Regression

Mr. C. R. Barde

Regressions

Simple Linear Regression

$$y = b_0 + b_1 * x_1$$

Dependent variable (DV)   Independent variable (IV)

Simple Linear Regression

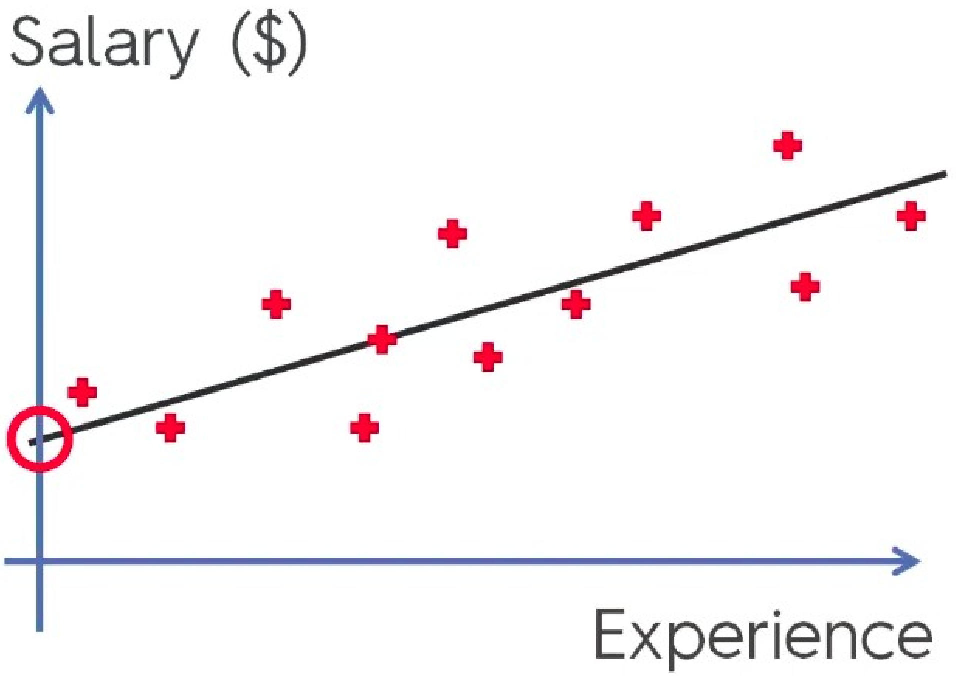
Constant   Coefficient

$$y = b_0 + b_1{*}x_1$$

Dependent variable (DV)   Independent variable (IV)

Regressions

Simple Linear Regression:

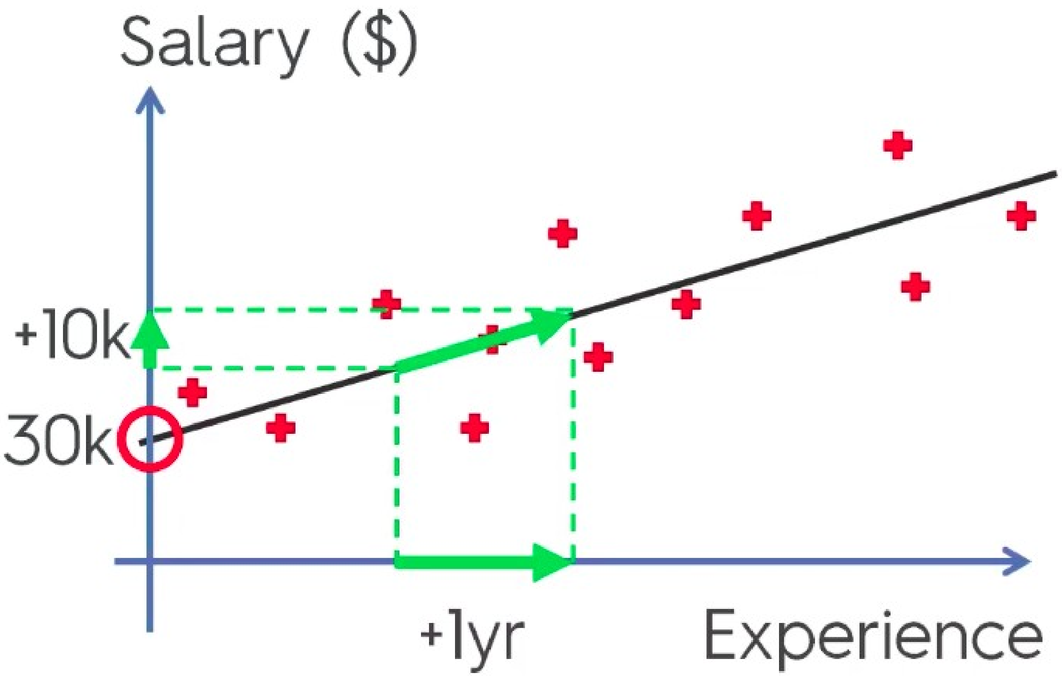


$$y = b_0 + b_1 * x$$

$$\text{Salary} = b_0 + b_1 * \text{Experience}$$

Regressions

Simple Linear Regression:

