

Correlation Coefficient Matrix

Age	Experience	Education Received	Salary
32	8	6	\$ 8,000
40	15	8	\$ 12,000
35	6	8	\$ 10,000

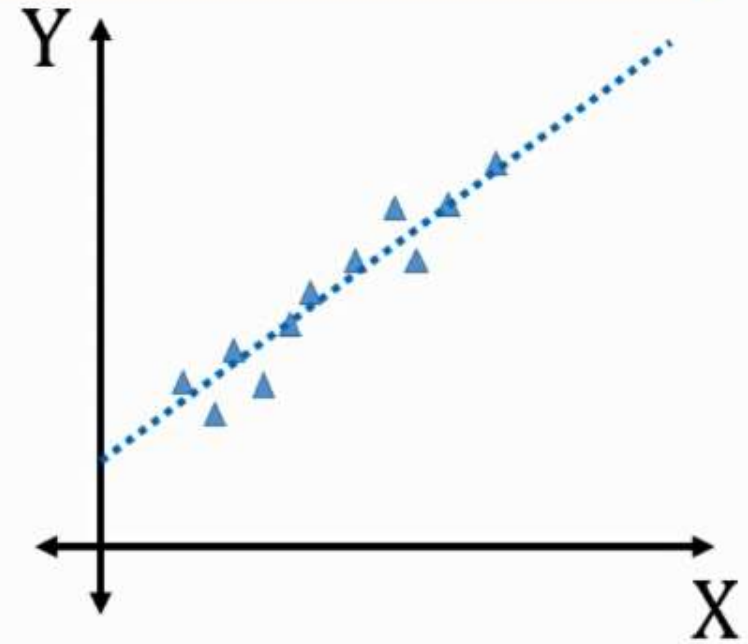
	Age	Experience	Education Received	Salary
Age	1	0.9	0.2	0.7
Experience	0.9	1	0.15	0.72
Education Received	0.2	0.15	1	0.85
Salary	0.7	0.72	0.85	1

Statistically Correlated

- Strength of the correlation – Coefficient of Correlation
- Direction of correlation – Sign of the Coefficient

Pearson Correlation
Coefficient

$$r = \frac{\sum (x - \bar{x}) * (y - \bar{y})}{(N - 1) * \sigma_x * \sigma_y}$$



Correlation Coefficient Matrix

Age	Experience	Education Received	Salary
32	8	6	\$ 8,000
40	15	8	\$ 12,000
35	6	8	\$ 10,000

	Age	Experience	Education Received	Salary
Age	1	0.88	0.2	0.7 X
Experience	0.88	1	0.15	0.72 ✓
Education Received	0.2	0.15	1	0.85
Salary	0.7	0.72	0.85	1

QuickFix

- Start with All the variable
- Remove unwanted ones using Adjusted R-Squared or feature selection methods

Multiple Linear Regression

Hrs Studied (X1)	Hrs Slept (X2)	Marks (Y)
0	8	40
2	8	52
3	7.5	53
4	7	55
4	9	56
5	8.5	72
6	9	71
6	7	88
7	6	56
7	7	74
8	9	89
9	6	67
9	9	89

$$y = b_0 + b_1 x_1 + b_2 x_2$$

Dependent Variable
Marks Obtained

Independent Variable
Hrs Studied
Hrs Slept

Categorical Variable in Regression

Hrs Studied (X1)	Hrs Slept (X2)	Major Subject (X3)	Marks (Y)
0	8	Math	40
2	8	Science	52
3	7.5	Art	53
4	7	Math	55
4	9	Math	56
5	8.5	Math	72
6	9	Science	71
6	7	Art	88
7	6	Art	56
7	7	Science	74
8	9	Science	89
9	6	Math	67
9	9	Art	89

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

Dependent Variable
Marks Obtained

Independent Variable
Hrs Studied
Hrs Slept
Major Subject

Label Encoding

Hrs Studied (X1)	Hrs Slept (X2)	Major Subject (X3)	Marks (Y)
0	8	1	40
2	8	2	52
3	7.5	3	53
4	7	1	55
4	9	1	56
5	8.5	1	72
6	9	2	71
6	7	3	88
7	6	3	56
7	7	2	74
8	9	2	89
9	6	1	67
9	9	3	89

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

Dependent Variable

Marks Obtained

Independent Variable

Hrs Studied

Hrs Slept

Major Subject

Math – 1

Science – 2

Art – 3

Hot Encoding or Dummy Variable

Hrs Studied (X1)	Hrs Slept (X2)	Math (X3)	Science (X4)	Art (X5)	Marks (Y)
0	8	1	0	0	40
2	8	0	1	0	52
3	7.5	0	0	1	53
4	7	1	0	0	55
4	9	1	0	0	56
5	8.5	1	0	0	72
6	9	0	1	0	71
6	7	0	0	1	88
7	6	0	0	1	56
7	7	0	1	0	74
8	9	0	1	0	89
9	6	1	0	0	67
9	9	0	0	1	89

Dummy Variable Trap

Hrs Studied (X1)	Hrs Slept (X2)	Math (X3)	Science (X4)	Art (X5)	Marks (Y)
0	8	1	0	0	40
2	8	0	1	0	52
3	7.5	0	0	1	53
4	7	1	0	0	55
4	9	1	0	0	56
5	8.5	1	0	0	72
6	9	0	1	0	71
6	7	0	0	1	88
7	6	0	0	1	56
7	7	0	1	0	74
8	9	0	1	0	89
9	6	1	0	0	67
9	9	0	0	1	89

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4$$

Dummy Variable Trap

Hrs Studied (X1)	Hrs Slept (X2)	Math (X3)	Science (X4)	Art (X5)	Marks (Y)
0	8	1	0	0	40
2	8	0	1	0	52
3	7.5	0	0	1	53
4	7	1	0	0	55
4	9	1	0	0	56
5	8.5	1	0	0	72
6	9	0	1	0	71
6	7	0	0	1	88
7	6	0	0	1	56
7	7	0	1	0	74
8	9	0	1	0	89
9	6	1	0	0	67
9	9	0	0	1	89

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4$$

```
pd.get_dummies(data, drop_first=True)
```



?

Quiz





Question 1:

We should use Simple Linear Regression to predict the winner of a football game

- True
- False



Question 2:

What is the class used in Python to create a simple linear regressor ?

☐ SimpleLinear

☐ LinearRegression

☐ LinReg

☐ SimpleLinearRegression



Question 3:

Which of the following formulas is not a simple linear regression model ?

