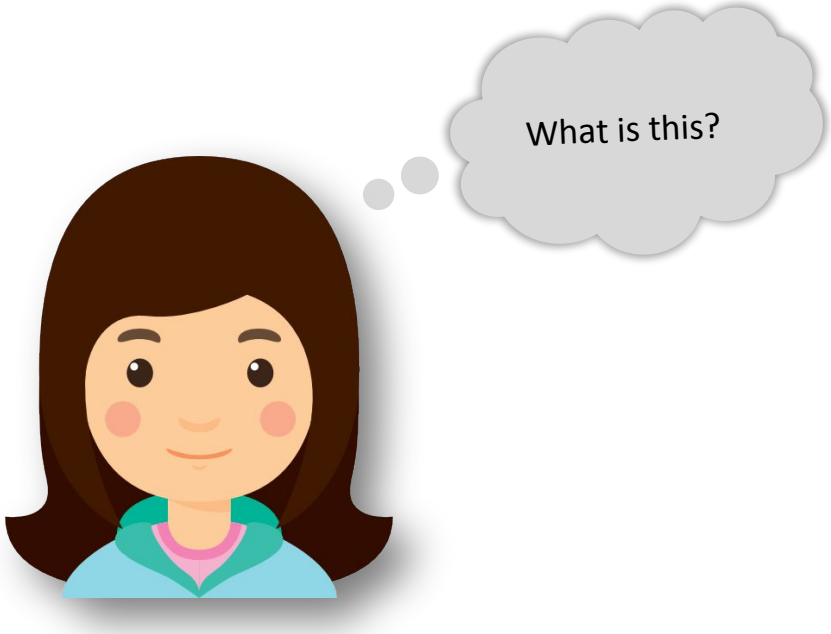
Agenda

greatlearning
Learning for Life

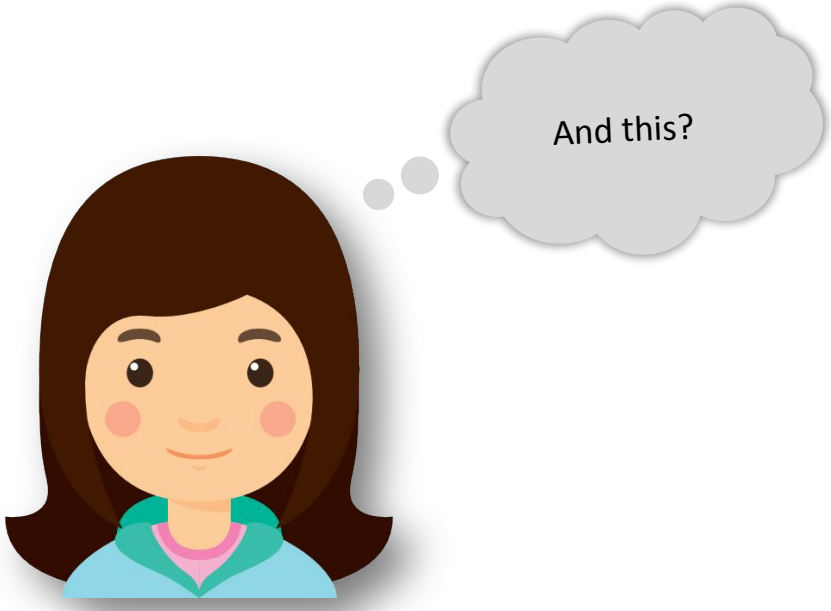
1. Intro to Machine Learning
2. Linear Regression
3. Logistic Regression & Decision Tree
4. Fake News Detection

What is Machine Learning?

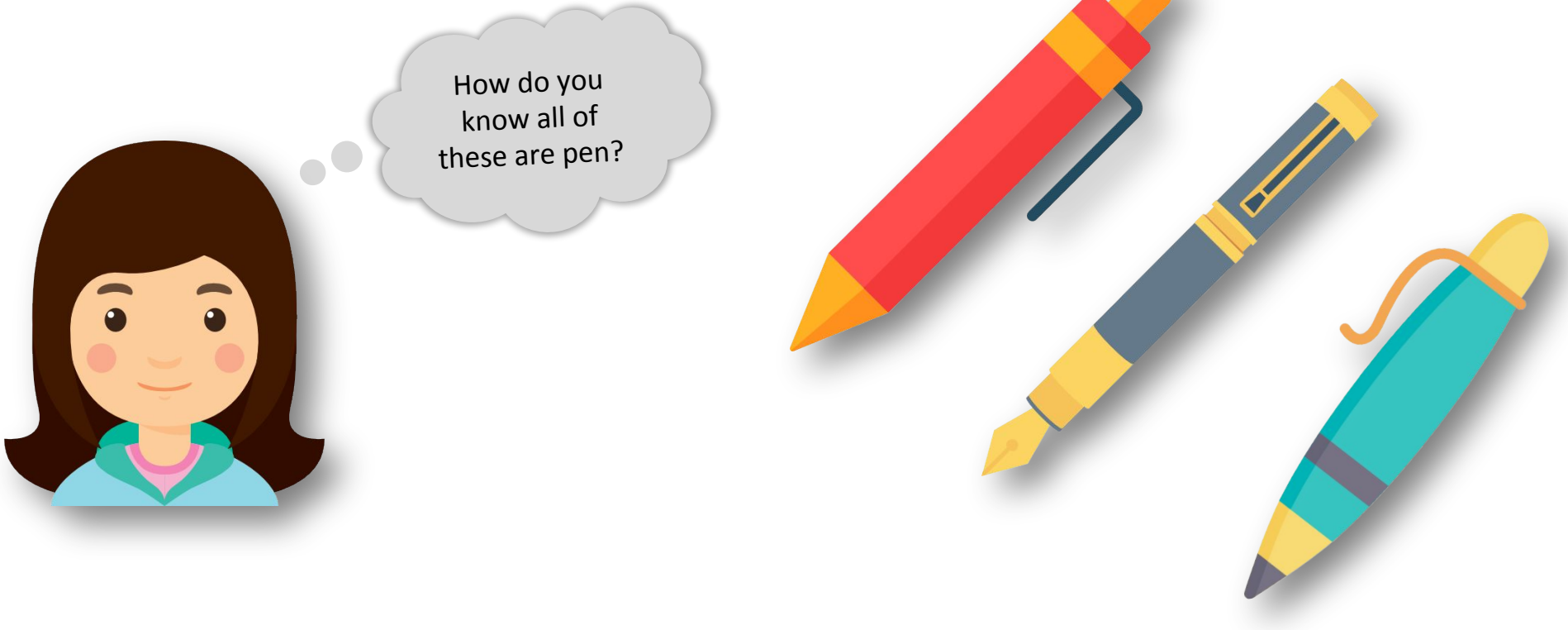


What is Machine Learning?