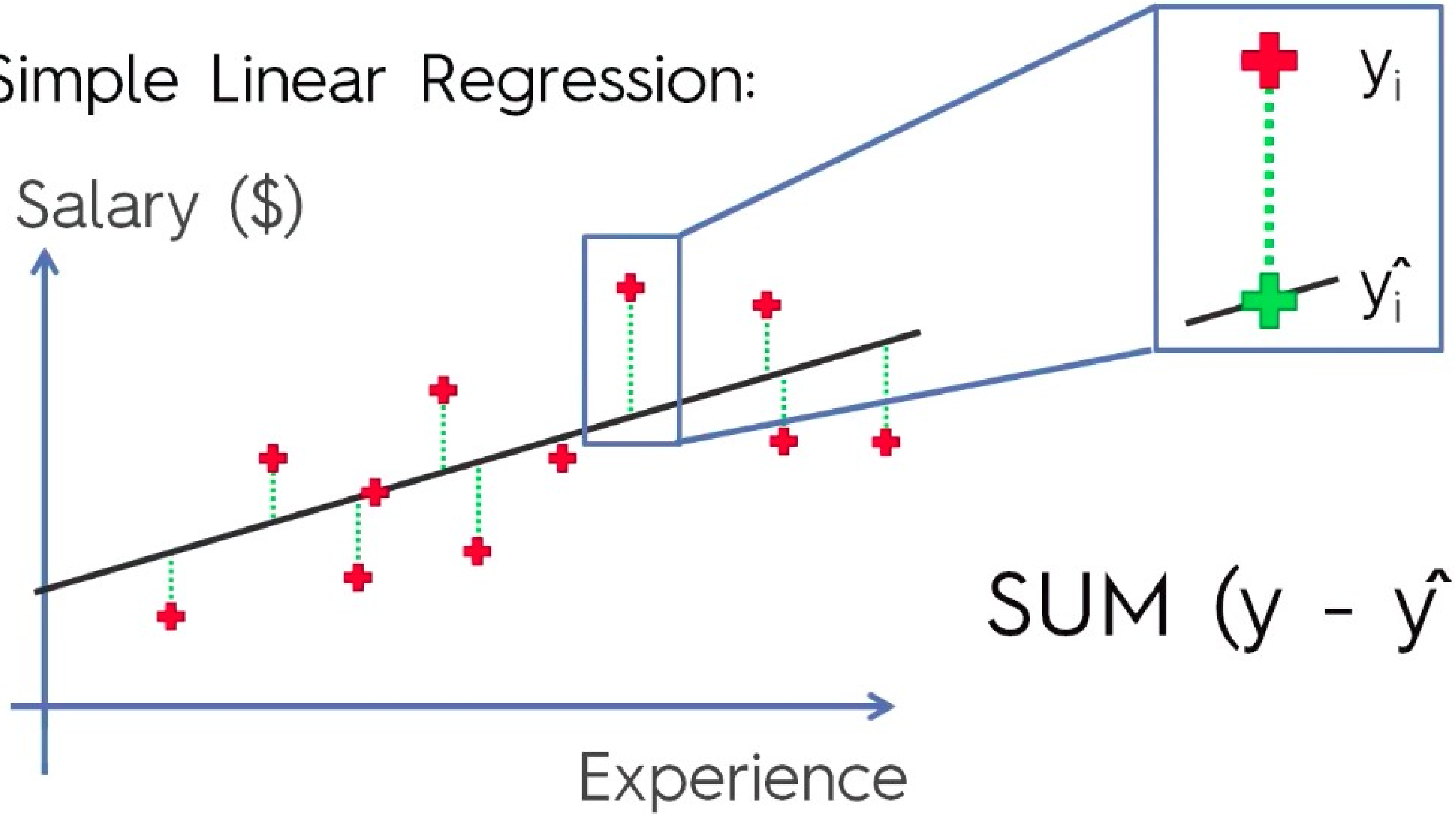


$y = b_0 + b_1 * x$

$Salary = b_0 + b_1 * Experience$

Ordinary Least Squares

Simple Linear Regression:



SUM $(y - \hat{y})^2$ -> min

Regressions

Simple Linear Regression

$$y = b_0 + b_1{}^*x_1$$

Multiple Linear Regression

$$y = b_0 + b_1{}^*x_1 + b_2{}^*x_2 + \ldots + b_n{}^*x_n$$

Regressions

Simple Linear Regression

$$y = b_0 + b_1{}^*x_1$$

Multiple Linear Regression

Dependent variable (DV)   Independent variables (IVs)

$$y = b_0 + b_1{}^*x_1 + b_2{}^*x_2 + \ldots + b_n{}^*x_n$$

Dummy Variables

Profit	R&D Spend	Admin	Marketing	State
192,261.83	165,349.20	136,897.80	471,784.10	New York
191,792.06	162,597.70	151,377.59	443,898.53	California
191,050.39	153,441.51	101,145.55	407,934.54	California
182,901.99	144,372.41	118,671.85	383,199.62	New York
166,187.94	142,107.34	91,391.77	366,168.42	California

Dummy Variables

Profit	R&D Spend	Admin	Marketing	State
192,261.83	165,349.20	136,897.80	471,784.10	New York
191,792.06	162,597.70	151,377.59	443,898.53	California
191,050.39	153,441.51	101,145.55	407,934.54	California
182,901.99	144,372.41	118,671.85	383,199.62	New York
166,187.94	142,107.34	91,391.77	366,168.42	California

Dummy Variables

New York	California
1	0
0	1
0	1
1	0
0	1

$$y = b_0 + b_1{}^*x_1 + b_2{}^*x_2 + b_3{}^*x_3 \qquad + b_4{}^*D_1$$

Dummy Variable Trap

Profit	R&D Spend	Admin	Marketing	State
192,261.83	165,349.20	136,897.80	471,784.10	New York
191,792.06	162,597.70	151,377.59	443,898.53	California
191,050.39	153,441.51	101,145.55	407,934.54	California
182,901.99	144,372.41	118,671.85	383,199.62	New York
166,187.94	142,107.34	91,391.77	366,168.42	California

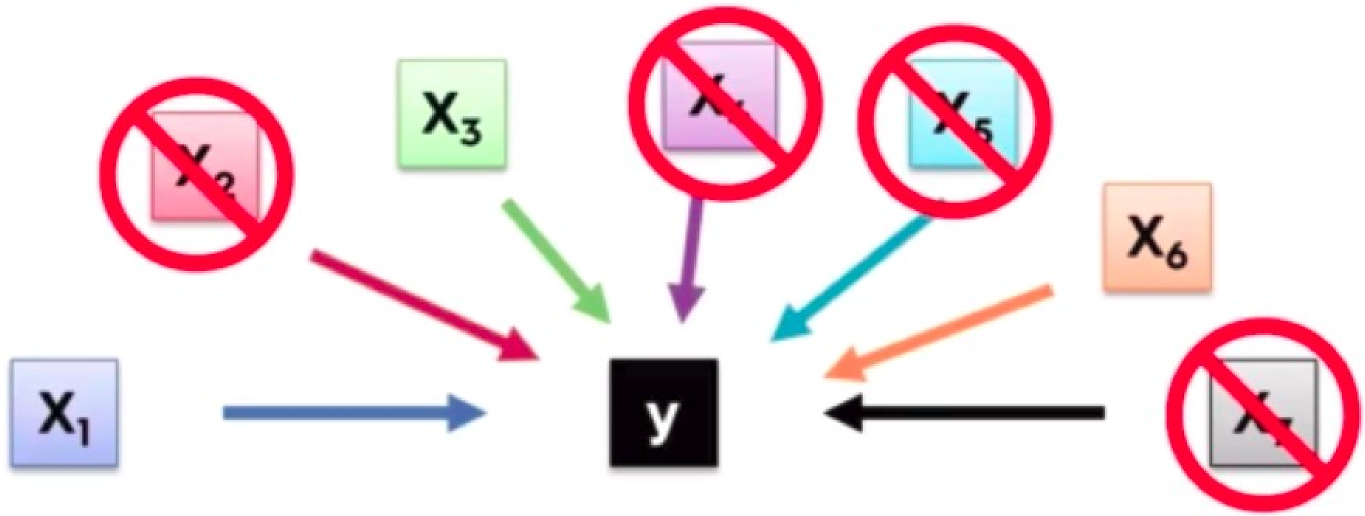
Dummy Variables

New York	California
1	0
0	1
0	1
1	0
0	1

$$y = b_0 + b_1{*}x_1 + b_2{*}x_2 + b_3{*}x_3 \qquad + b_4{*}D_1 + \cancel{b_5{*}D_2}$$

Always omit one dummy variable

Building A Model



Why?

Building A Model

5 methods of building models:

1. All-in
2. Backward Elimination
3. Forward Selection
4. Bidirectional Elimination
5. Score Comparison

Stepwise Regression

Building A Model

Backward Elimination

STEP 1: Select a significance level to stay in the model (e.g. SL = 0.05)

STEP 2: Fit the full model with all possible predictors

STEP 3: Consider the predictor with the <u>highest</u> P-value. If P > SL, go to STEP 4, otherwise go to FIN

STEP 4: Remove the predictor

STEP 5: Fit model without this variable*

Building A Model

Forward Selection

STEP 1: Select a significance level to enter the model (e.g. SL = 0.05)