☐ Salary = a * Experience

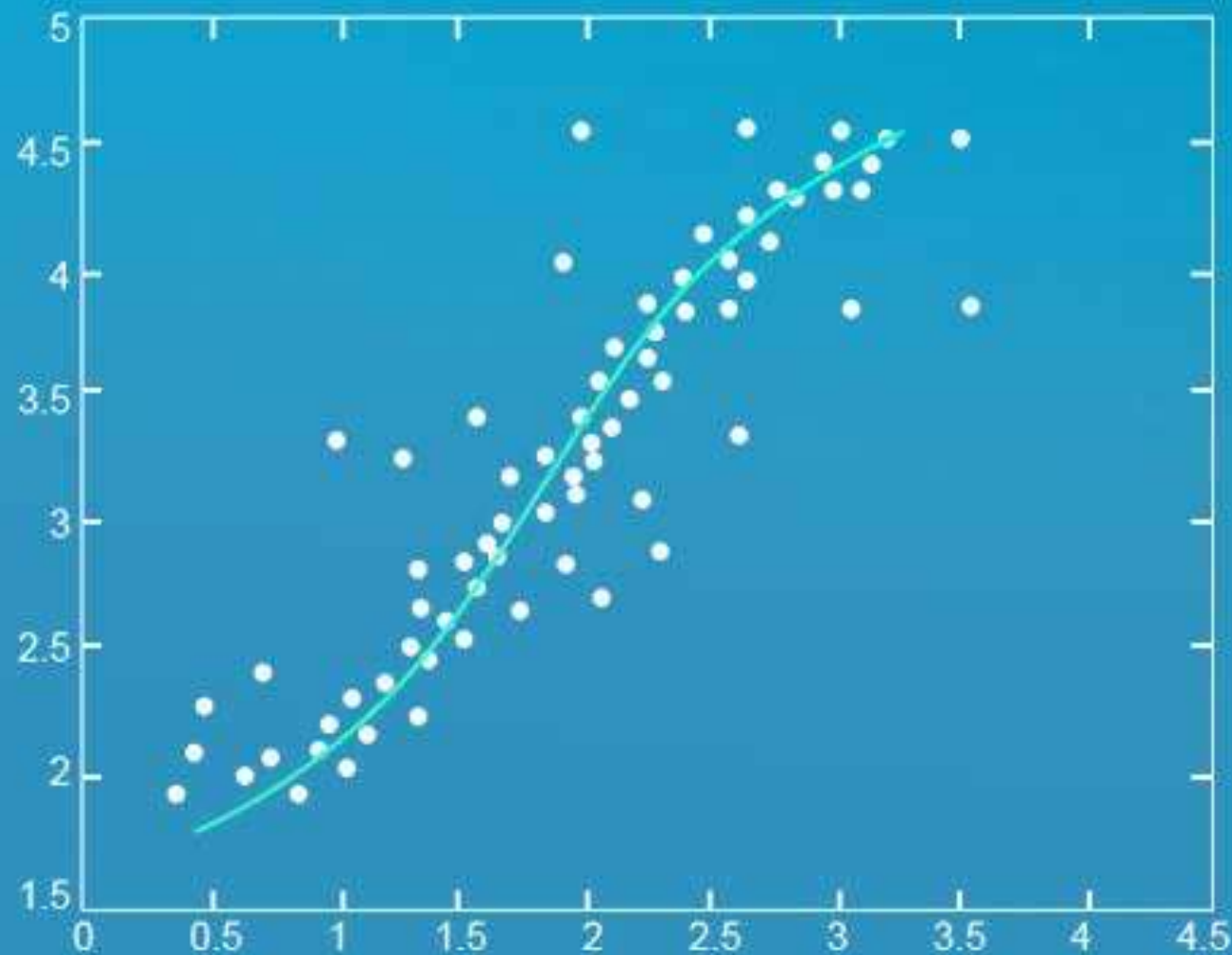
☐ Salary = a * Experience + b

☐ Salary = a * Experience + b * Age

Task: Create ML model to predict the Power output for (Combined Cycle Power Plant - CCPP)