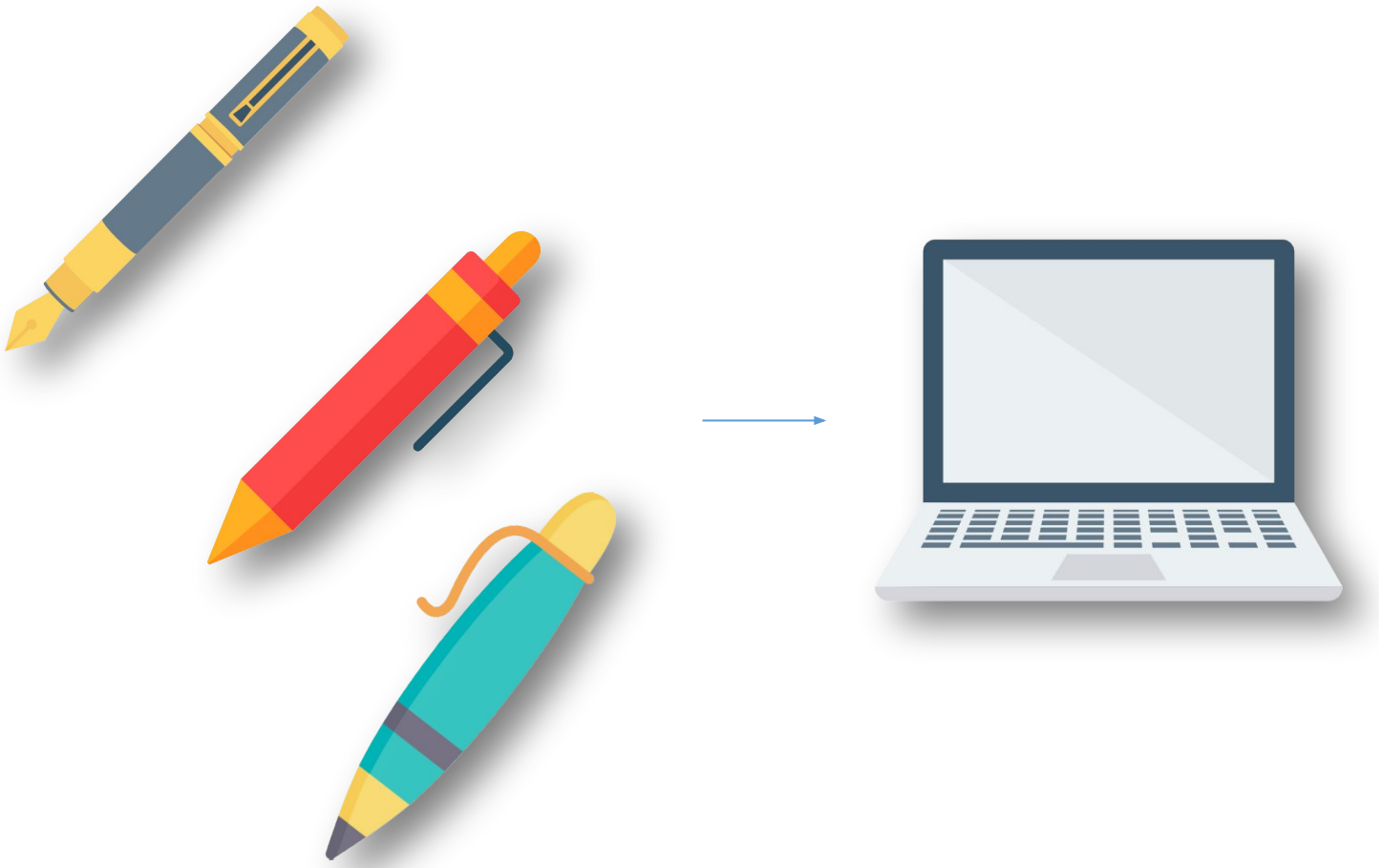
What is Machine Learning?



What is Machine Learning?

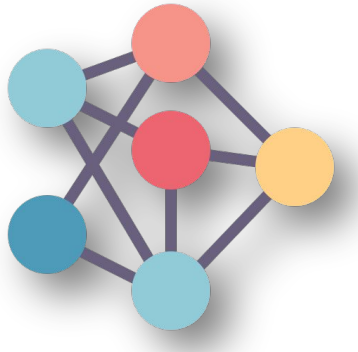


What is Machine Learning?



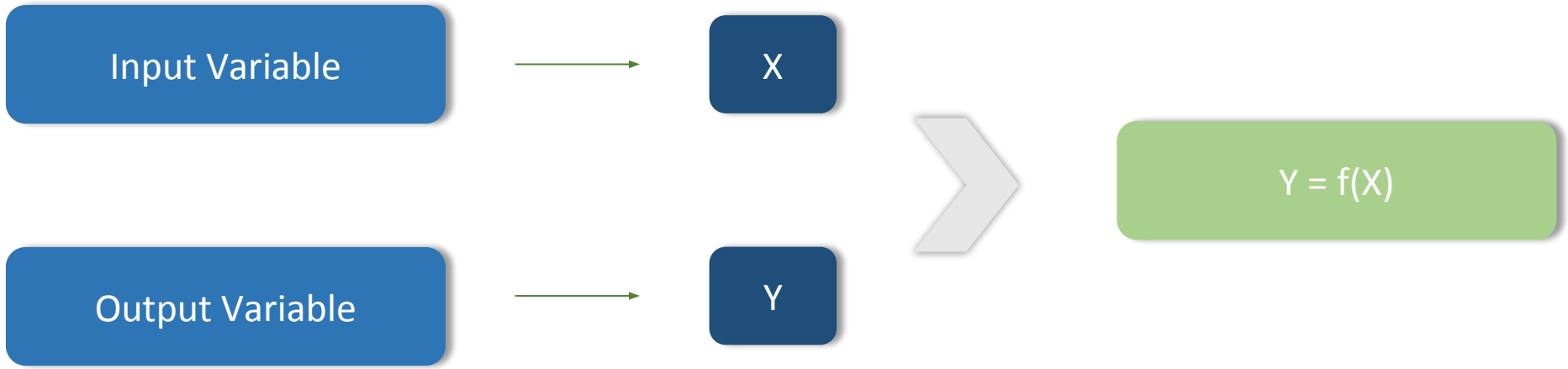
What is Machine Learning?

Categories of Machine Learning

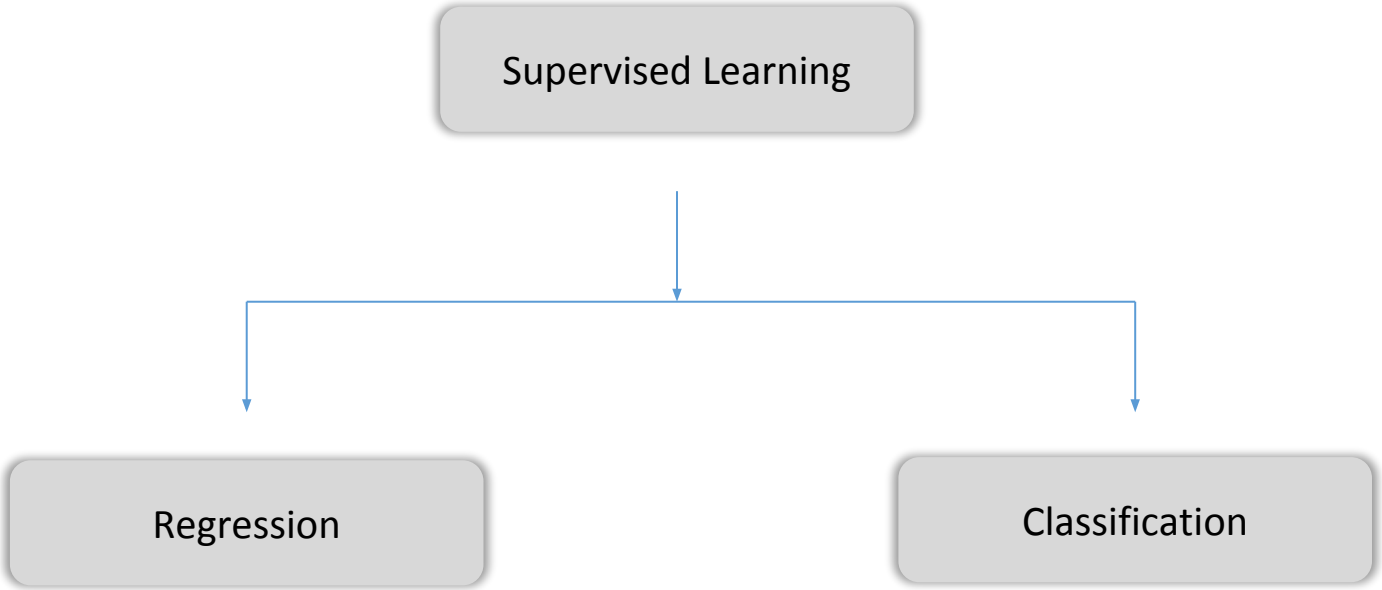


- Supervised Learning
- Unsupervised Learning

Supervised Learning

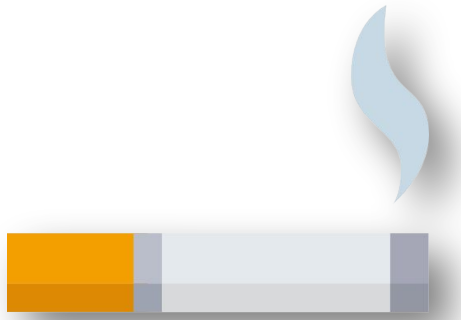


Categories of Supervised Learning



Classification

Classification is the process of predicting the class of a new variable



Smoke(Yes/No)