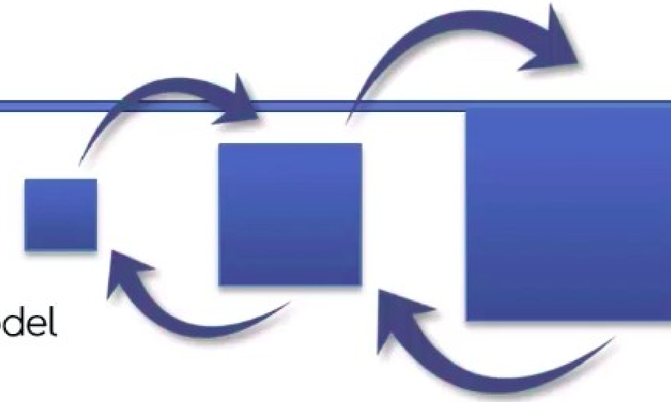
STEP 2: Fit all simple regression models $\mathbf{y} \sim \mathbf{x}_n$ Select the one with the lowest P-value

STEP 3: Keep this variable and fit all possible models with one extra predictor added to the one(s) you already have

STEP 4: Consider the predictor with the lowest P-value. If ~~P < SL,~~ go to STEP 3, otherwise go to FIN

FIN: Keep the previous model

Building A Model



Bidirectional Elimination

STEP 1: Select a significance level to enter and to stay in the model
e.g.: SLENTER = 0.05, SLSTAY = 0.05

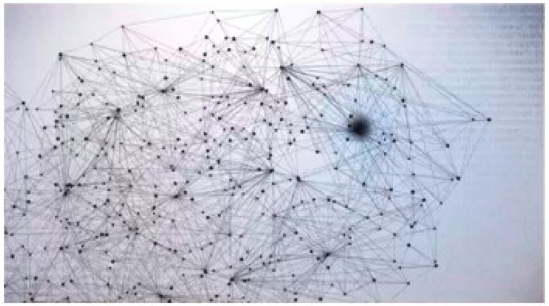
STEP 2: Perform the next step of Forward Selection (new variables must have: P < SLENTER to enter)

STEP 3: Perform ALL steps of Backward Elimination (old variables must have P < SLSTAY to stay)

STEP 4: No new variables can enter and no old variables can exit

FIN: Your Model Is Ready

Building A Model



All Possible Models

STEP 1: Select a criterion of goodness of fit (e.g. Akaike criterion)

STEP 2: Construct All Possible Regression Models: 2^N-1 total combinations

STEP 3: Select the one with the best criterion

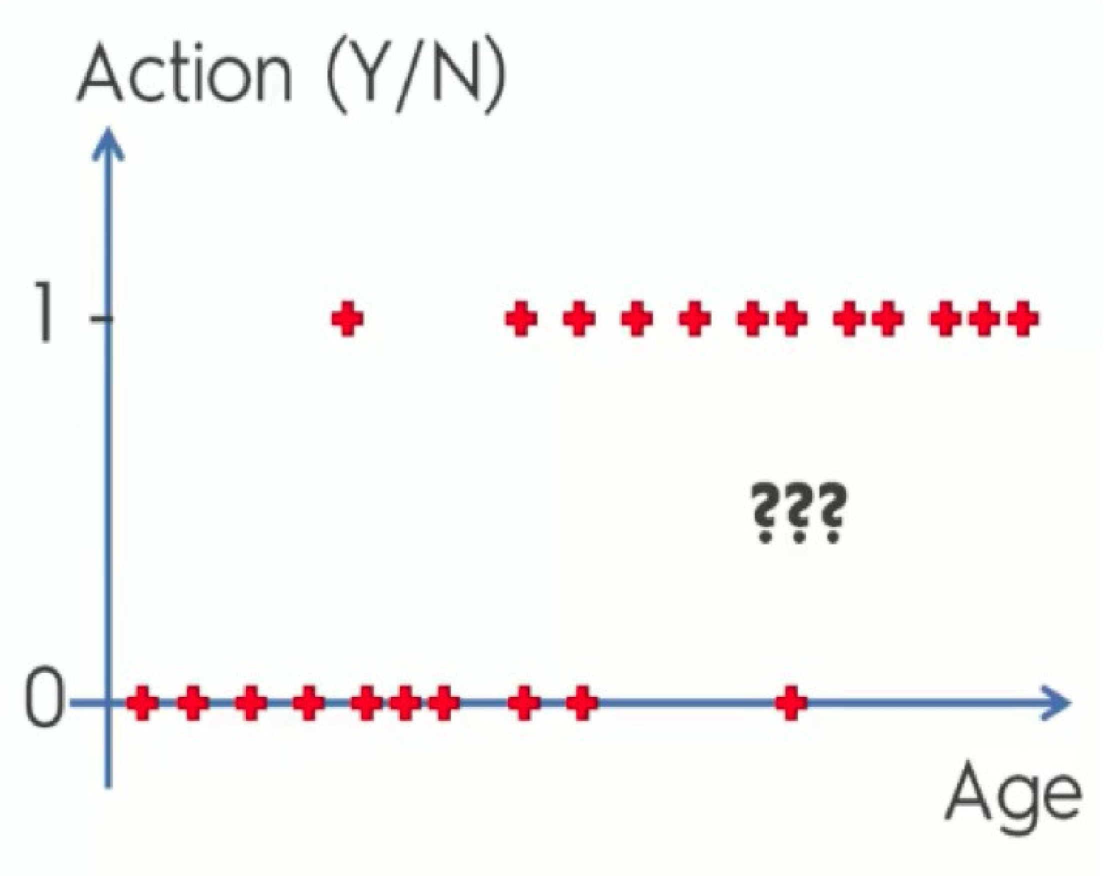
FIN: Your Model Is Ready

Example:
10 columns means
1,023 models

Logistic Regression

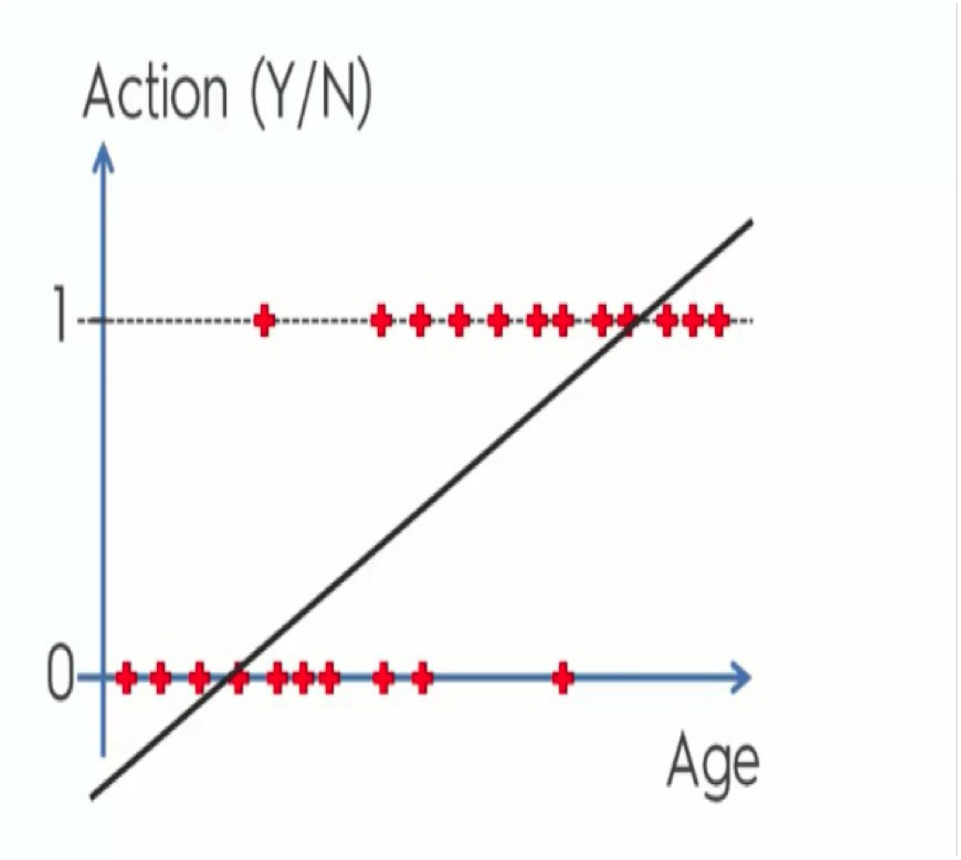
Say a company is sending out emails to customers or potential customers trying to persuade them to buy certain products and providing them with offers.