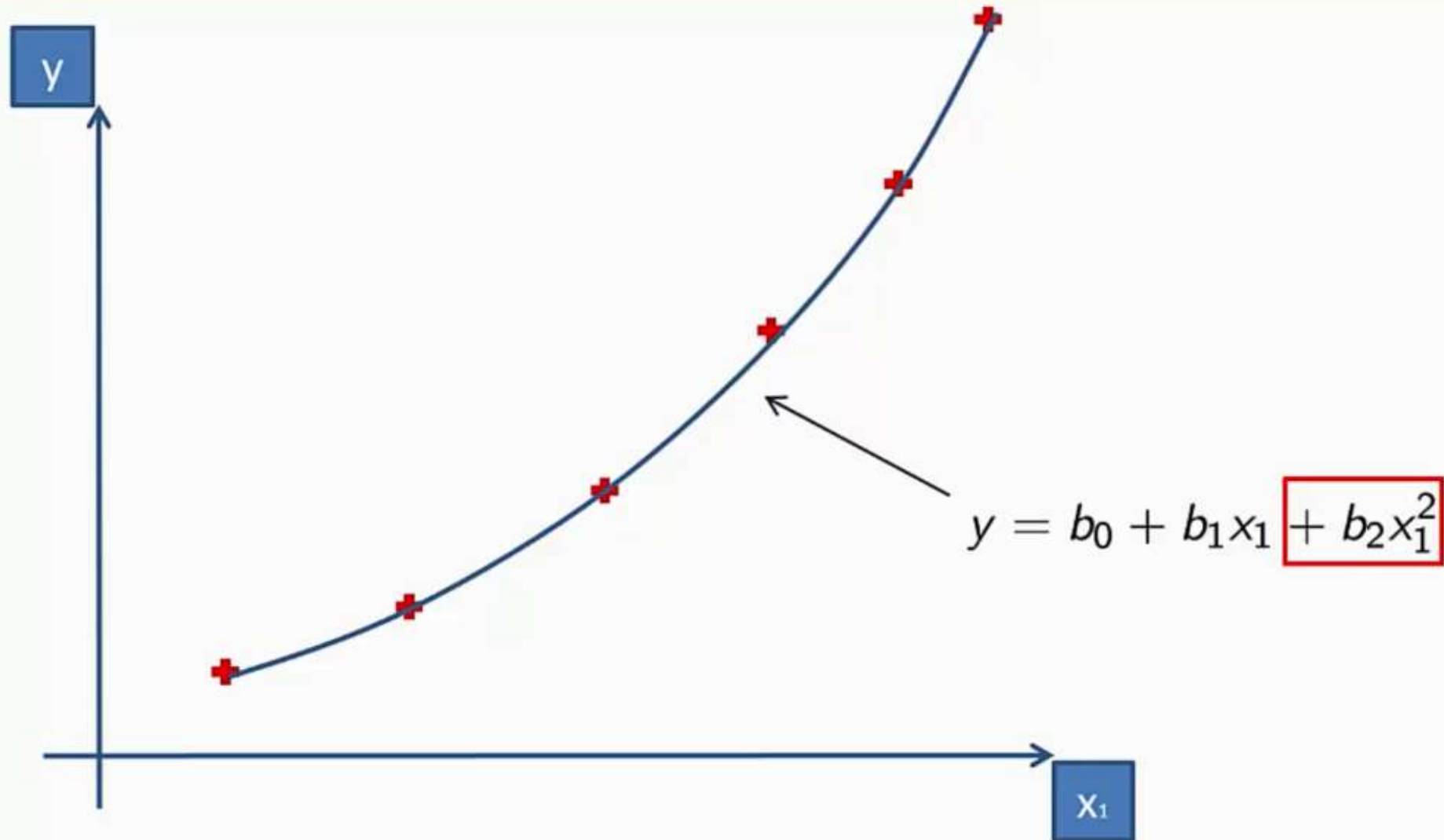


SVM Linear Classifier with $C=1$



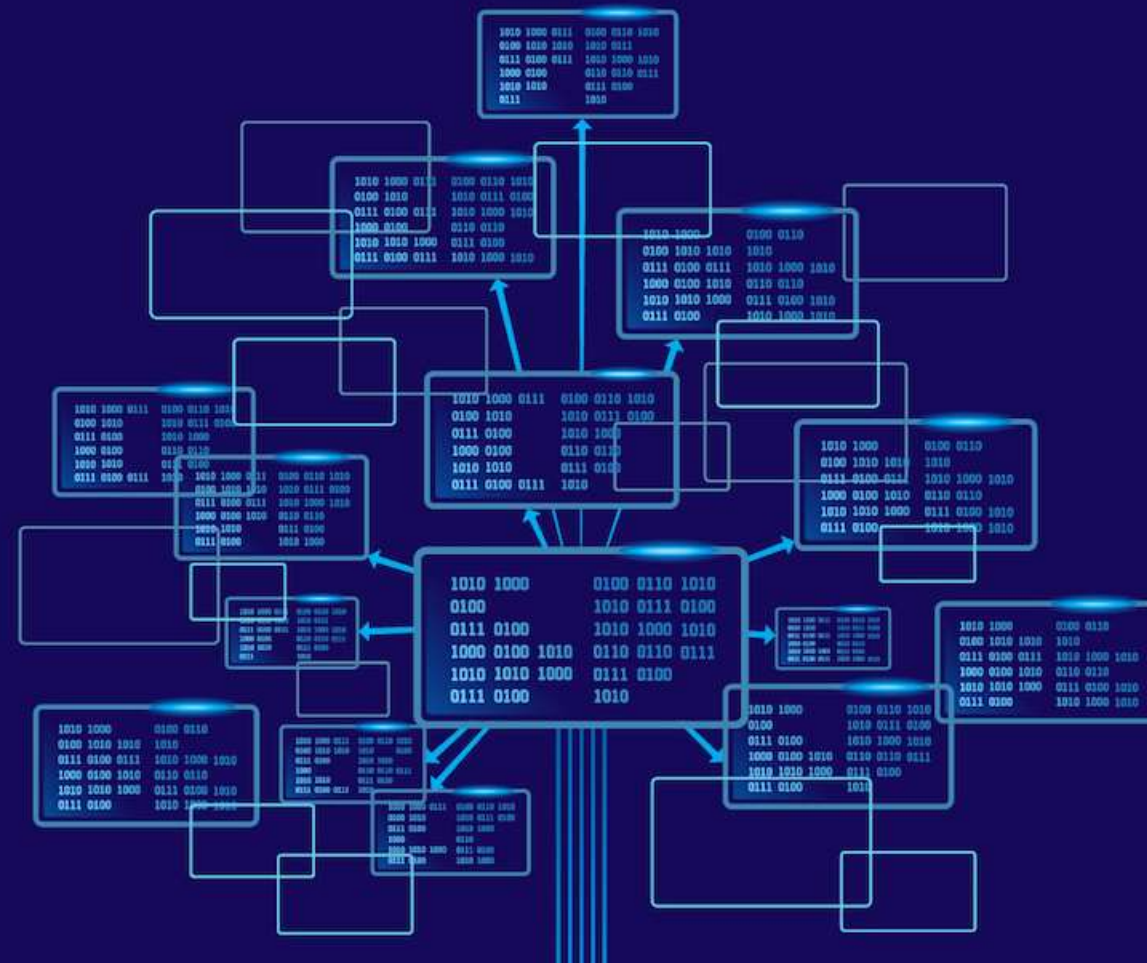
Polynomial Regression

Polynomial Regression



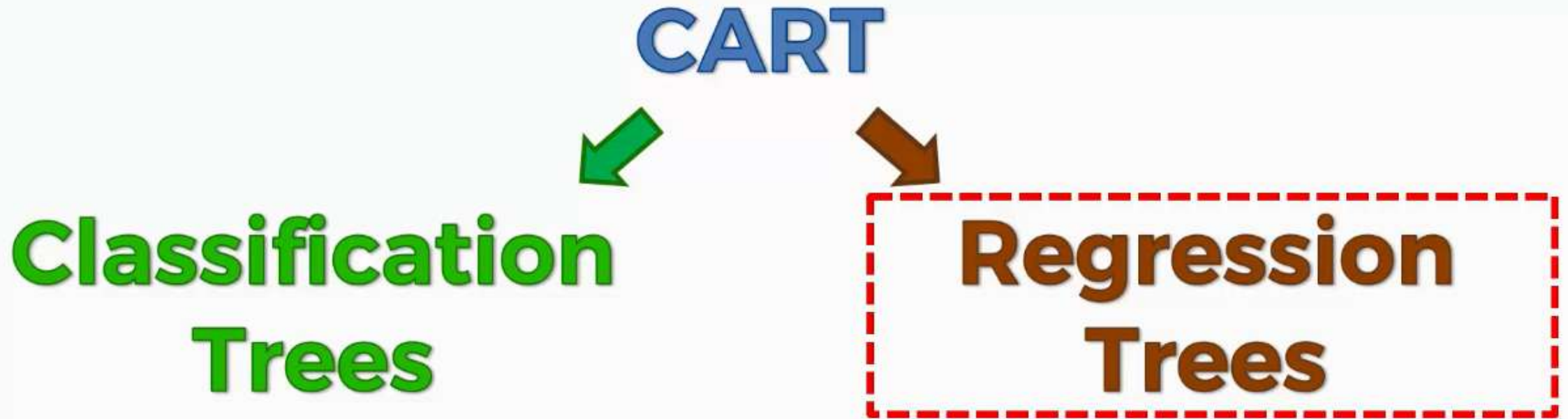
Demo: Create ML using Polynomial Regression and compare it with Multiple regression.