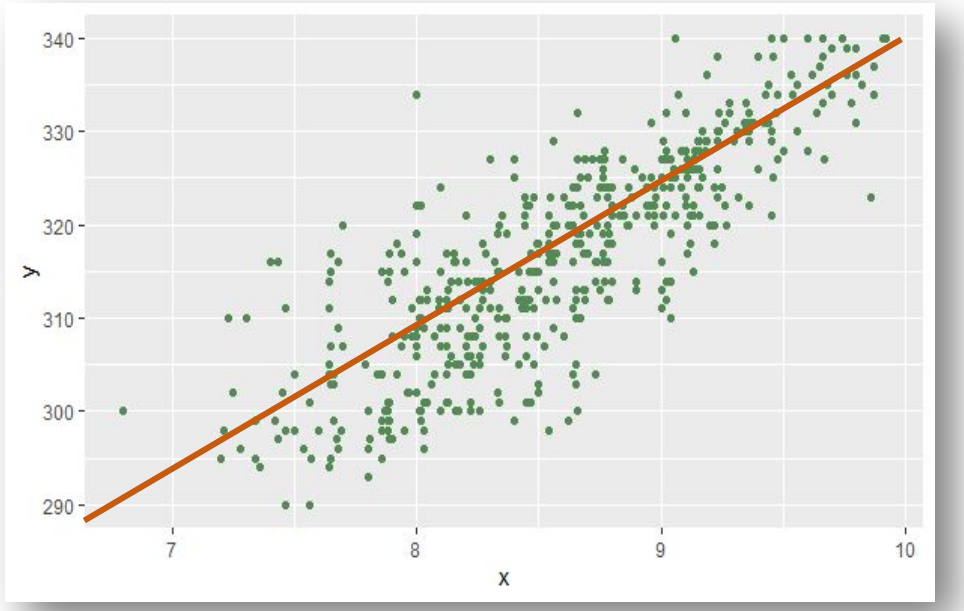
Cancer(Yes/No)

Regression

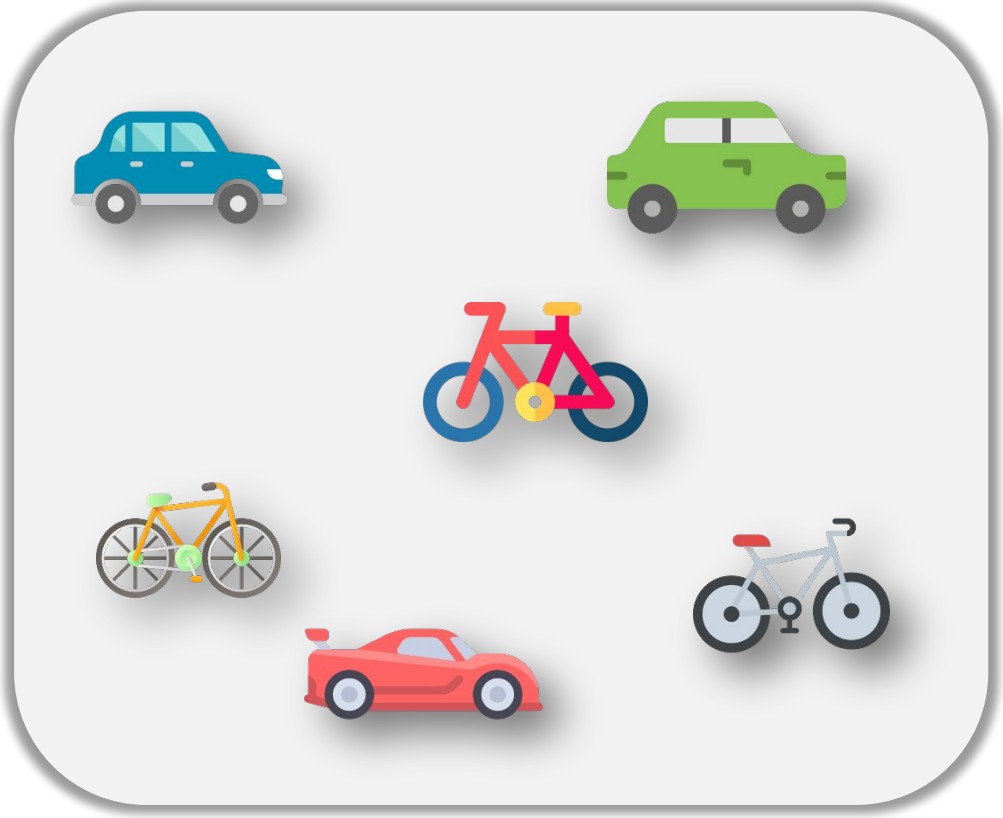
This method is used to estimate the relationship between different entities

Dependent Variable ← → Independent Variable

$$Y=f(x)$$

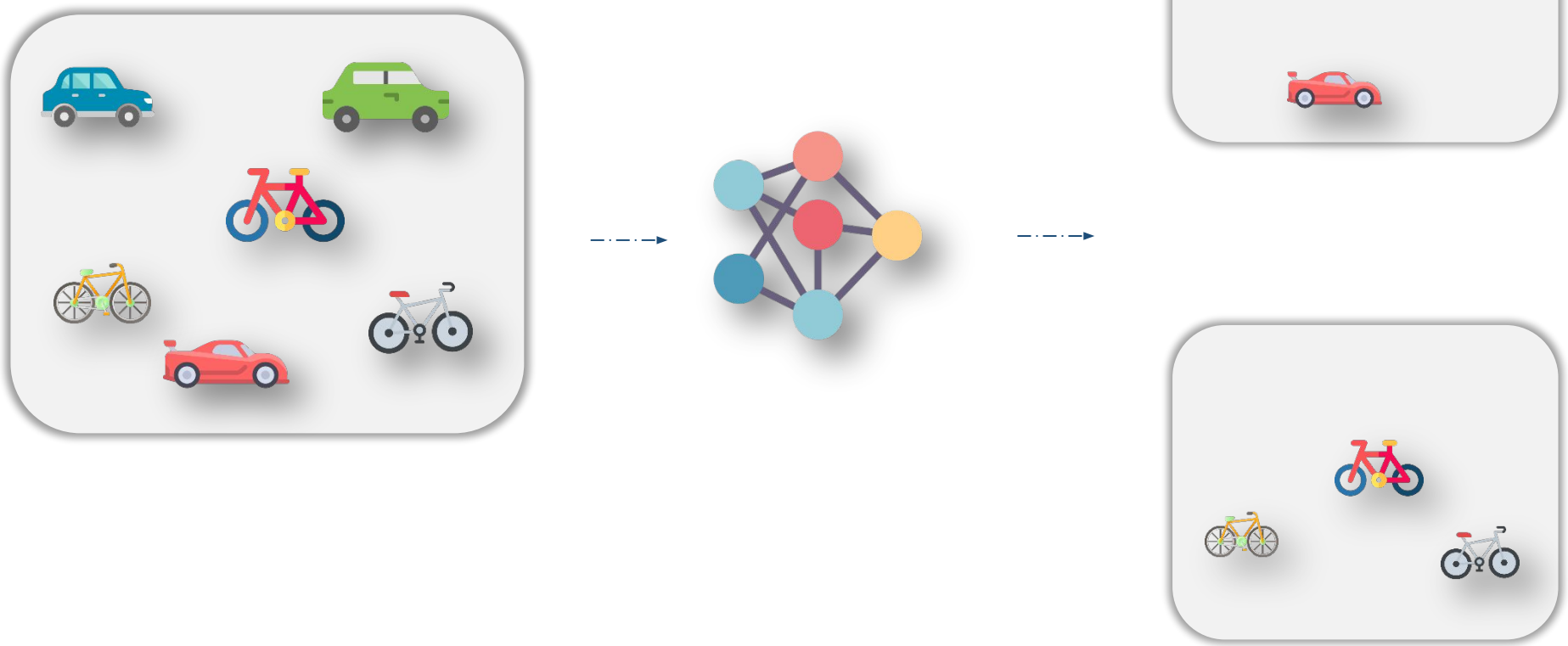


Unsupervised Learning (Clustering)

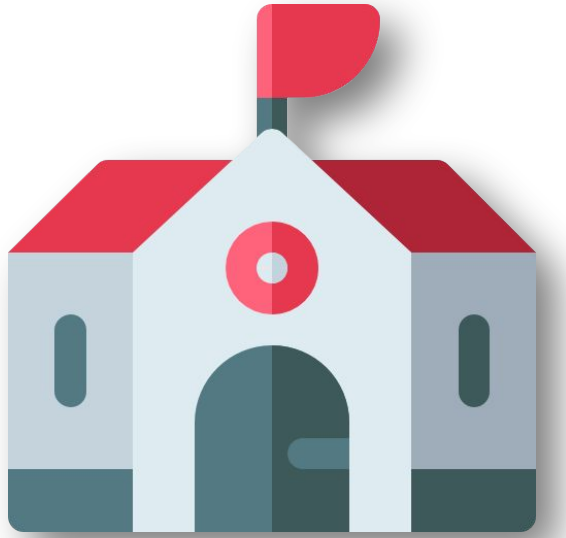


Input Data with no class labels

Unsupervised Learning (Clustering)



