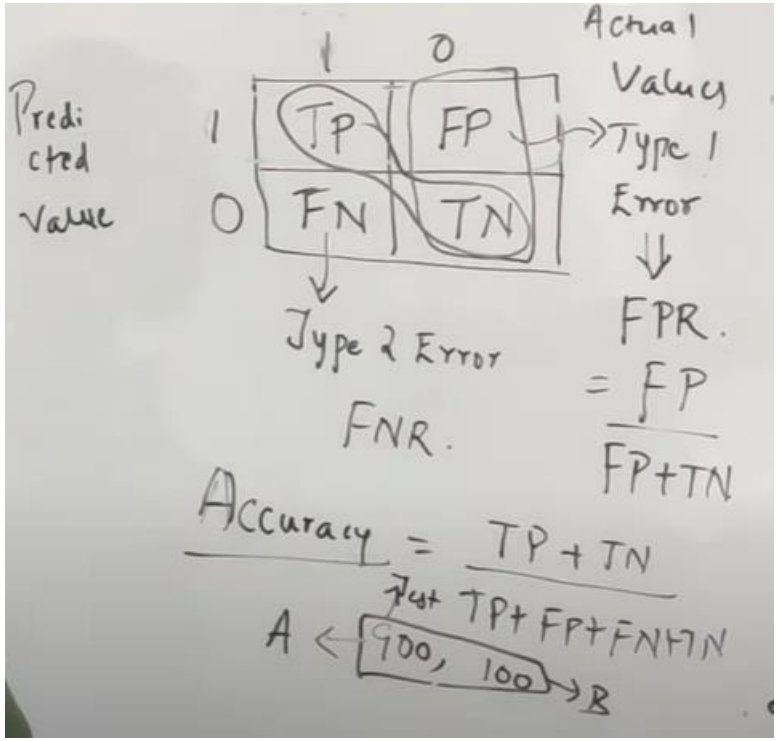


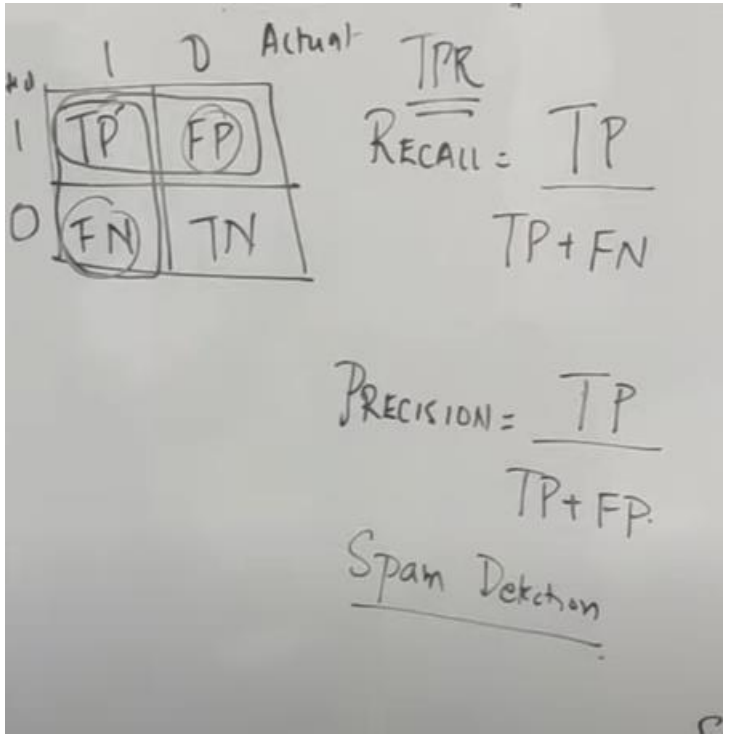
Metrics In Classification

- 1. Confusion Matrix
- 2. False Positive Rate – FPR (Type 1 Error)
- 3. False Negative Rate – FNR (Type 2 Error)
- 4. Recall (TPR,Sensitivity)
- 5. Precision (+ve Prod Val)
- 6. Accuracy
- 7. F1 Score (F Beta)
- 8. Cohen Kappa
- 9. RoC Curve,AUC Score
- 10. PR Curve