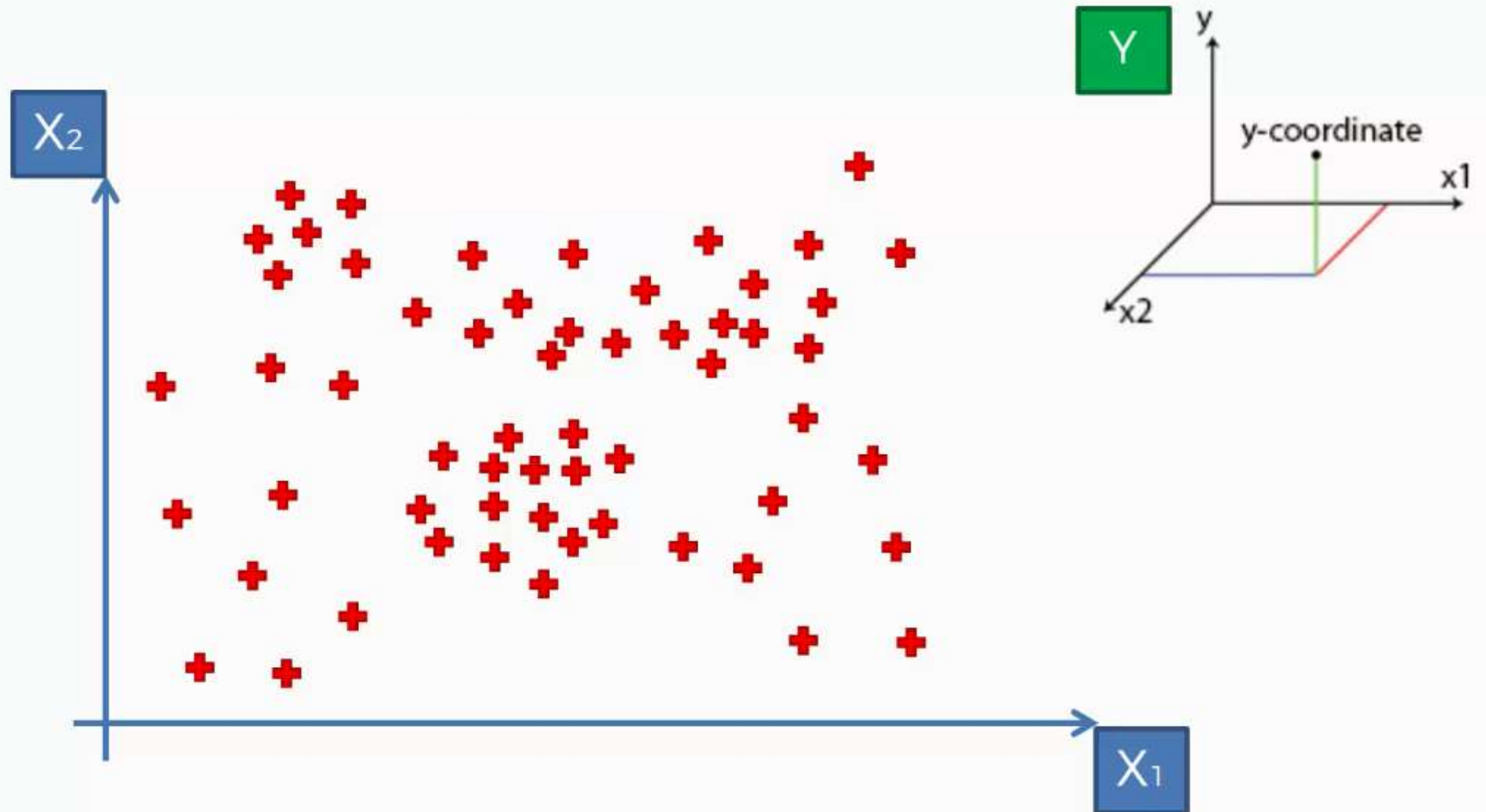


Decision Tree Regression

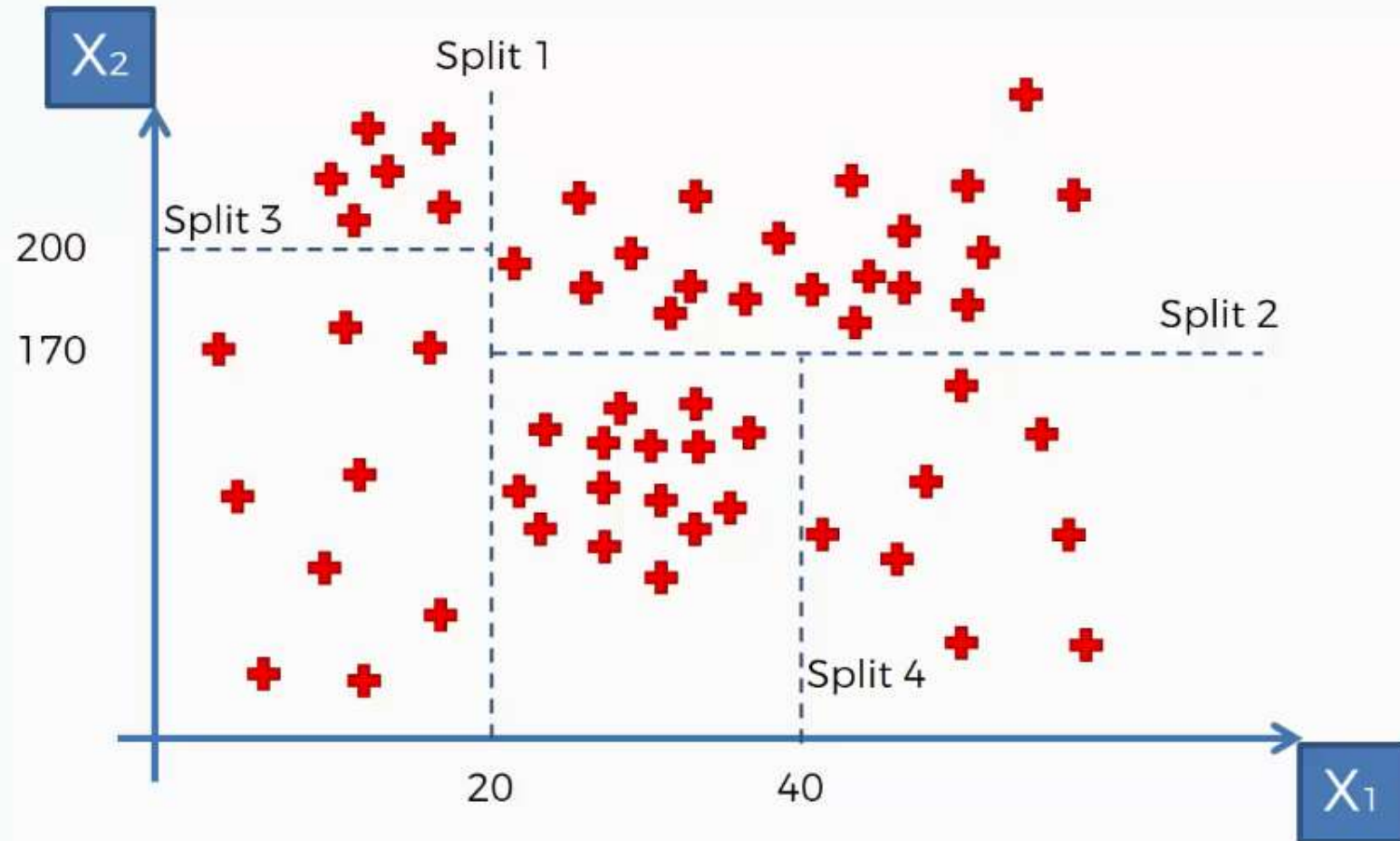
Decision Tree Intuition



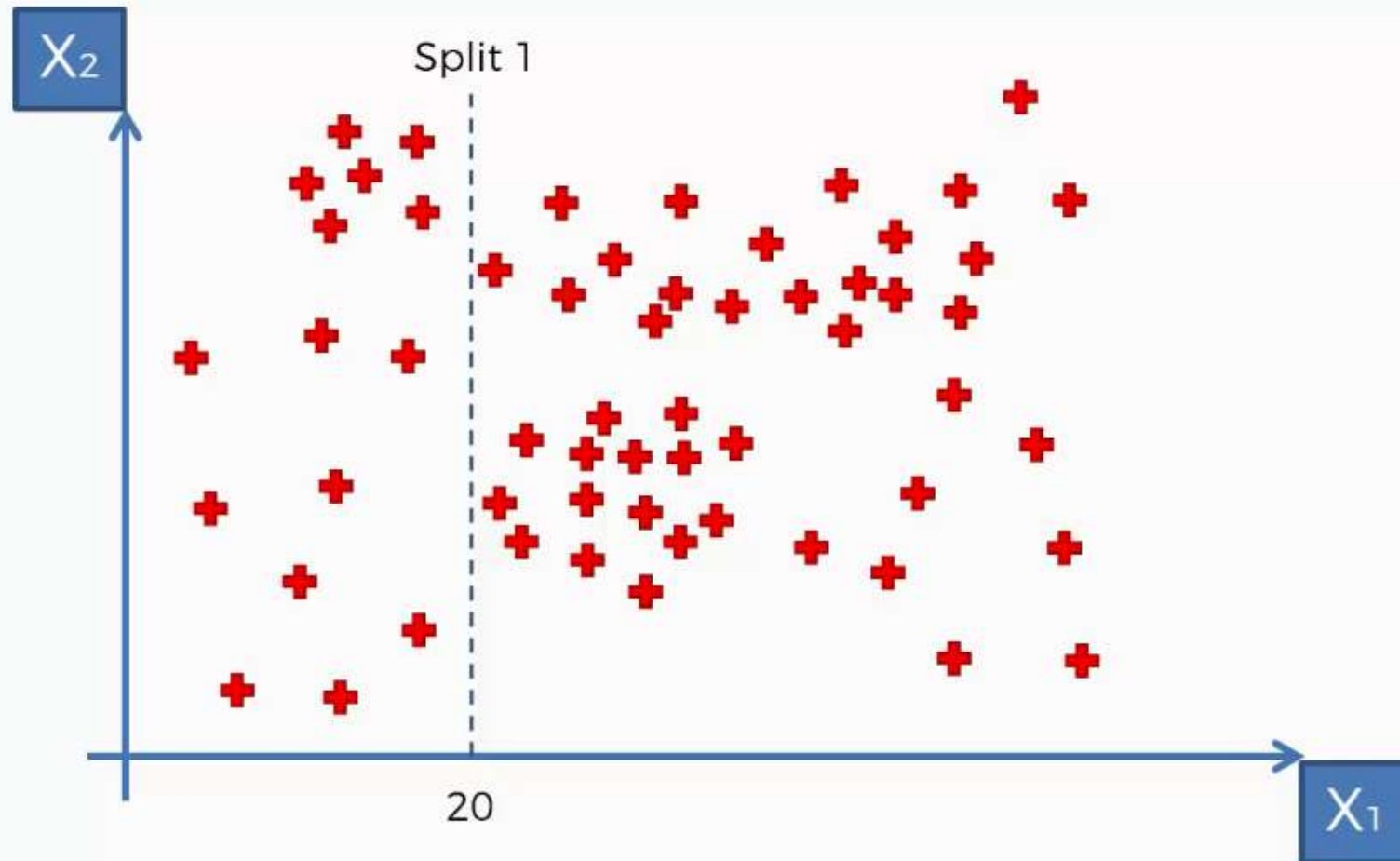
Decision Tree Intuition



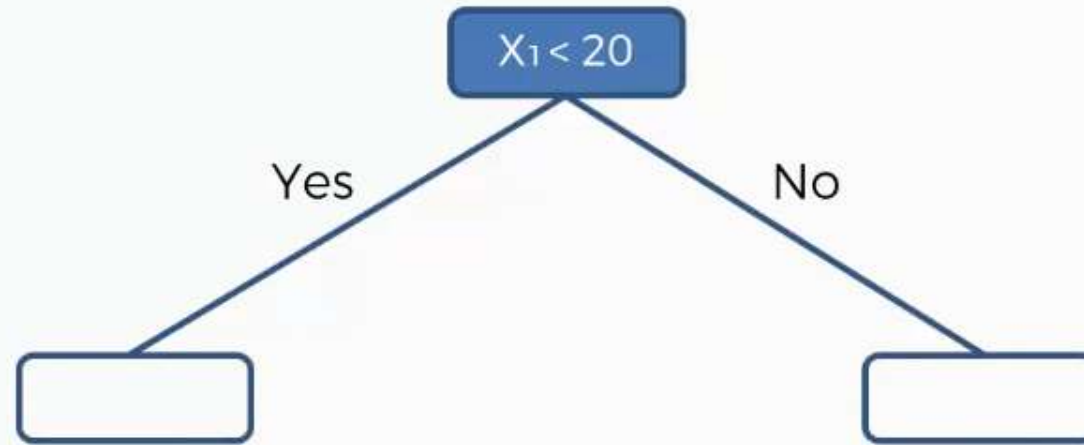
Decision Tree Intuition



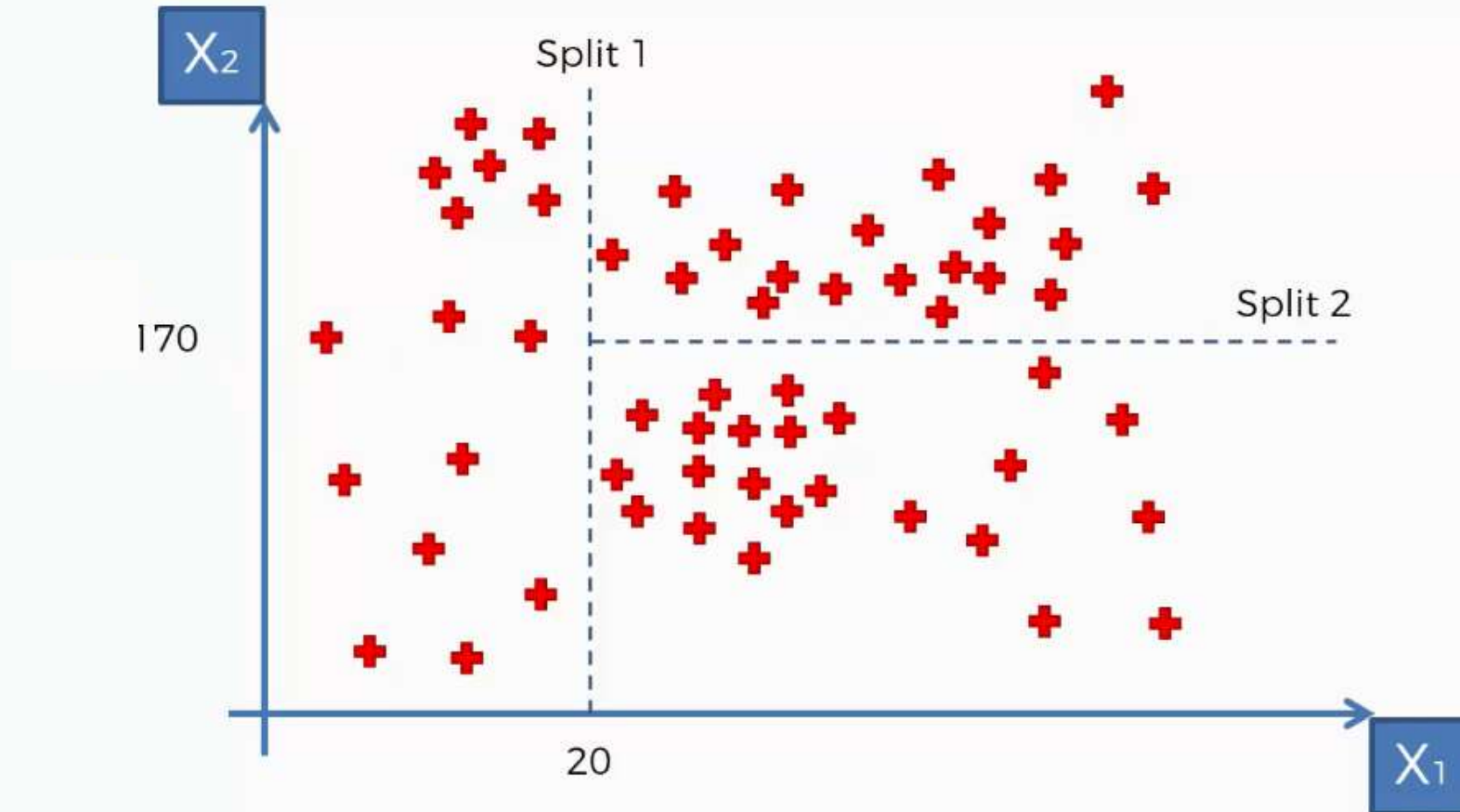
Decision Tree Intuition