You are conducting a case-study on a set of college students to understand if students with high CGPA also get a high GRE score



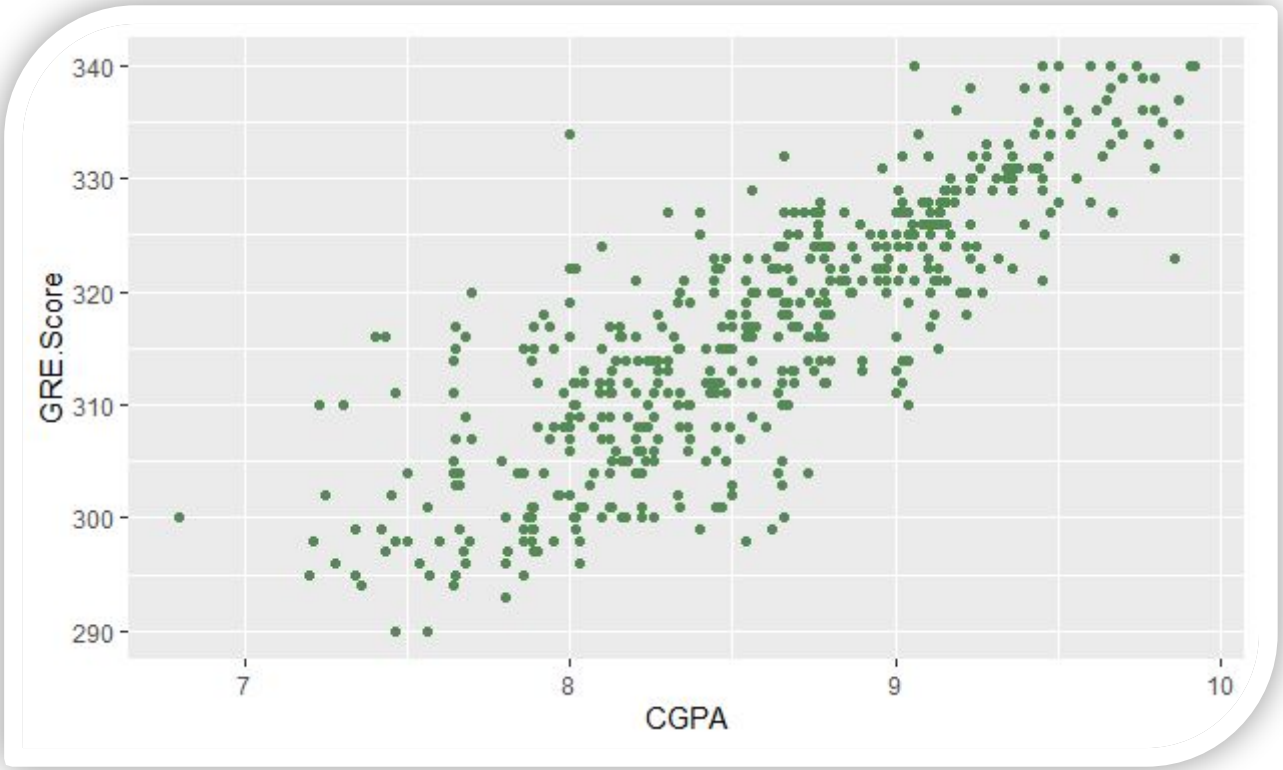
Linear Regression

GRE Score	CGPA
337	9.65
324	8.87
316	8.00
322	8.67
314	8.21
330	9.34
321	8.20
308	7.90
302	8.00
323	8.60

Case Study

greatlearning
Learning for Life

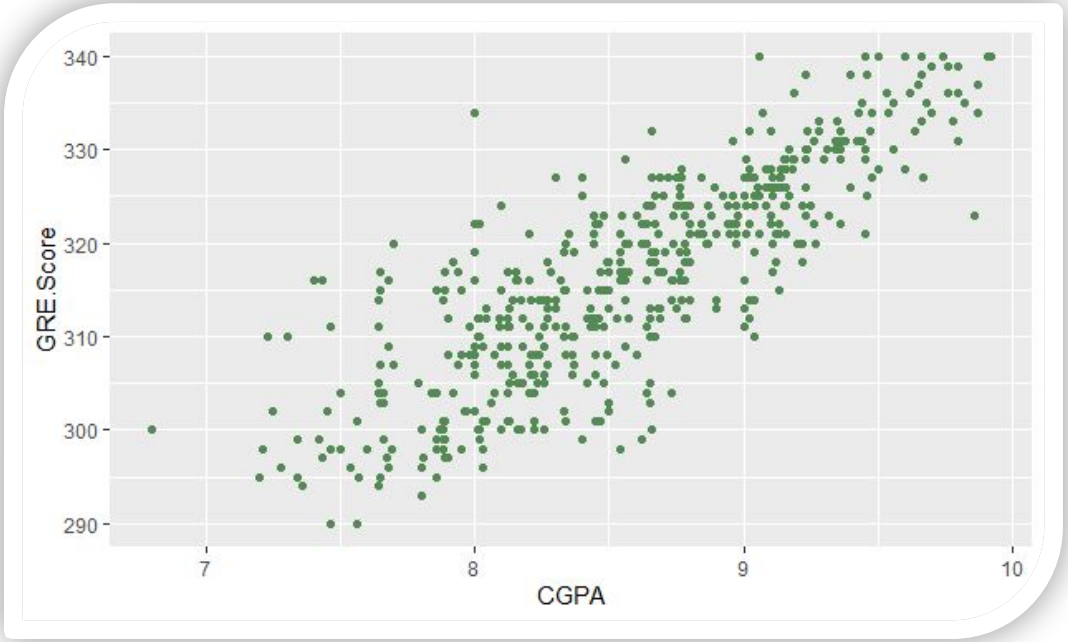
GRE.Score	CGPA
337	9.65
324	8.87
316	8.00
322	8.67
314	8.21
330	9.34
321	8.20
308	7.90
302	8.00
323	8.60