In the chart below, we have the contacted customers lined up horizontally. Those who are lined up along the x-axis are customers who declined the offers, whereas the other group consists of those who agreed to make a purchase.



Logistic Regression

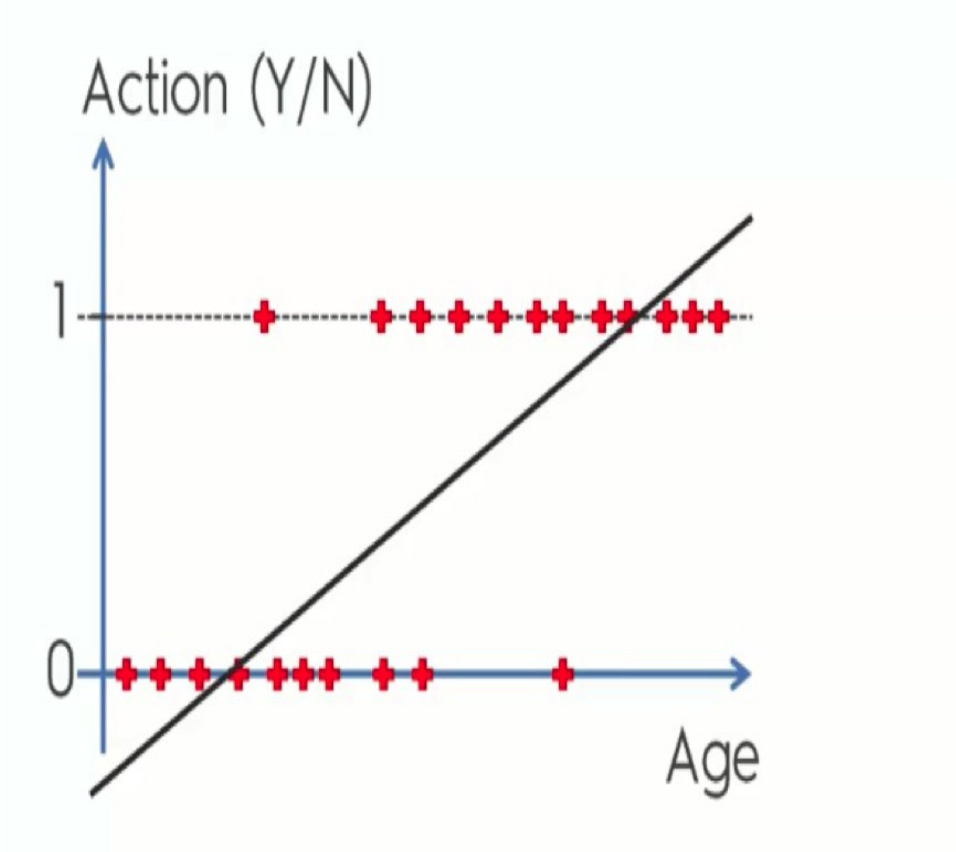
As you can see, the main pattern exhibited on the chart is that the older the customers are, the more likely they are to make the purchase. And, in contrast, the younger they are, the more likely they are to ignore the offer.



Logistic Regression

Instead of trying to predict each customer's action, we'll calculate the probabilities. The graph shows the Y-axis numbered from 0 to 1 and it so happens that probability falls between 0 and 1. What a match!

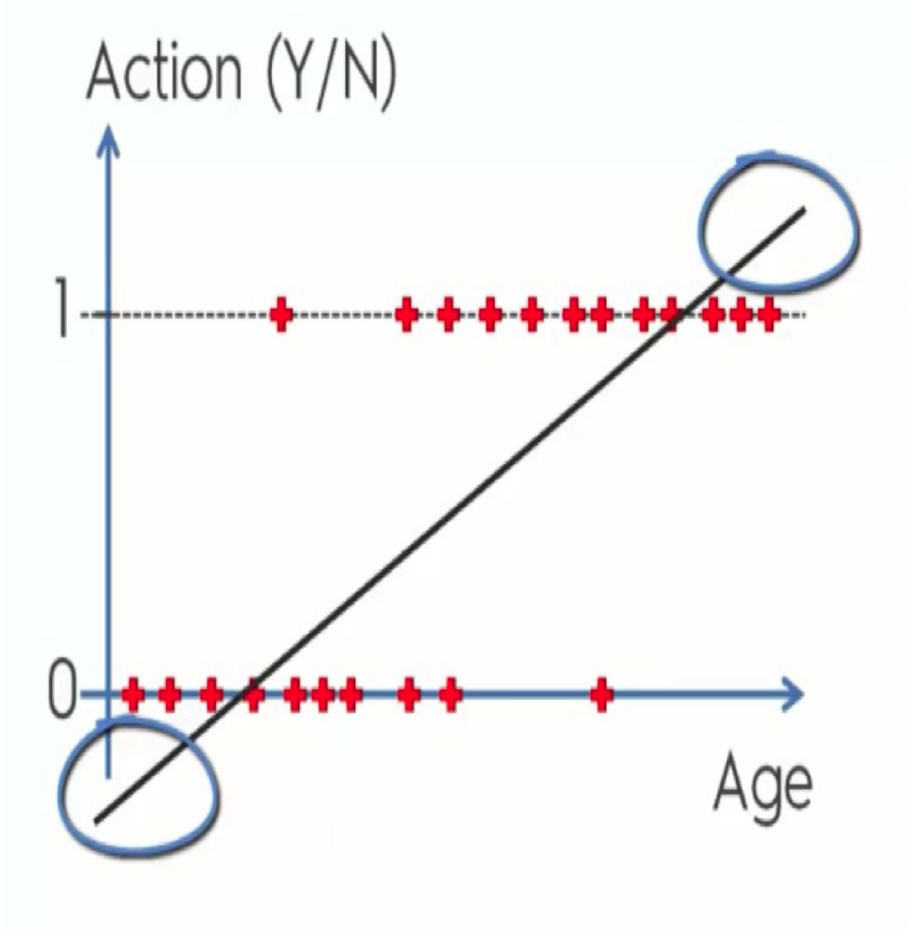
The reason that our current data only stands at either 0 or 1 is that we already have the actual results from our observation. When we're predicting, though, our points would fall between 0 and 1, because in this case we're only working with probabilities and do not have any solid observations.