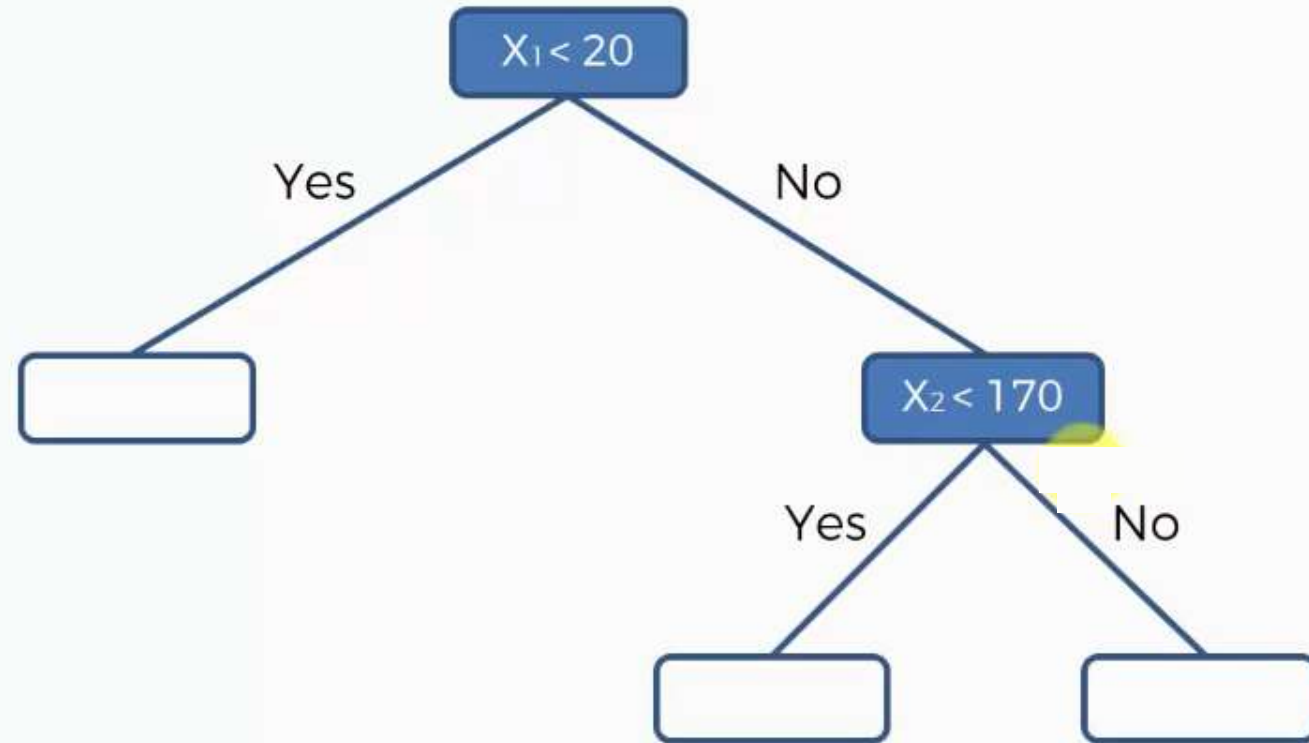
Decision Tree Intuition



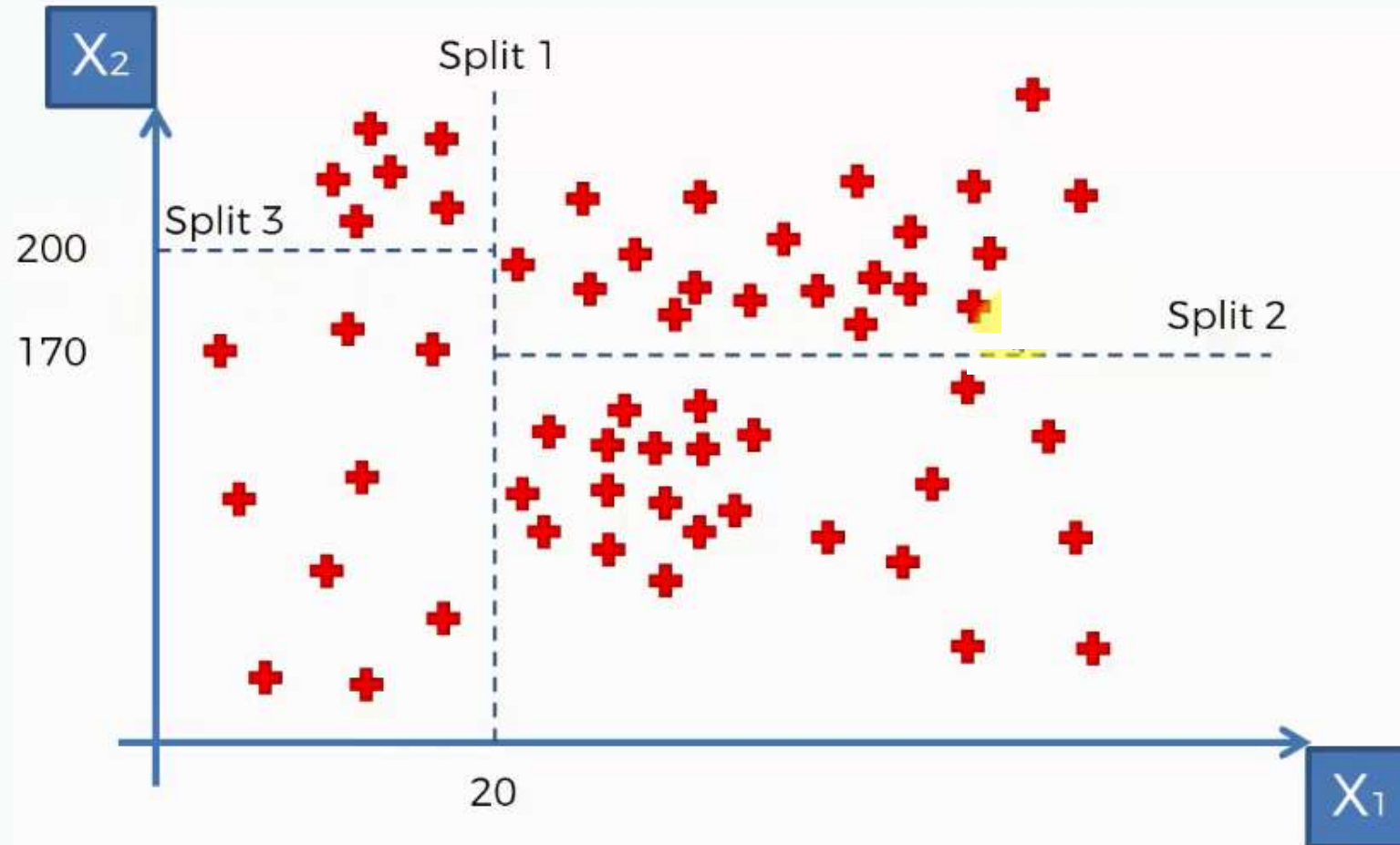
Decision Tree Intuition



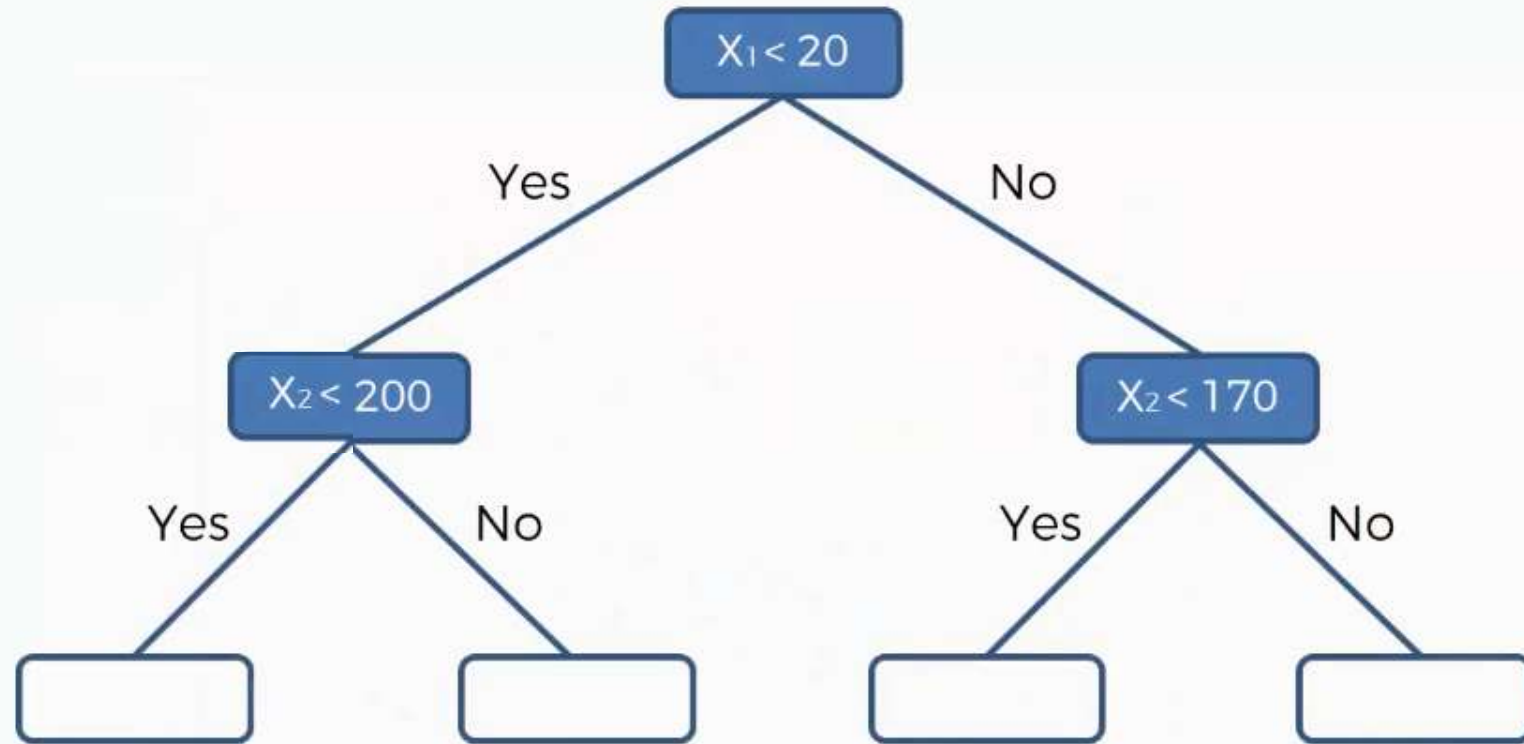
Decision Tree Intuition



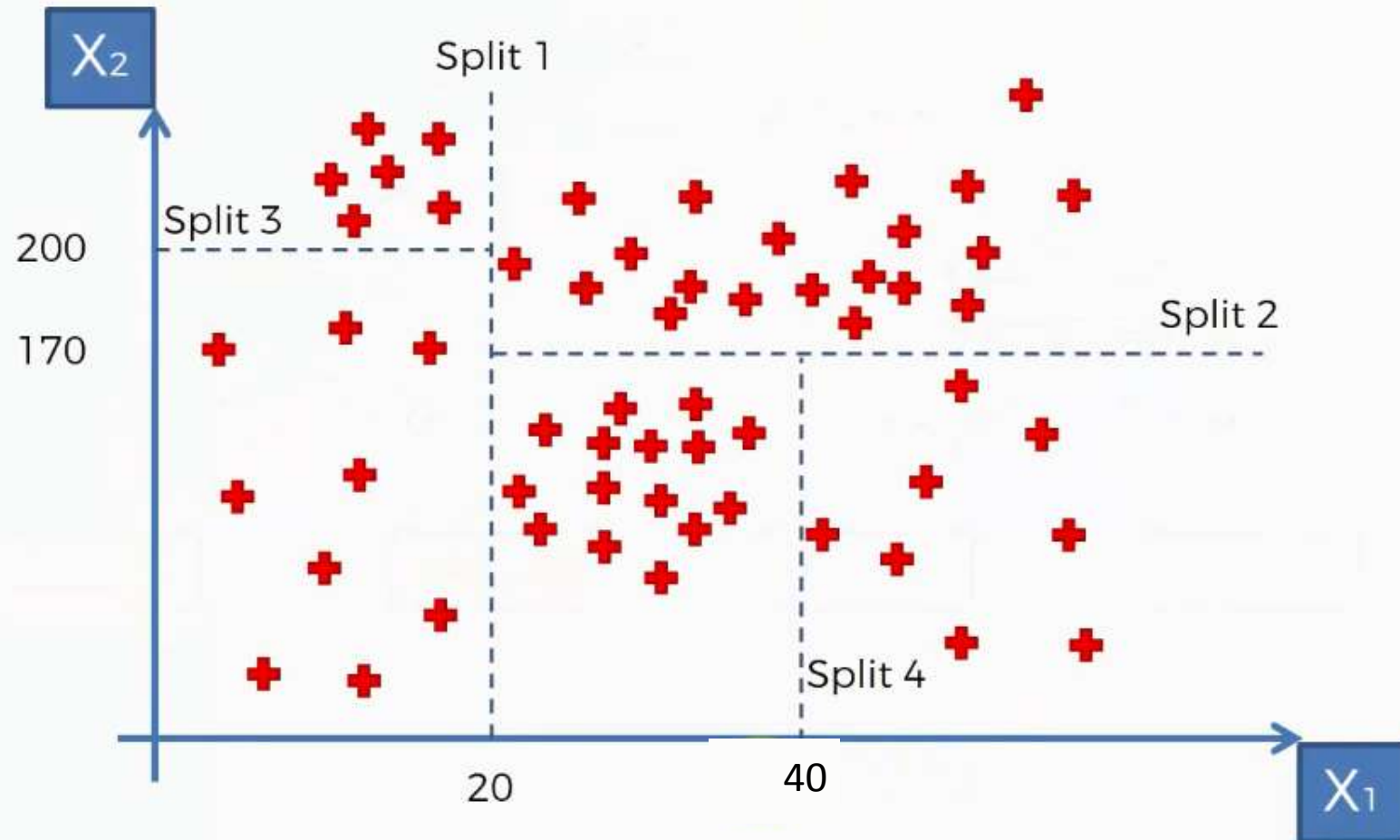
Decision Tree Intuition



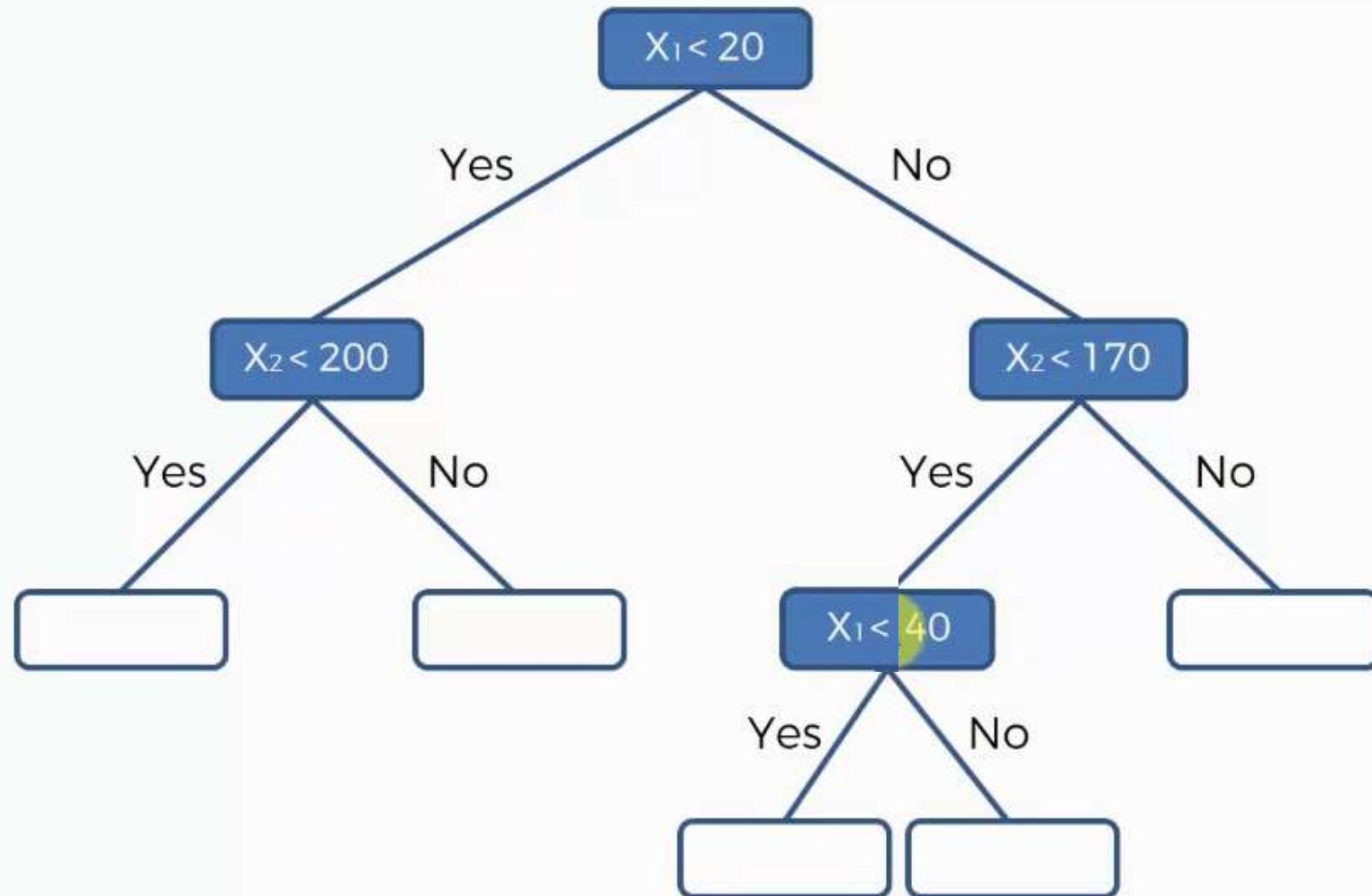