Y X

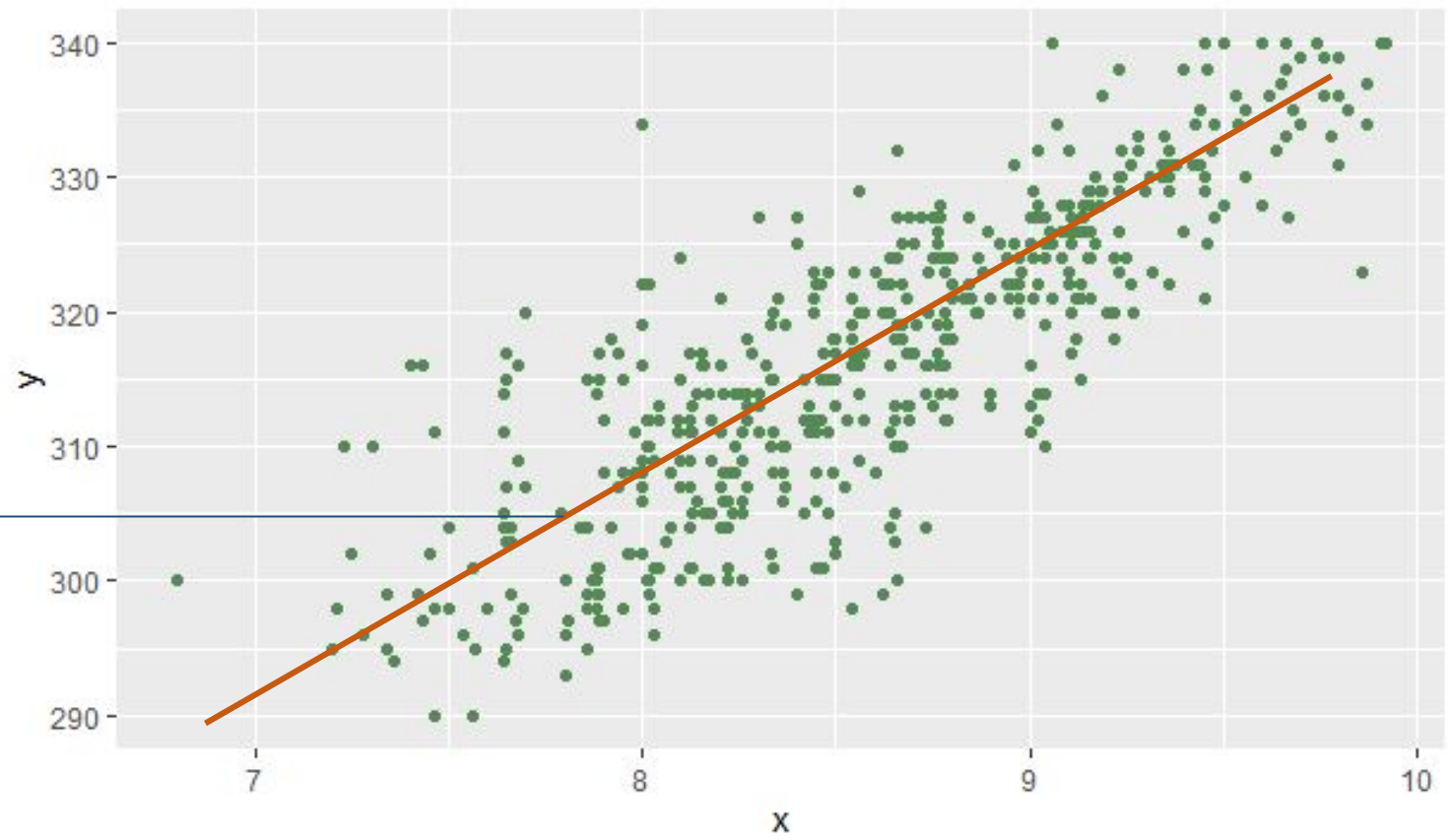
Need of Regression Analysis

What would be the GRE Score of a student if his CGPA is 8.32?