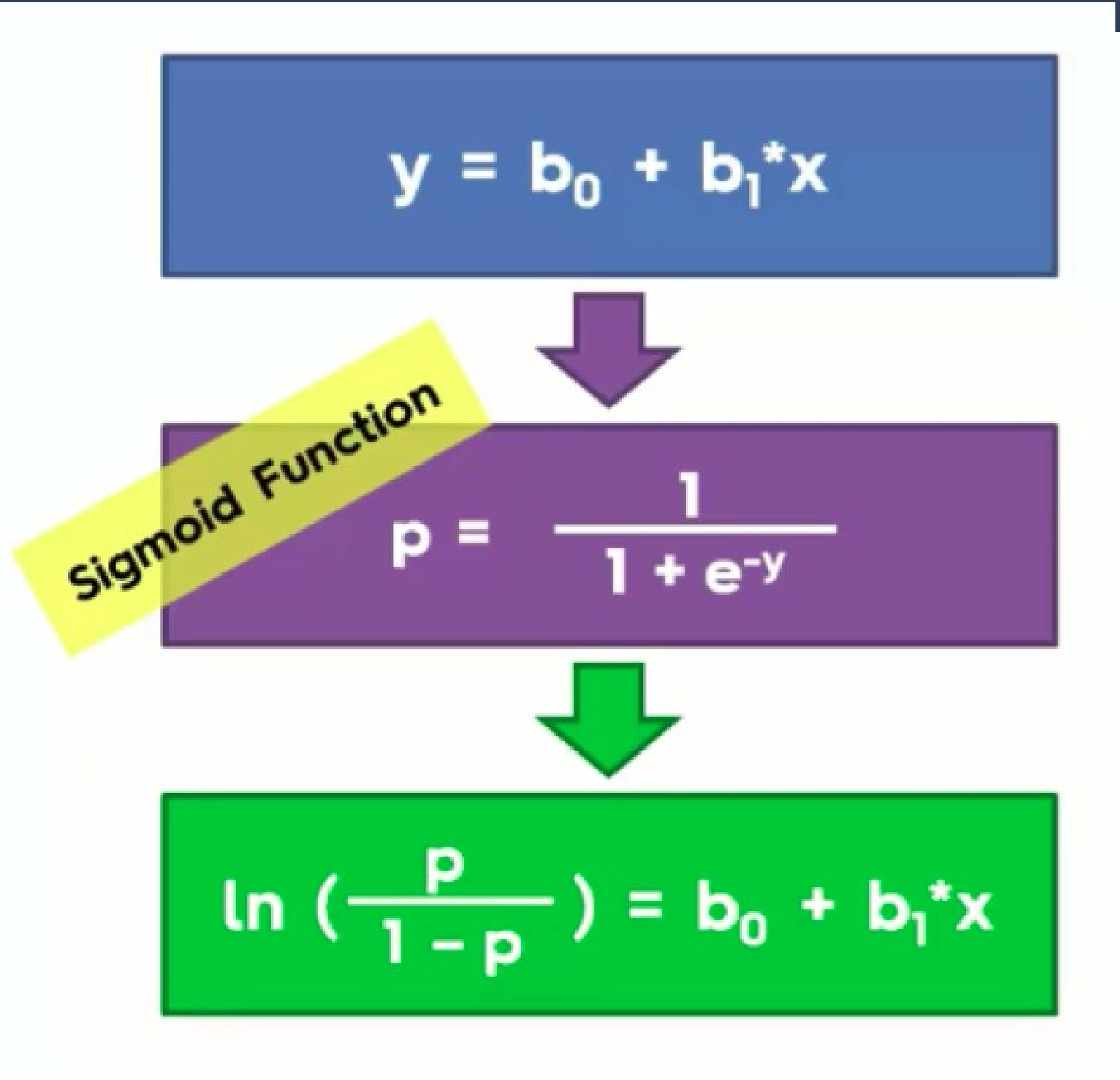
Logistic Regression

Now, let's say that the point where the line intersects with the x-axis stands for 35 years old customers. What this line does, in this case, is tell us that anyone who is 35 or more has a probability to buy one of the company's products, and the older they get the more probable they are to do it.

This way, we can learn that a 40-year old customer has a probability of 0.3 to make a purchase, for example, whereas a 55-year old would have a 0.7 probability.



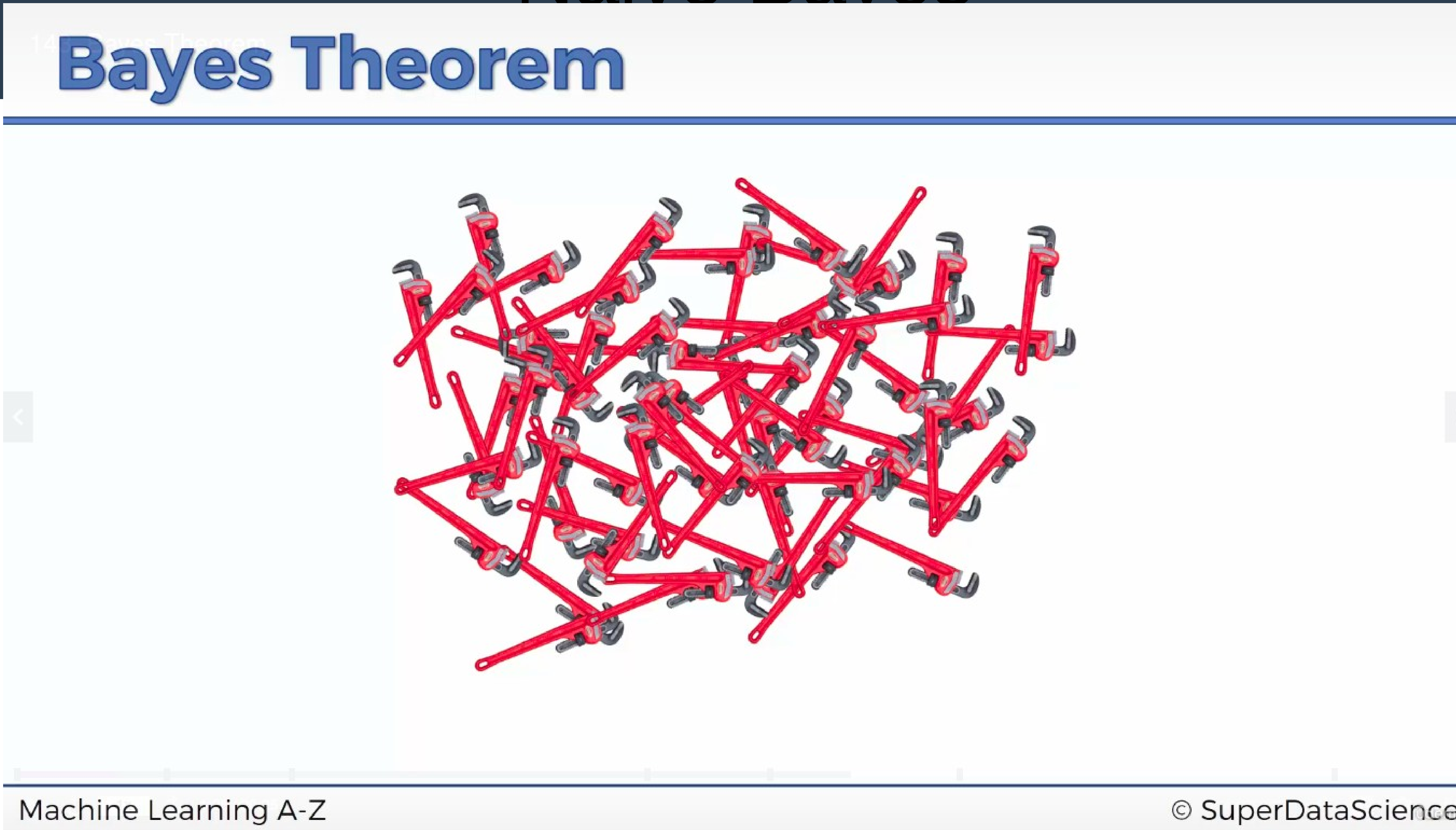
Logistic Regression



Naive Bayes

Bayes Theorem

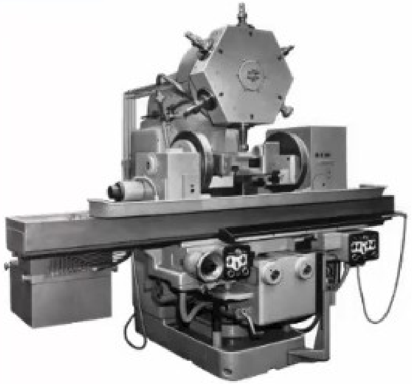
Naive Bayes



Naive Bayes

Bayes Theorem

What’s the probability?



m2

Naive Bayes

Bayes Theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Naive Bayes

Bayes Theorem

Mach1: 30 wrenches / hr
Mach2: 20 wrenches / hr

Out of all produced parts:
We can SEE that 1% are defective

Out of all defective parts:
We can SEE that 50% came from mach1
And 50% came from mach2

Question:
What is the probability that a part produced by mach2 is defective = ?