Decision Tree Intuition



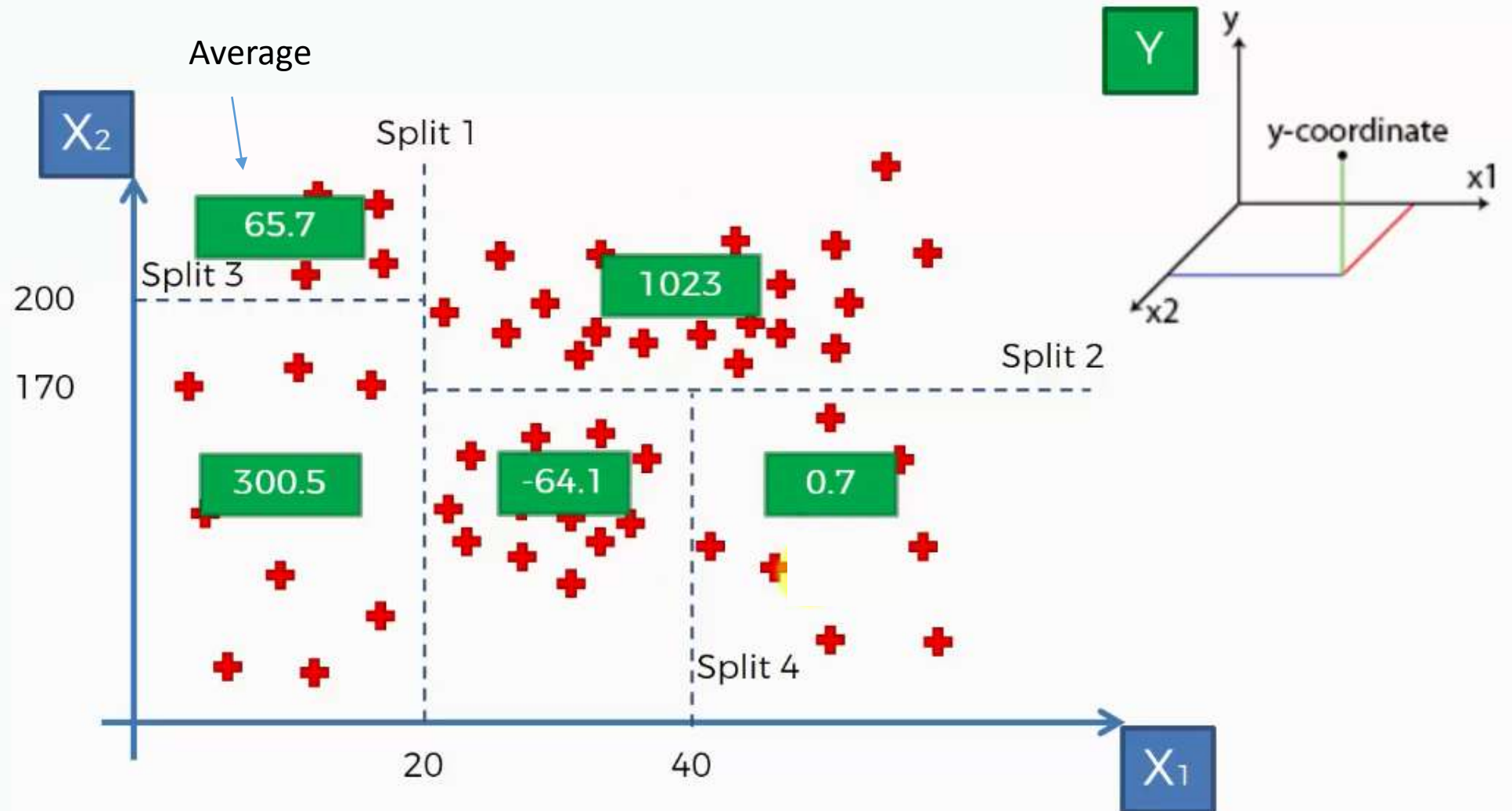
Decision Tree Intuition



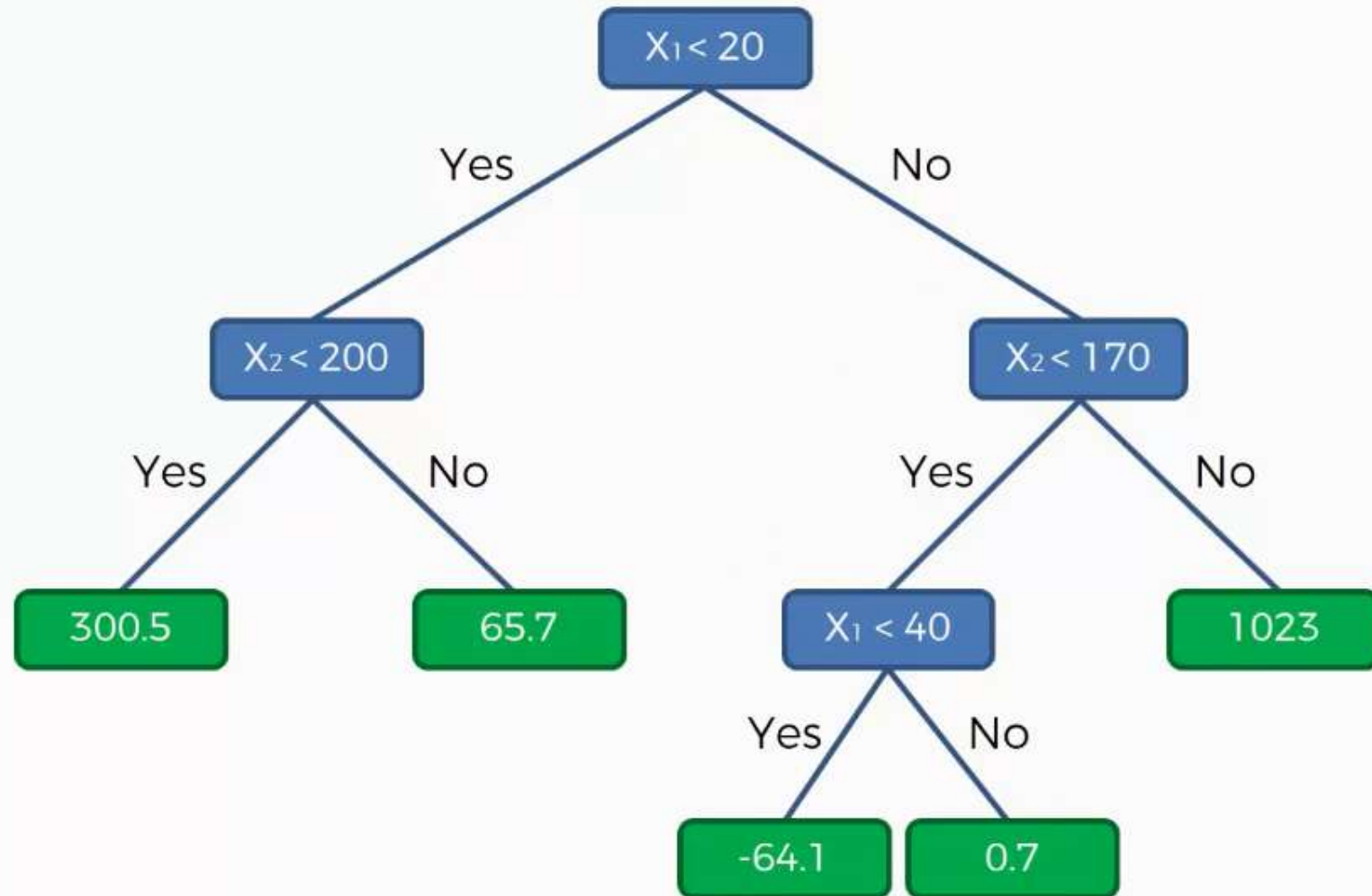
Decision Tree Intuition



Decision Tree Intuition



Decision Tree Intuition



Demo: Create ML using Decision Tree Regression to predict the salary based on the position.





Random Forest Regression

Random Forest Intuition (Ensemble Learning)



Random Forest Intuition (Ensemble Learning)

STEP 1: Pick at random K data points from the Training set.



STEP 2: Build the Decision Tree associated to these K data points.



STEP 3: Choose the number N_{tree} of trees you want to build and repeat STEPS 1 & 2



STEP 4: For a new data point, make each one of your N_{tree} trees predict the value of Y to for the data point in question, and assign the new data point the average across all of the predicted Y values.

Demo: Create ML using Random Forest Regression to predict the salary based on the position.



Demo: Create ML Regression model Templates.