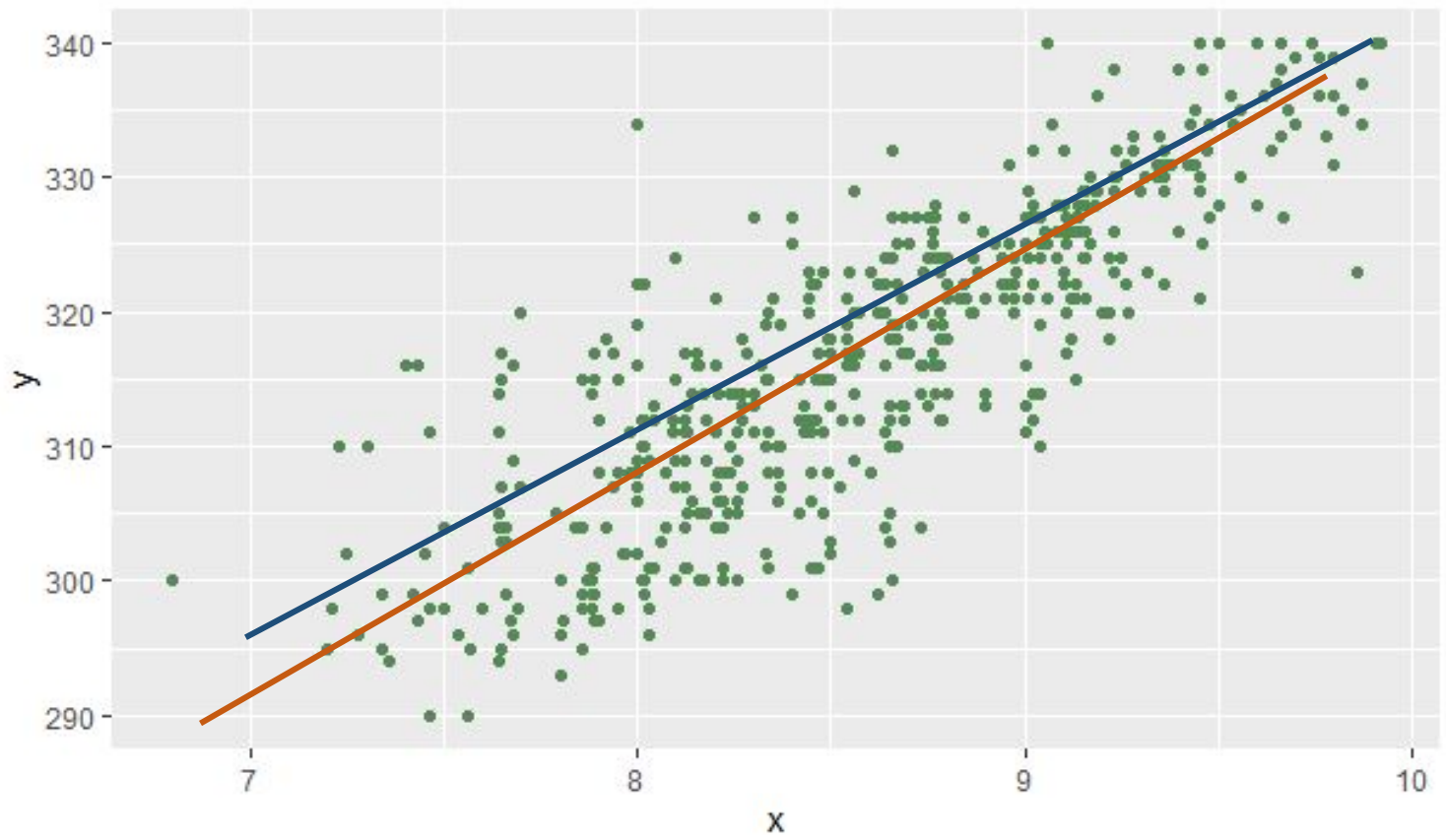


Best Fit Line

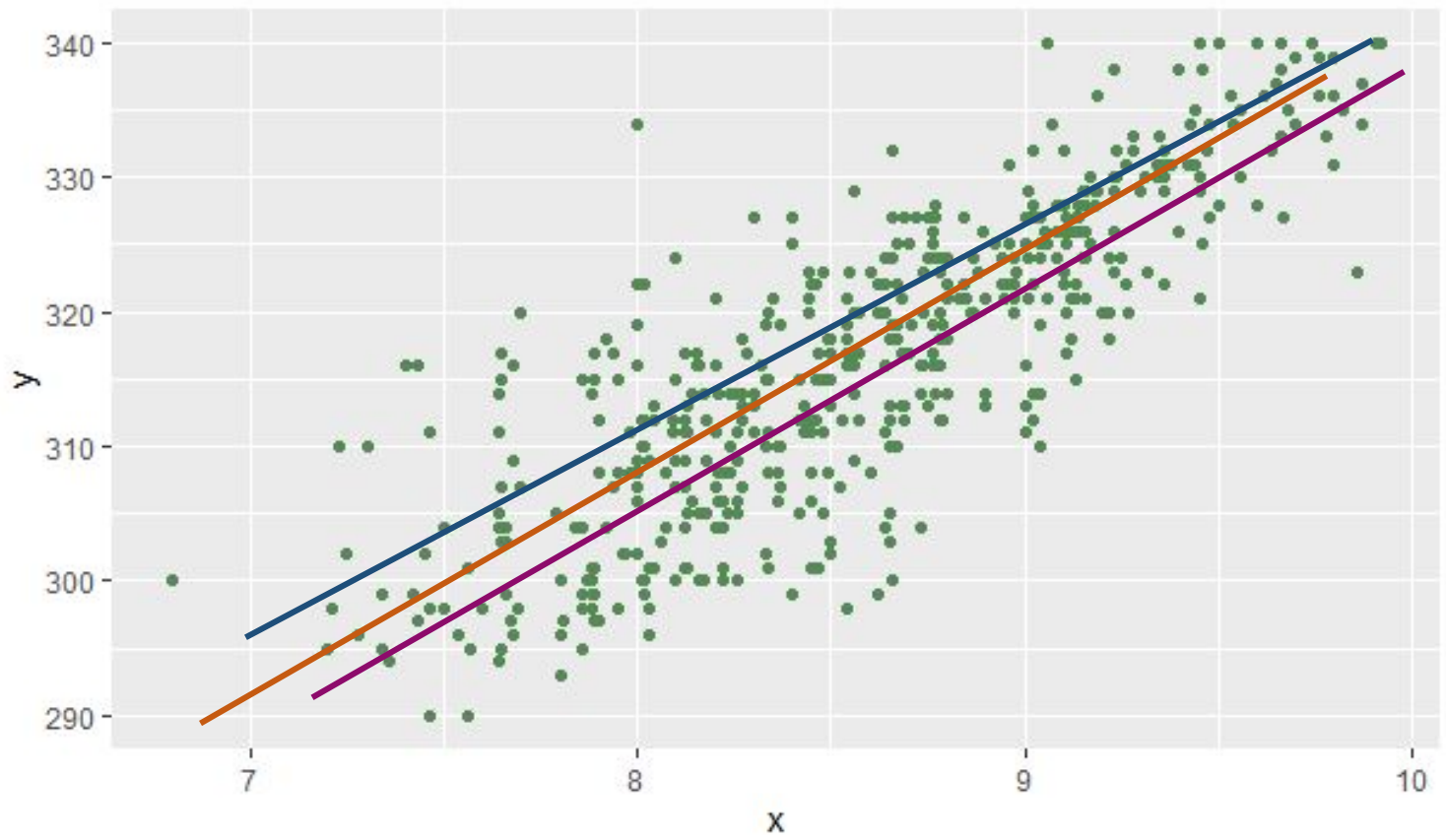
$Y = a + bx$



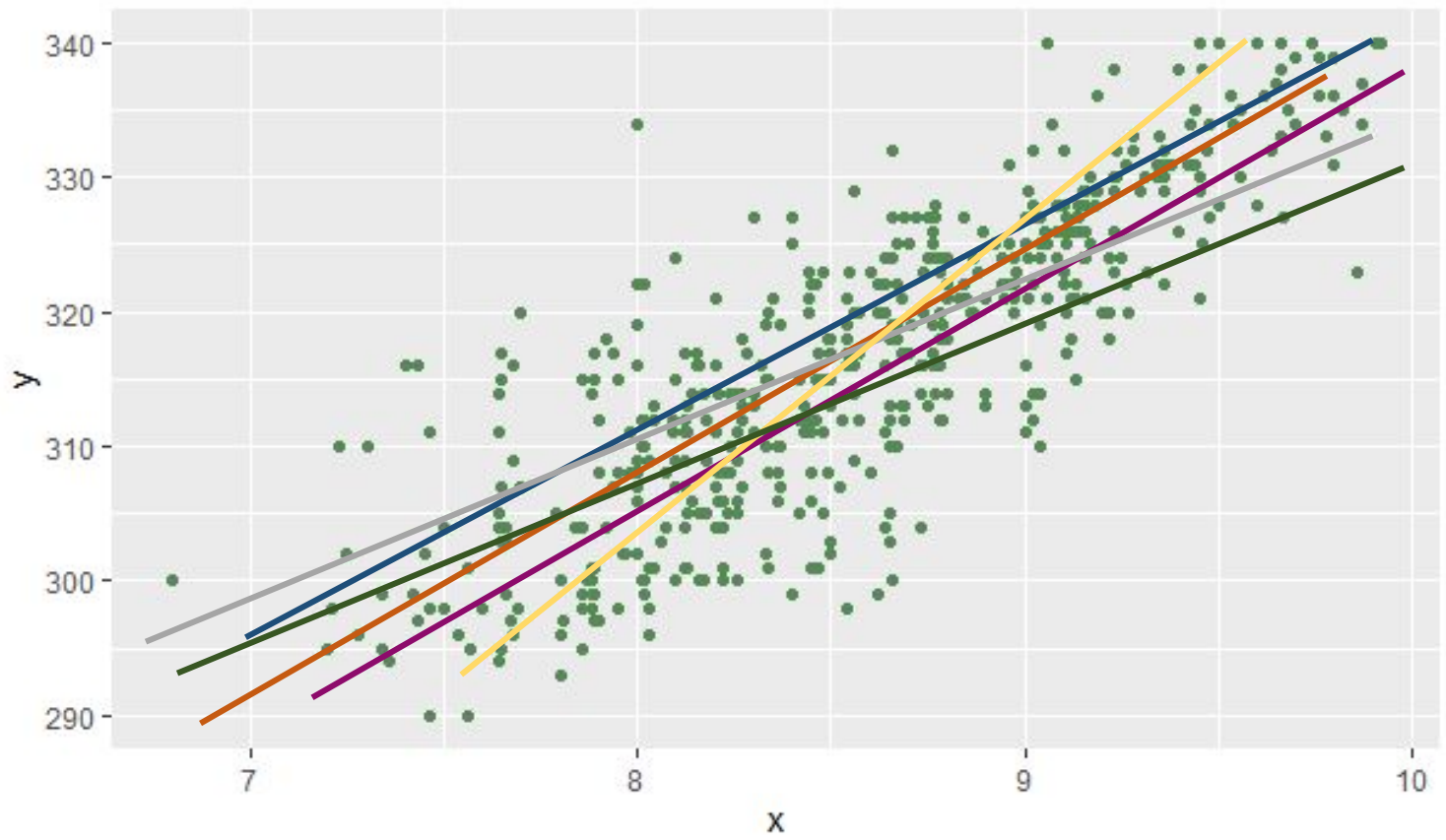
Best Fit Line



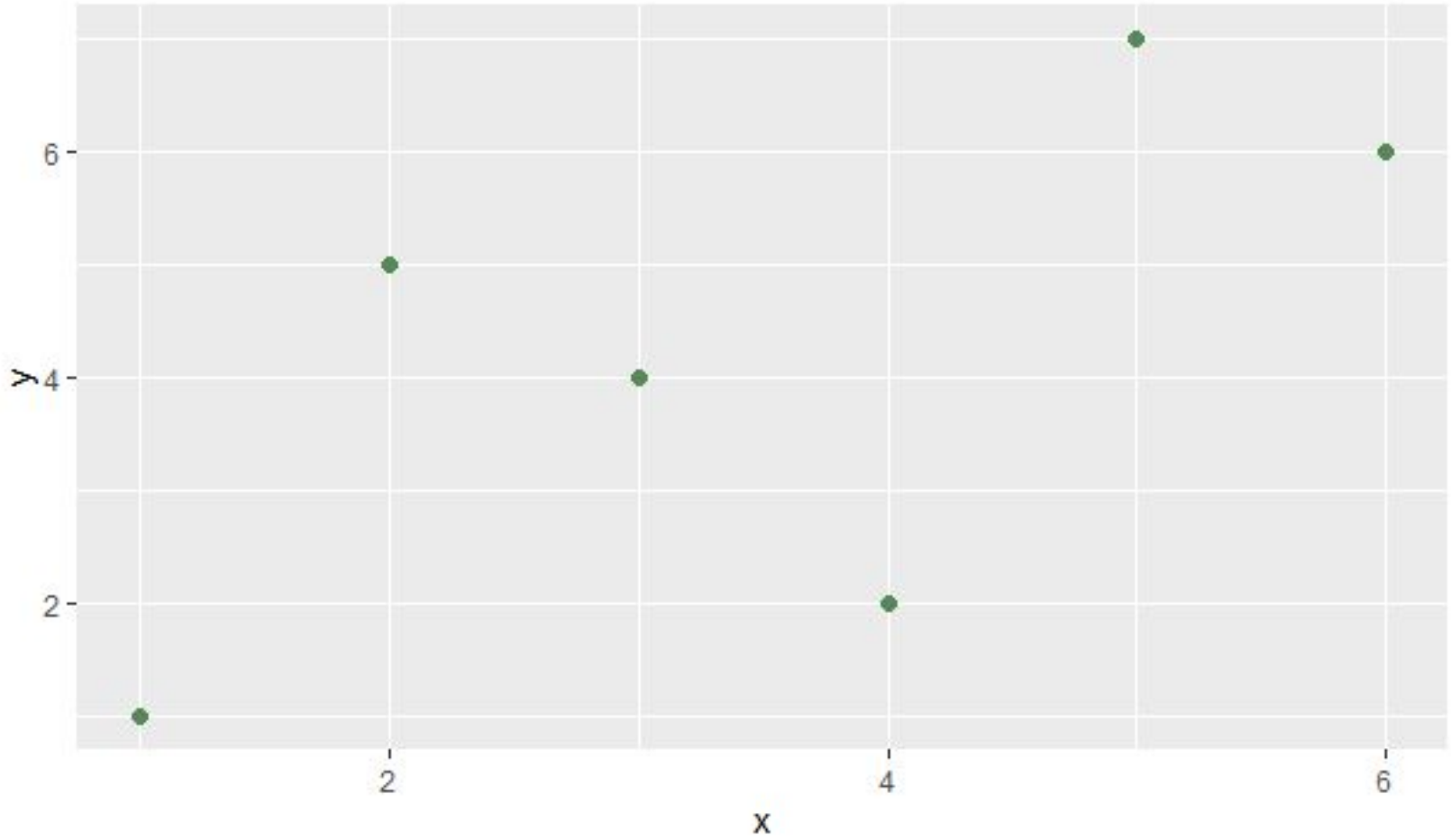
Best Fit Line



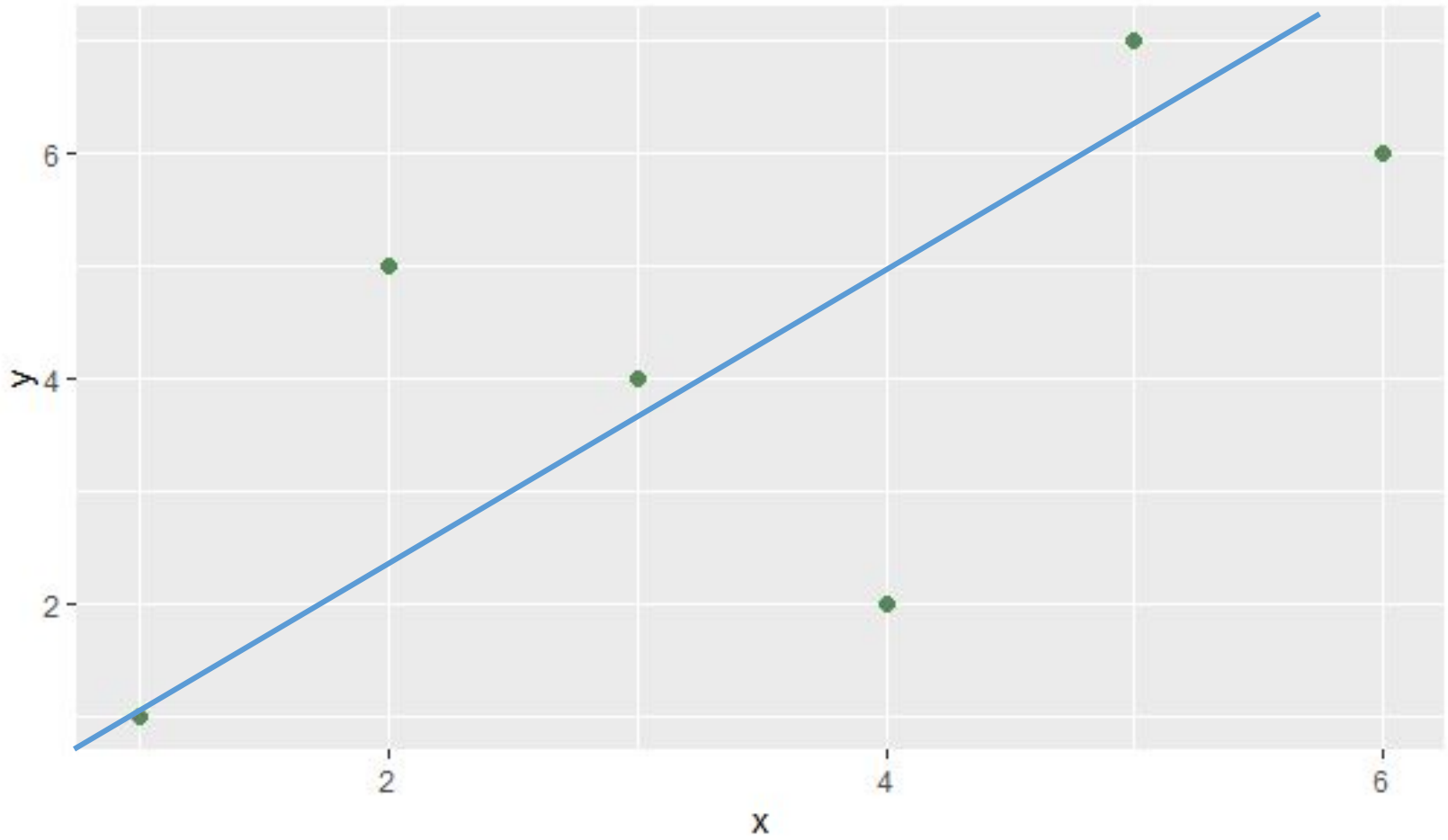
Best Fit Line



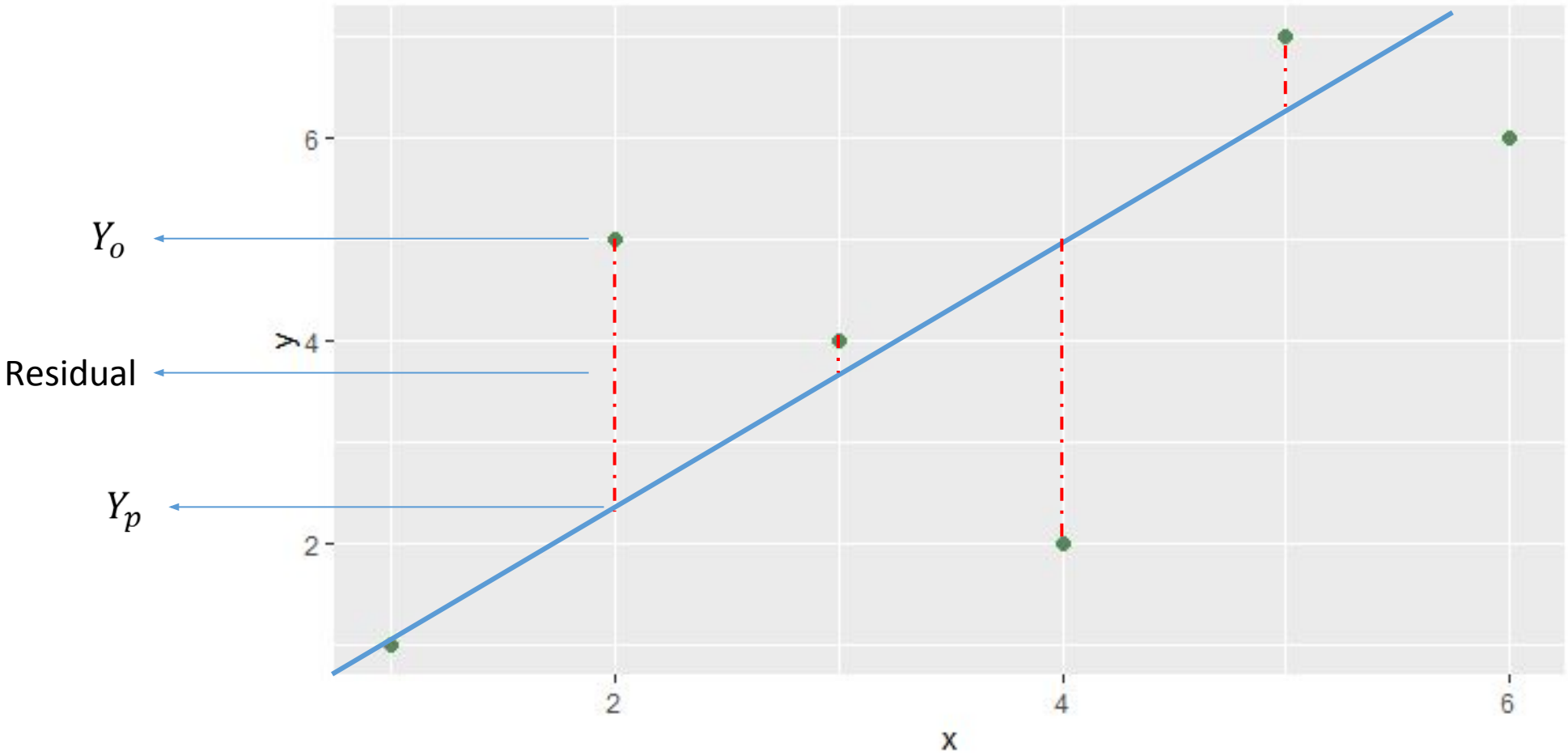
Measure of Good Fit