Naive Bayes

Bayes Theorem

Mach1: 30 wrenches / hr -> P(Mach1) = 30/50 = 0.6
Mach2: 20 wrenches / hr -> P(Mach2) = 20/50 = 0.4

Out of all produced parts:
We can SEE that 1% are defective -> P(Defect) = 1%

Out of all defective parts:
We can SEE that 50% came from mach1 -> P(Mach1 | Defect) = 50%
And 50% came from mach2 -> P(Mach2 | Defect) = 50%

Question:
What is the probability that a part produced by mach2 is defective = ? -> P(Defect | Mach2) = ?

Naive Bayes

Bayes Theorem

Mach1: 30 wrenches / hr
Mach2: 20 wrenches / hr
Out of all produced parts:
We can SEE that 1% are defective
Out of all defective parts:
We can SEE that 50% came from mach1
And 50% came from mach2
Question:
What is the probability that a part
produced by mach2 is defective = ?

-> P(Mach2) = 20/50 = 0.4
-> P(Defect) = 1%
-> P(Mach2 | Defect) = 50%
-> P(Defect | Mach2) = ?

$$P(Defect \mid Mach2) = \frac{0.5 \quad * \quad 0.01}{0.4} = 0.0125$$

Naive Bayes

It's intuitive!

$$P(\text{Defect} \mid \text{Mach2}) = \frac{P(\text{Mach2} \mid \text{Defect}) * P(\text{Defect})}{P(\text{Mach2})} = 1.25\%$$

Let's look at an example:

- 1000 wrenches
- 400 came from Mach2
- 1% have a defect = 10
- of them 50% came from Mach2 = 5
- % defective parts from Mach2 = 5/400 = 1.25%

Naive Bayes

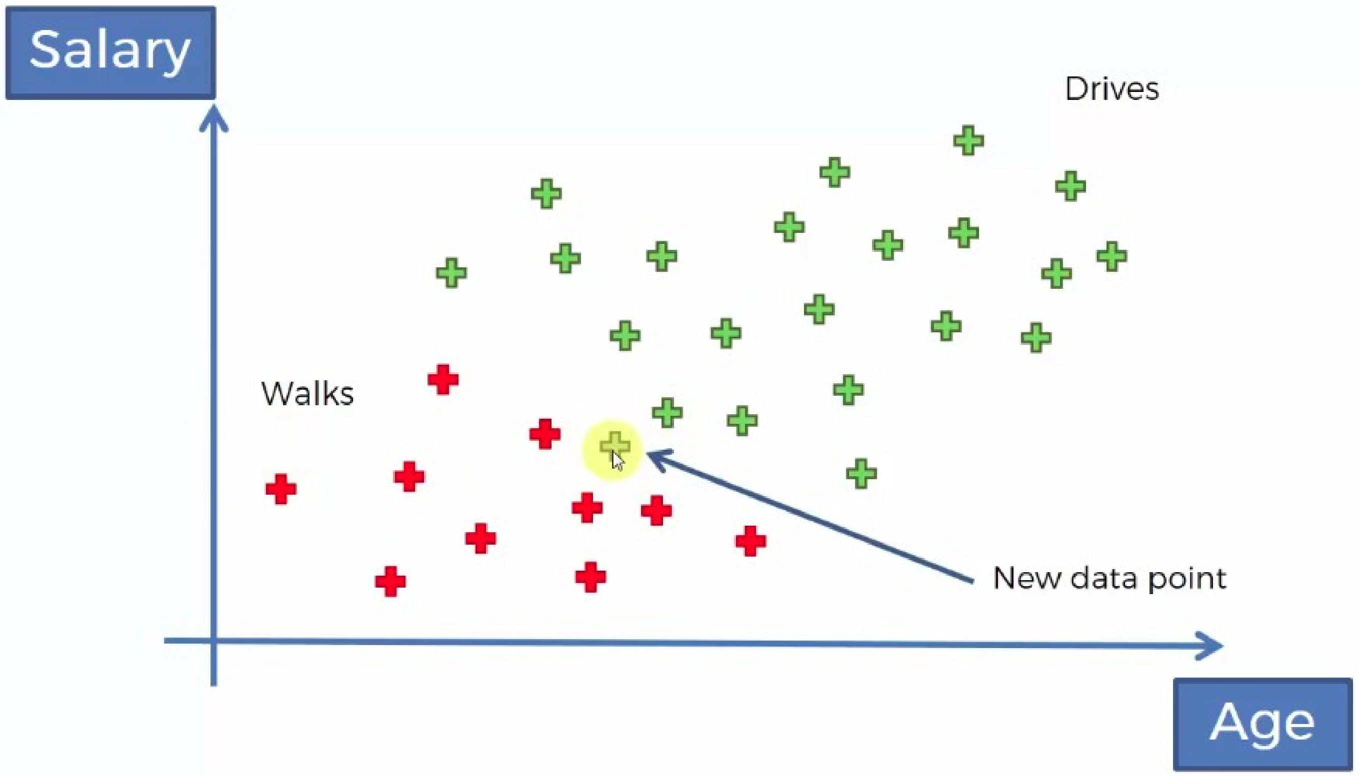
It's intuitive!

Obvious question:
If the items are labeled, why couldn't we just count the number of defective wrenches that came from Mach2 and divide by the total number that came from Mach2?