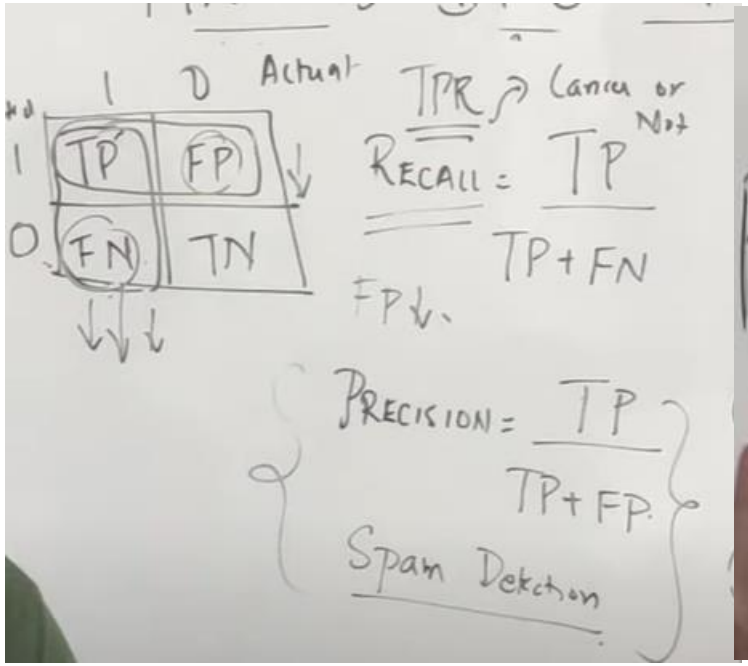
Metrics In Classification

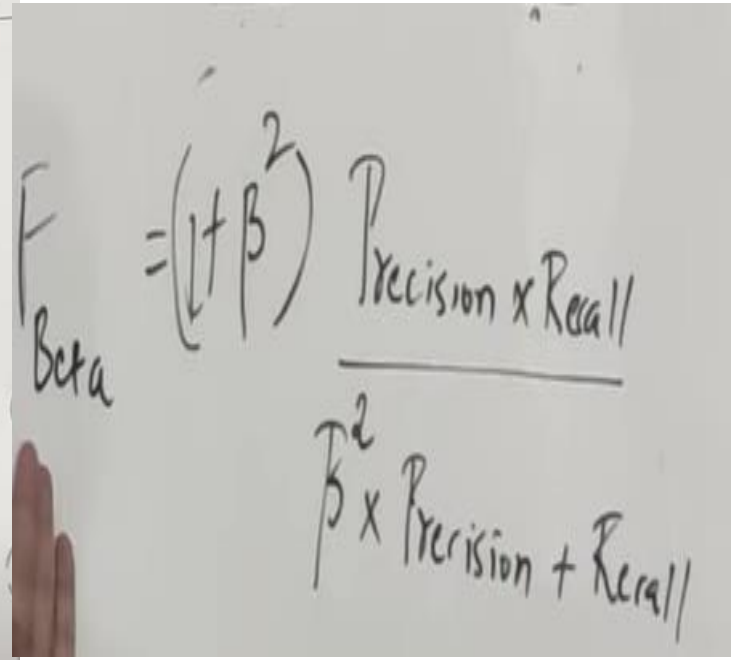


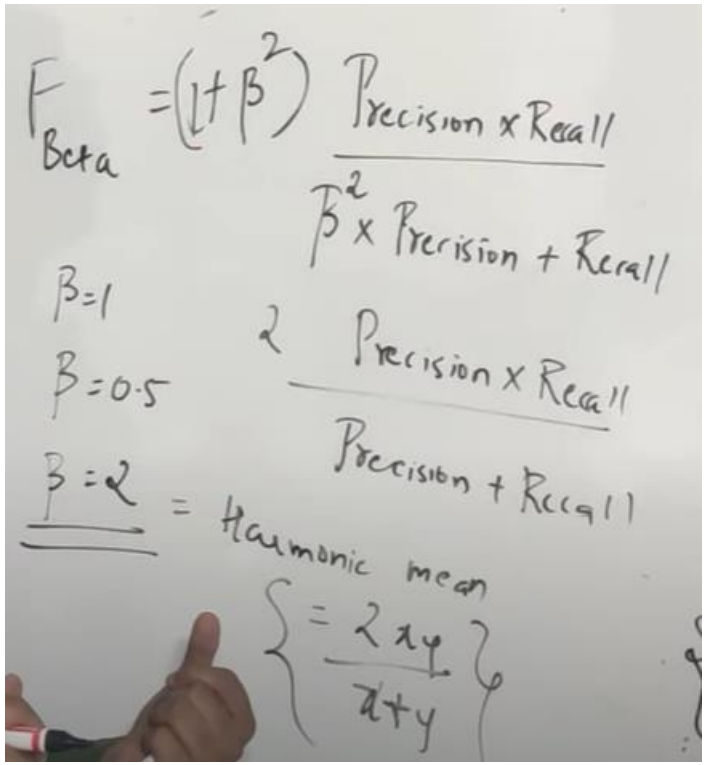




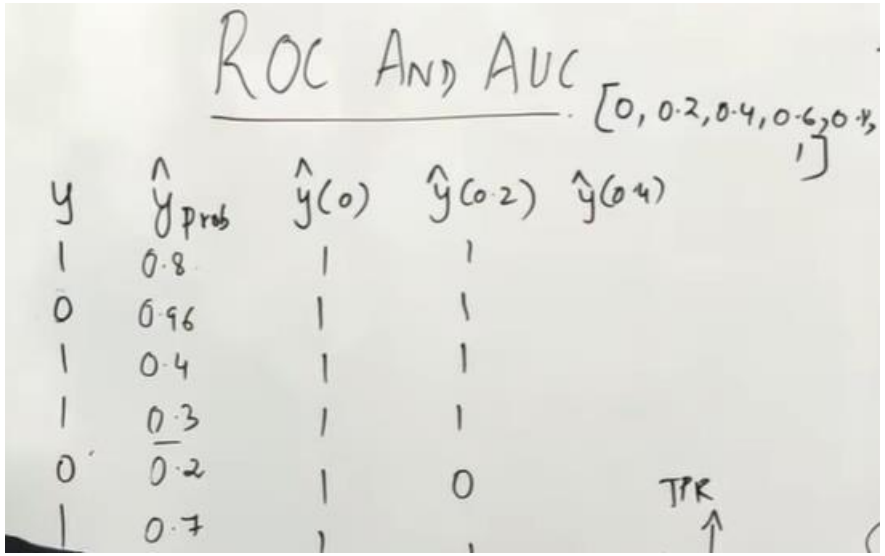
Metrics In Classification

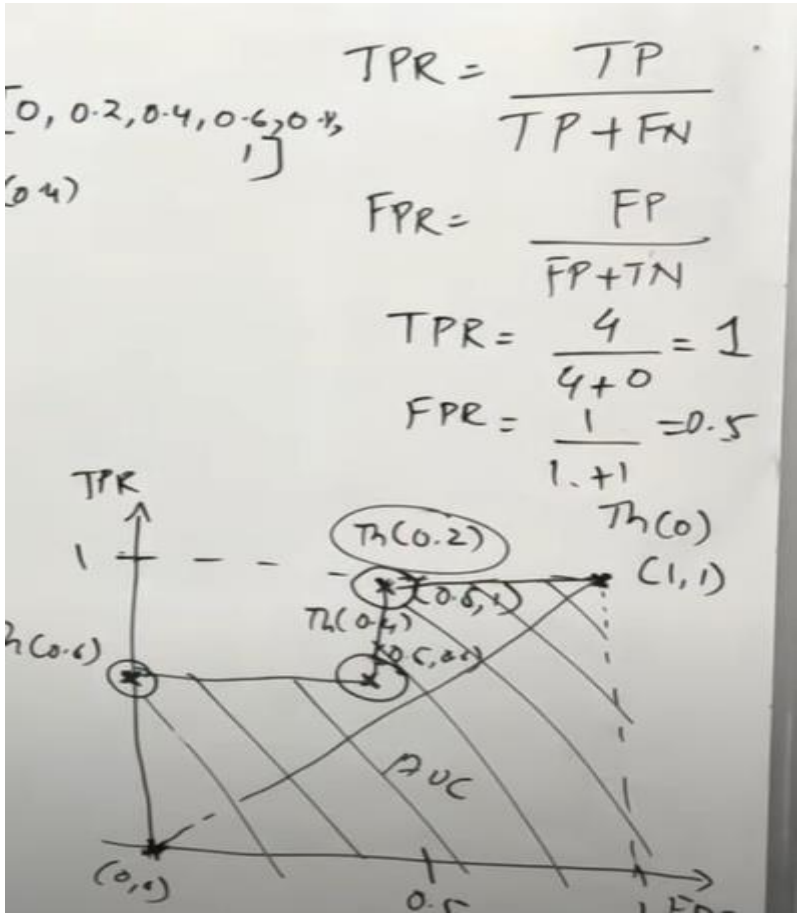