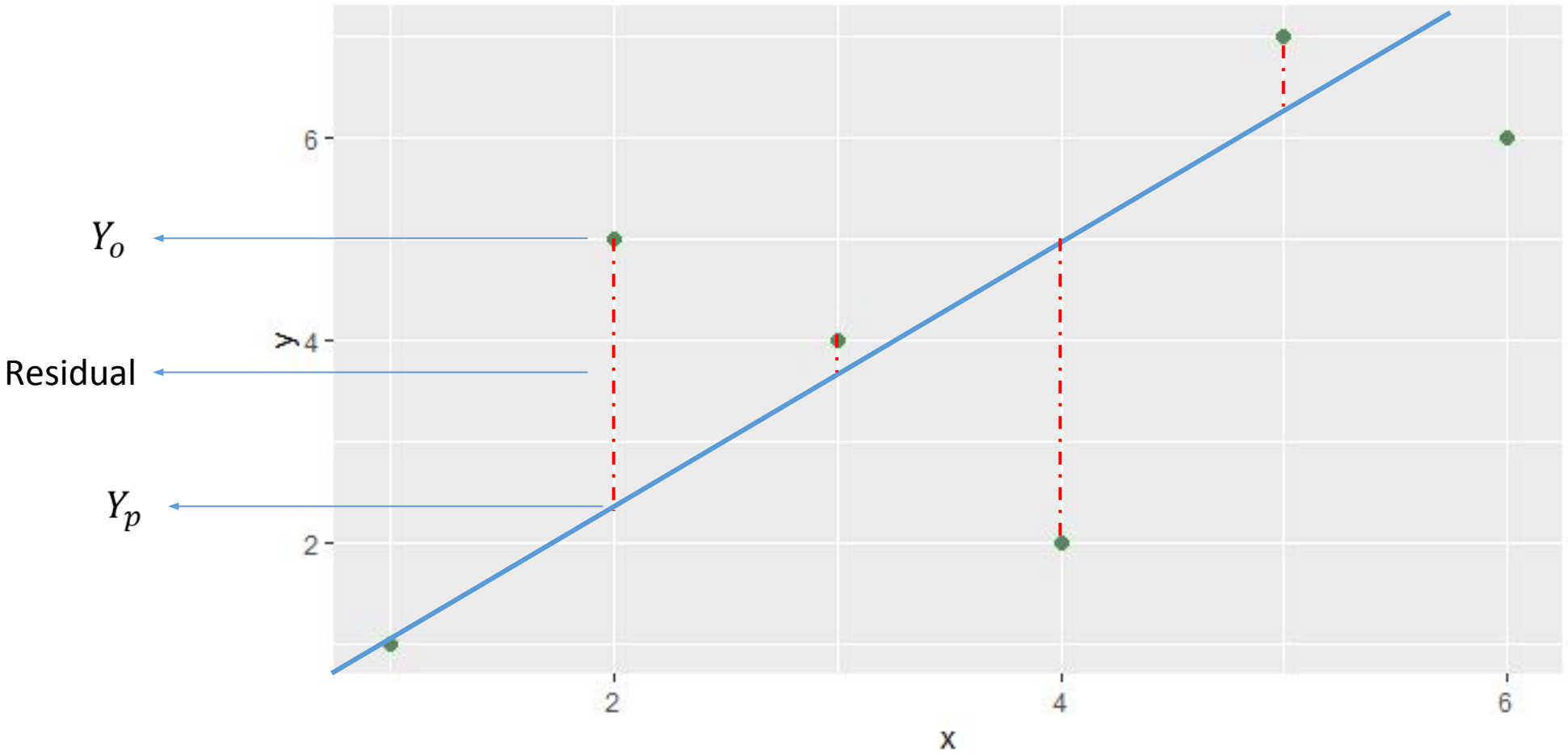
Measure of Good Fit



Measure of Good Fit

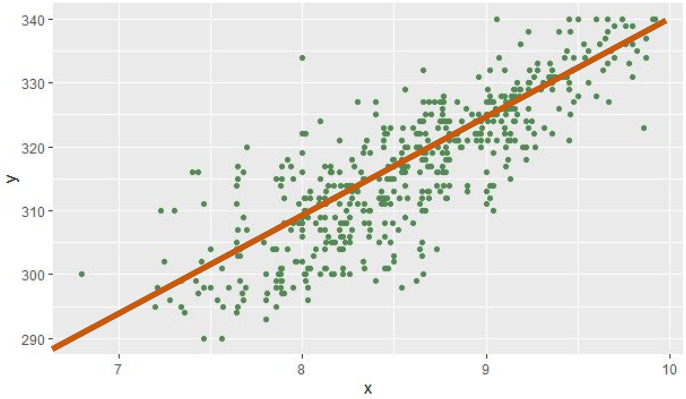


Measure of Good Fit

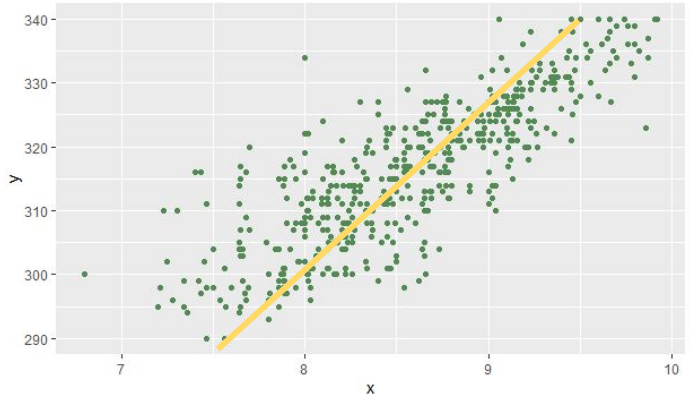


$$\sum (Y_0 - Y_P)^2$$

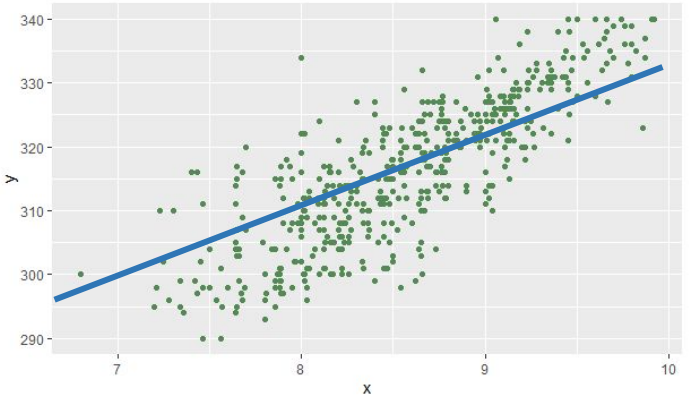
Measure of Good Fit



$\sum(Y_0 - Y_P)^2 = 28$



$\sum(Y_0 - Y_P)^2 = 22$

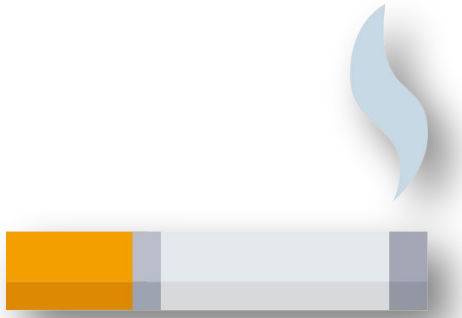


$\sum(Y_0 - Y_P)^2 = 24$