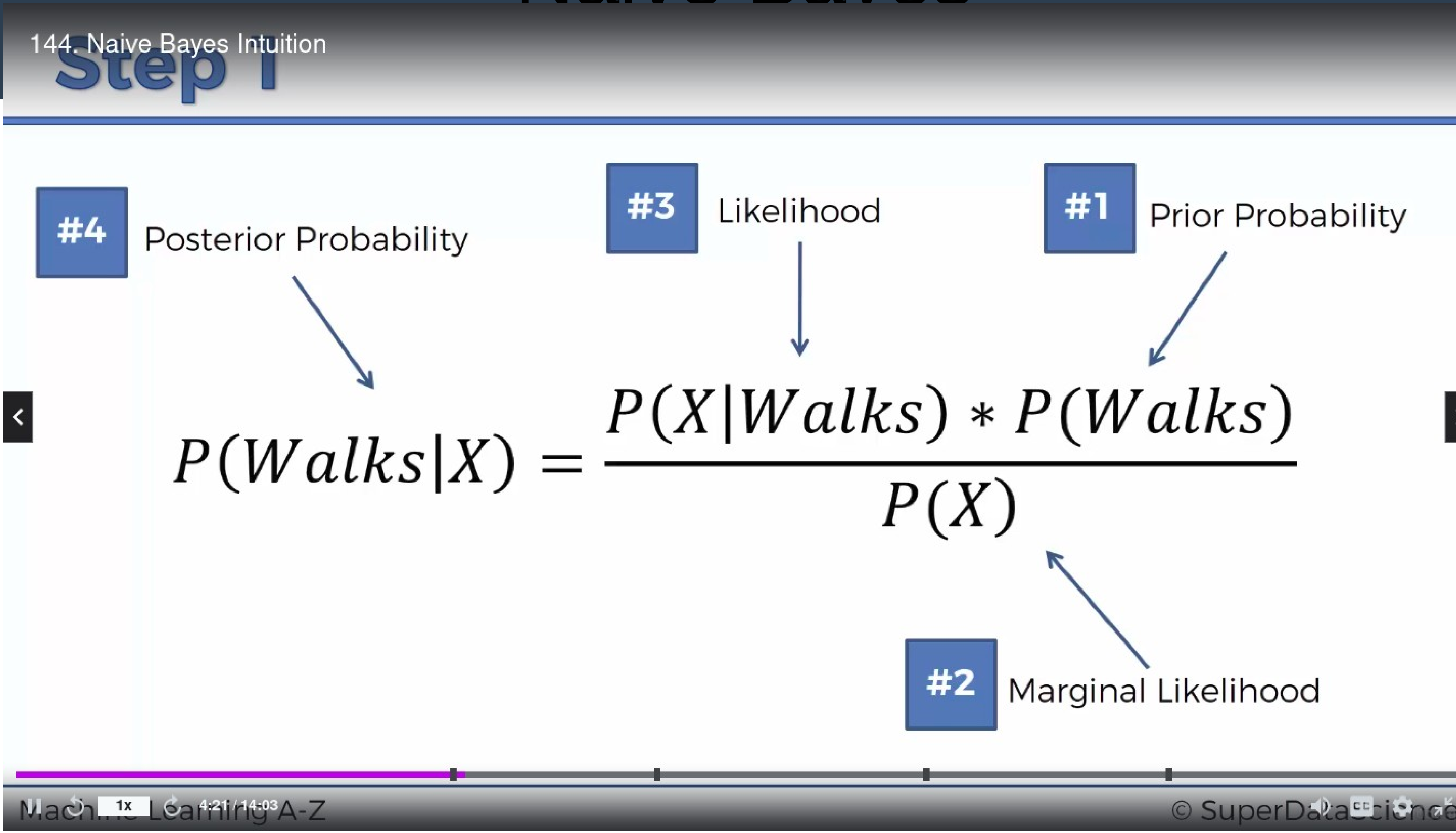
151 people have written a note here.

Naive Bayes

Naïve Bayes



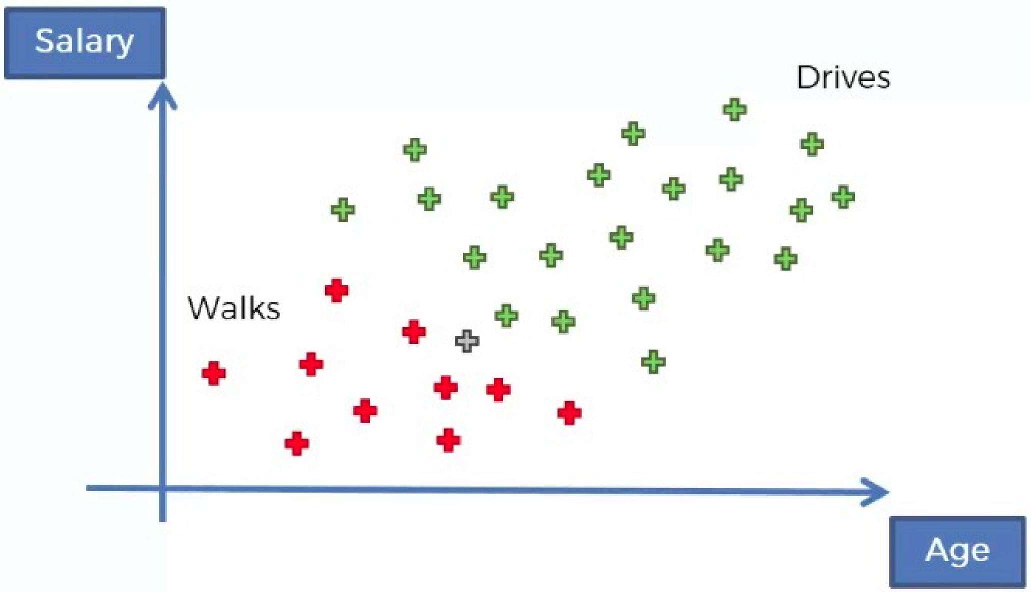
Naive Bayes


144. Naive Bayes Intuition
Step 1
#4 Posterior Probability
#3 Likelihood
#1 Prior Probability
$P(Walks|X) = \frac{P(X|Walks) * P(Walks)}{P(X)}$
#2 Marginal Likelihood
1x
4:21 / 14:03
Machine Learning A-Z
© SuperDataScience

Naive Bayes

Naïve Bayes: Step 1

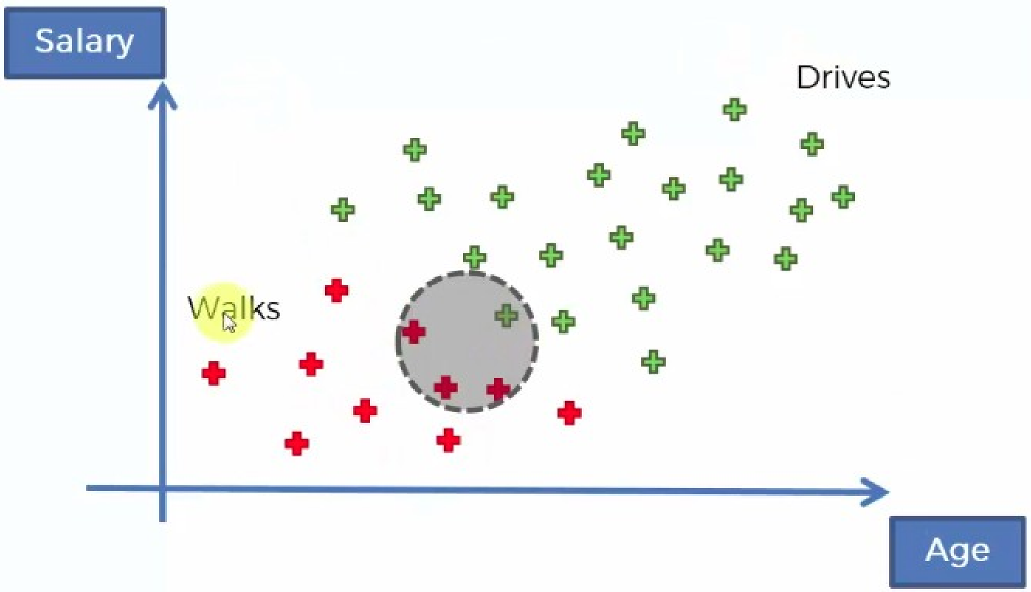
#1. P(Walks)



$$P(Walks) = \frac{Number\ of\ Walkers}{Total\ Observations}$$

$$P(Walks) = \frac{10}{30}$$

Naïve Bayes: Step 1



#2. P(X)