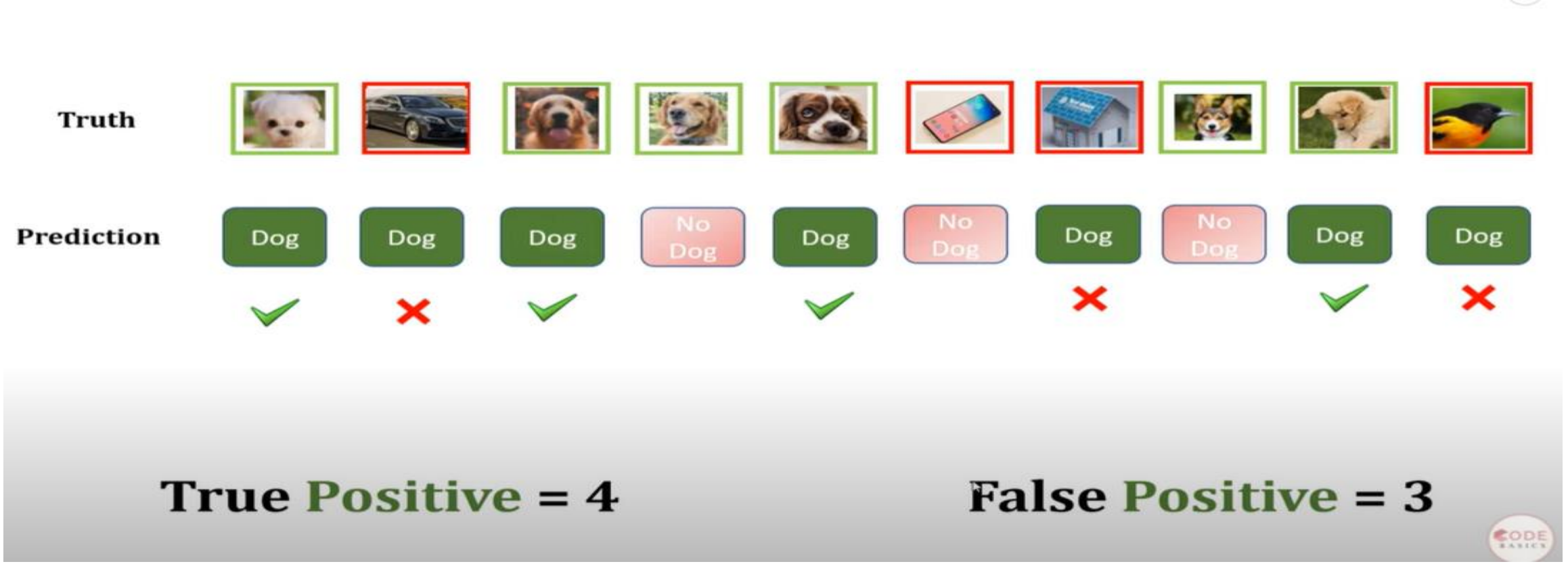


Metrics In Classification

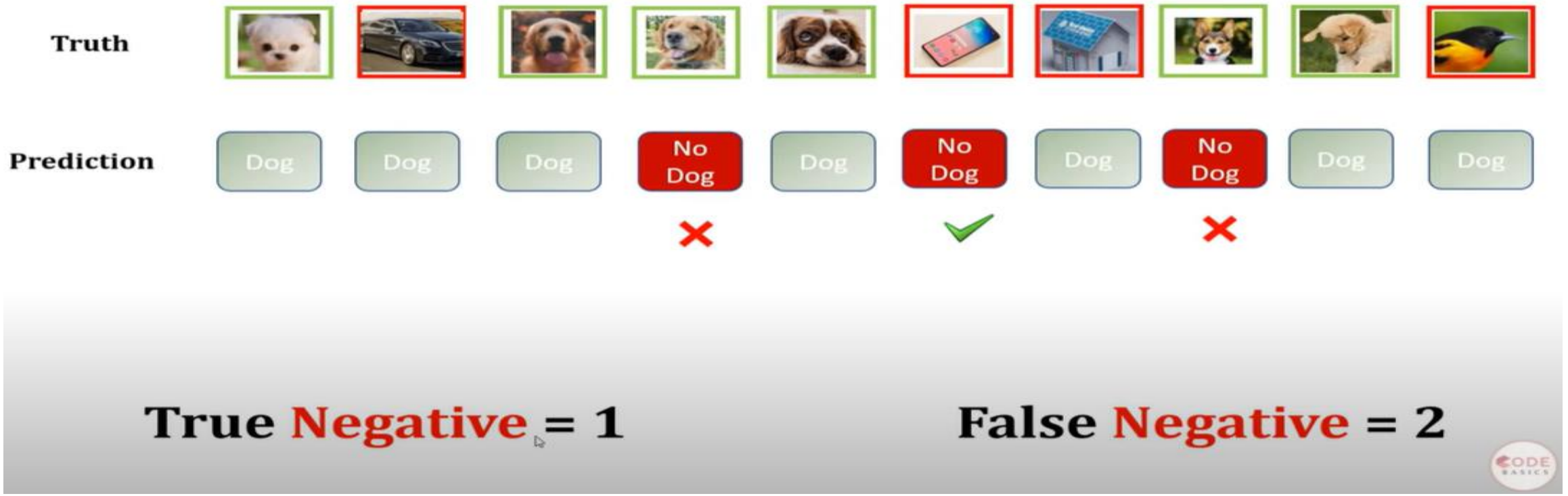




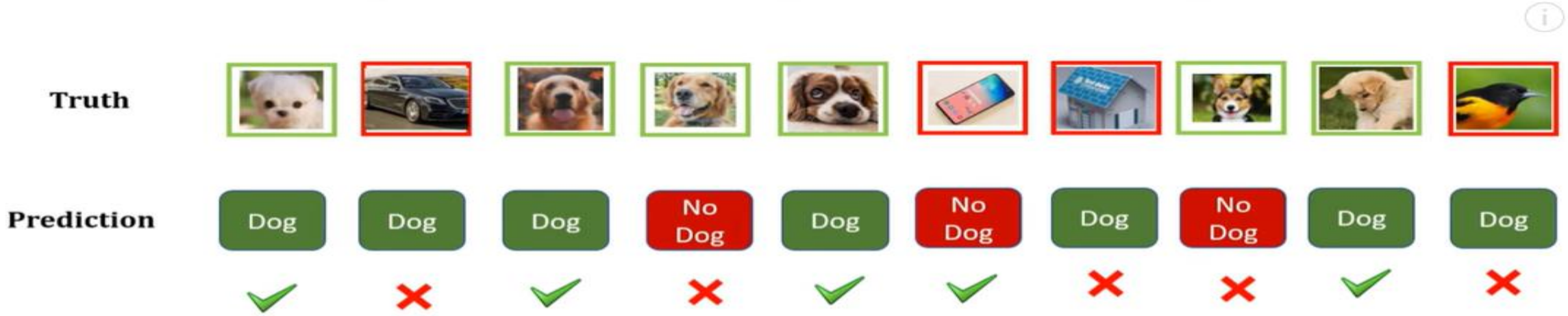
Precision,Recall,F1 Score



Precision,Recall,F1 Score



Precision,Recall,F1 Score



How many we got right? → 5

Accuracy → 5/10 → **0.5**

Precision,Recall,F1 Score

Precision,Recall,F1 Score

Precision,Recall,F1 Score

For precision, think about predictions as your base

For recall, think about truth as your base

Precision,Recall,F1 Score

F1 Score

- F1=2*(Precision*Recall)/(precision+recall)

Hands-on session

- Linear Regression
- Logistic Regression

References

- https://www.skillbasics.com/
- 3Blue1BrownSeries
- https://www.youtube.com/watch?v=IHZwWFHWa-w


Thank You