Best Fit Line

Logistic Regression

Logistic regression is used for classification

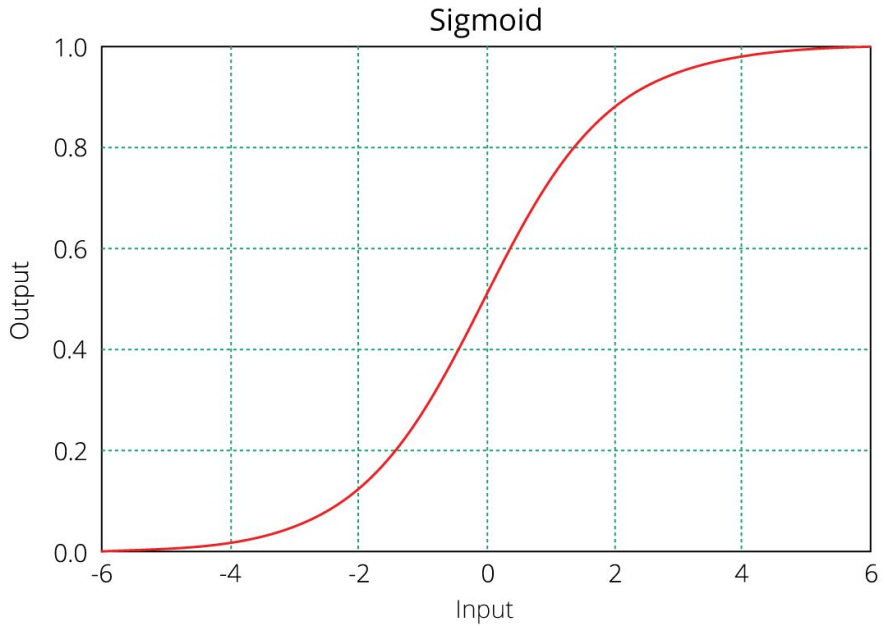


Smoke(Yes/No)

Cancer(Yes/No)

Logistic Regression

This link function follows a sigmoid function which limits its range of probabilities between 0 and 1.



Decision Tree

Decision Tree is used for both classification and Regression