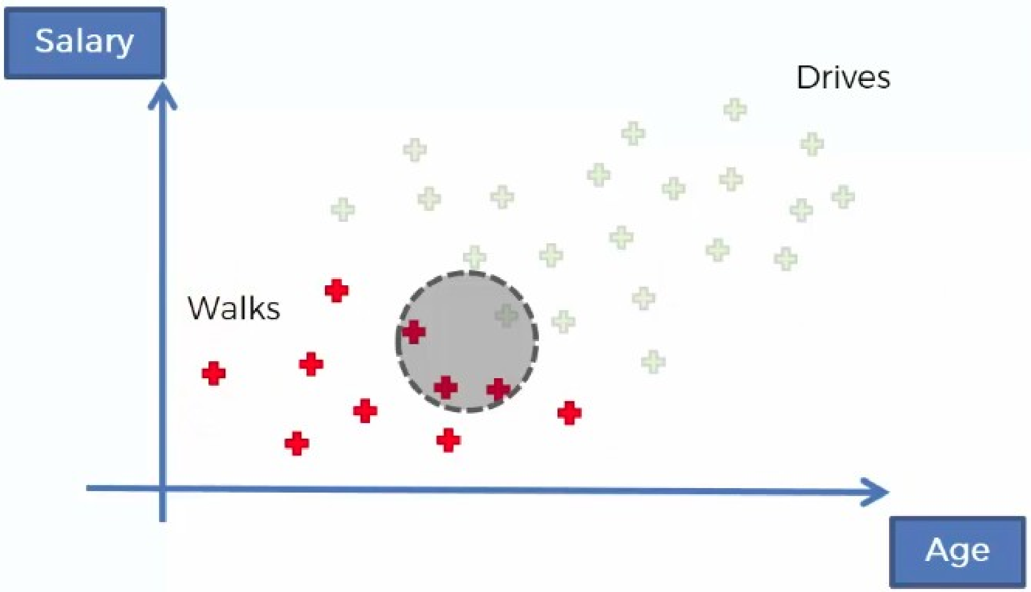
$$P(X) = \frac{Number\ of\ Similar\ Observations}{Total\ Observations}$$

$$P(X) = \frac{4}{30}$$

340 people have written a note here.

Naive Bayes

Naïve Bayes: Step 1

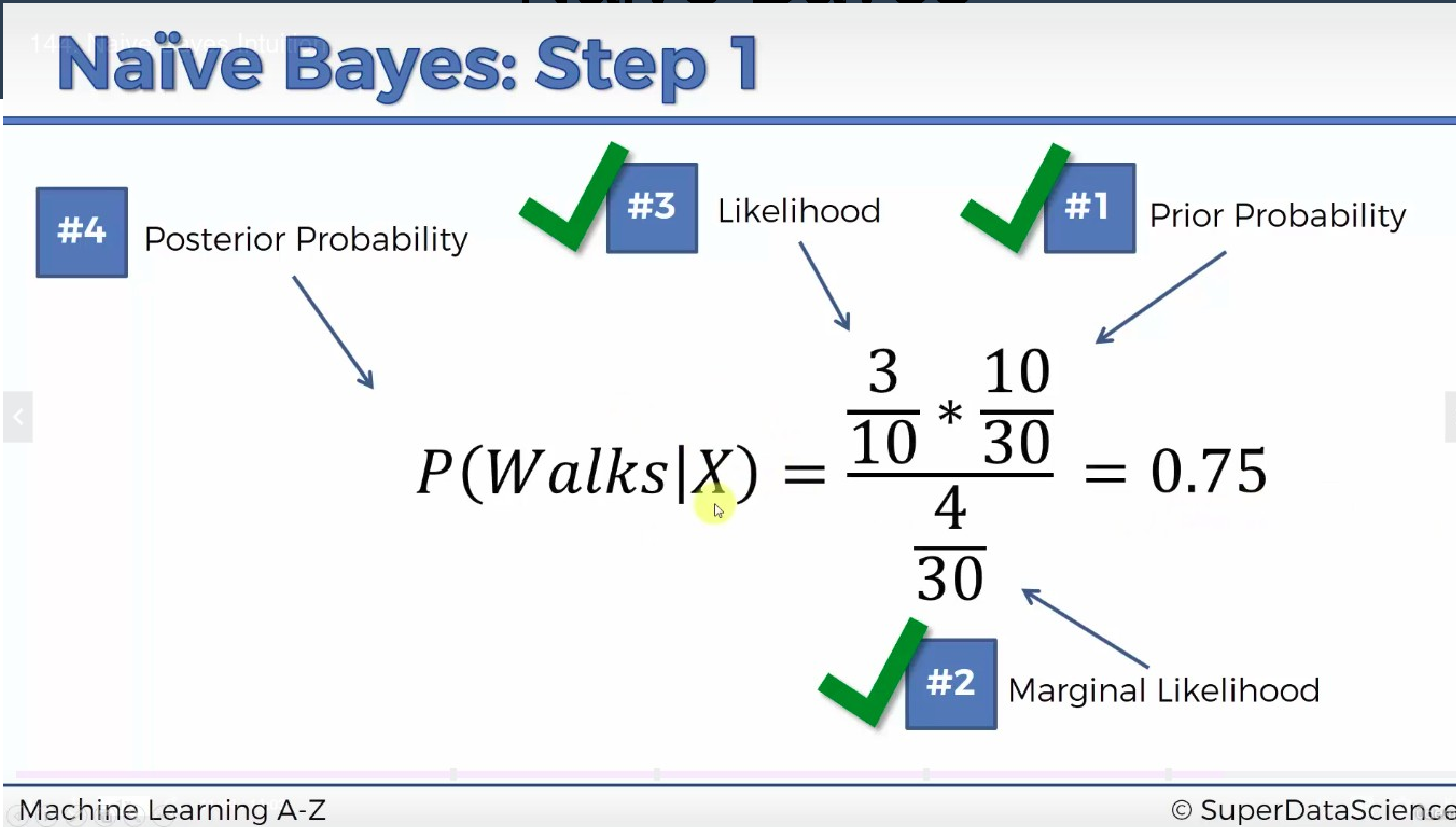


#3. P(X|Walks)

$$P(X|Walks) = \frac{\begin{matrix}Number\ of\ Similar \\ Observations \\ Among\ those\ who\ Walk\end{matrix}}{Total\ number\ of\ Walkers}$$

$$P(X|Walks) = \frac{3}{10}$$

Naive Bayes



Naïve Bayes: Step 1
#4 Posterior Probability
#3 Likelihood
#1 Prior Probability
$P(Walks|X) = \frac{\frac{3}{10} * \frac{10}{30}}{\frac{4}{30}} = 0.75$
#2 Marginal Likelihood
Machine Learning A-Z
© SuperDataScience

Naive Bayes

Step 2

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Drives|X) = \frac{P(X|Drives) * P(Drives)}{P(X)}$$

#2 Marginal Likelihood

Naive Bayes

Naïve Bayes: Step 2

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

#2 Marginal Likelihood

$$P(Drives|X) = \frac{\frac{1}{20} * \frac{20}{30}}{\frac{4}{30}} = 0.25$$

Naive Bayes

Step 3

$$P(Walks|X)\ v.s.\ P(Drives|X)$$

Naive Bayes

14. Naive Bayes Intuition

Step 3

$$P(Walks|X) > P(Drives|X)$$

Naive Bayes

Naïve Bayes