DATA SCIENCE COURSE

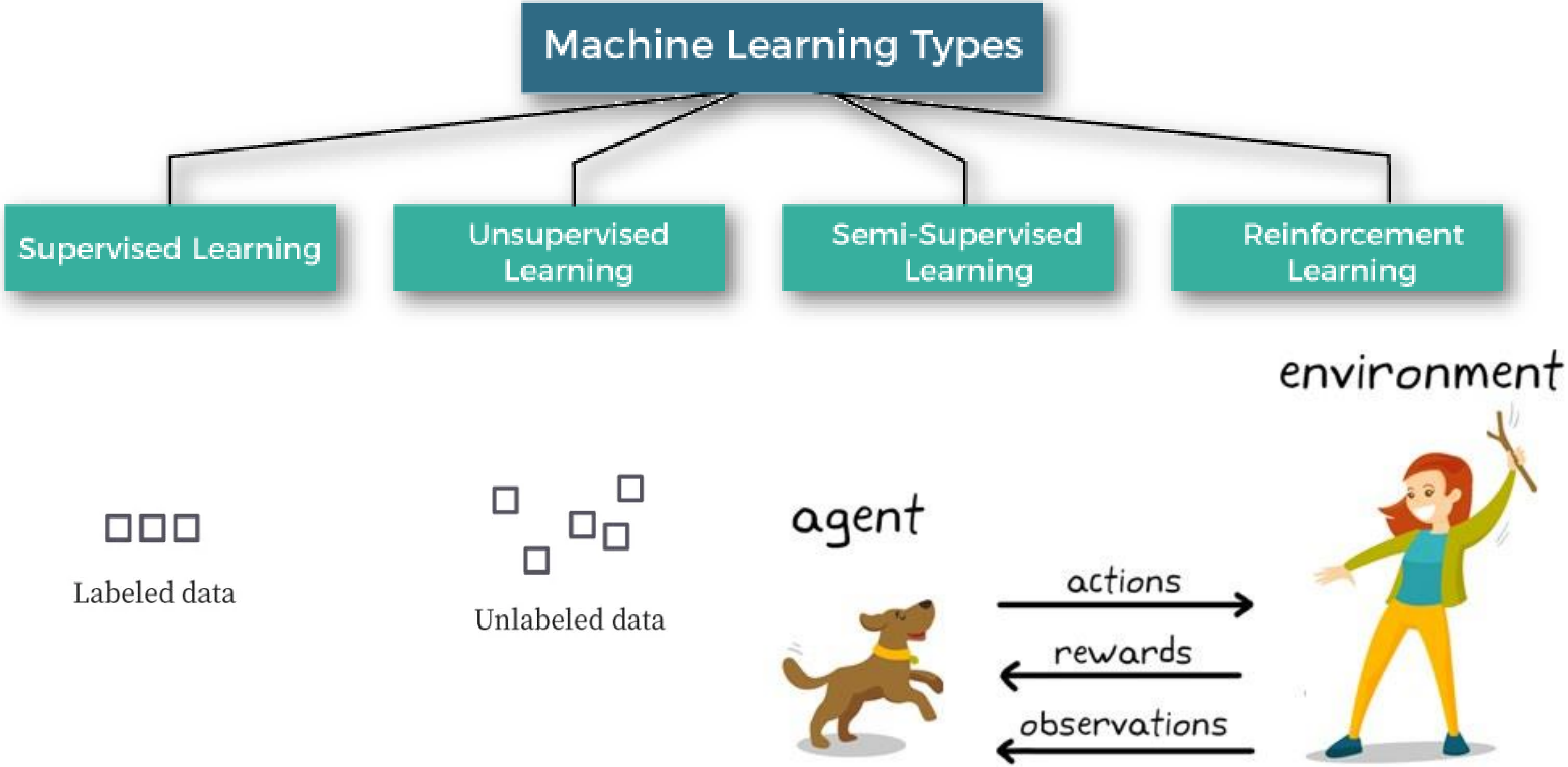
MACHINE LEARNING DAY 01

WHAT IS MACHINE LEARNING?

Experience

=

Data



Machine Learning Types
Supervised Learning
Unsupervised Learning
Semi-Supervised Learning
Reinforcement Learning
Labeled data
Unlabeled data
agent
environment
actions
rewards
observations

Splitting The Data

Model Evaluation and Validation

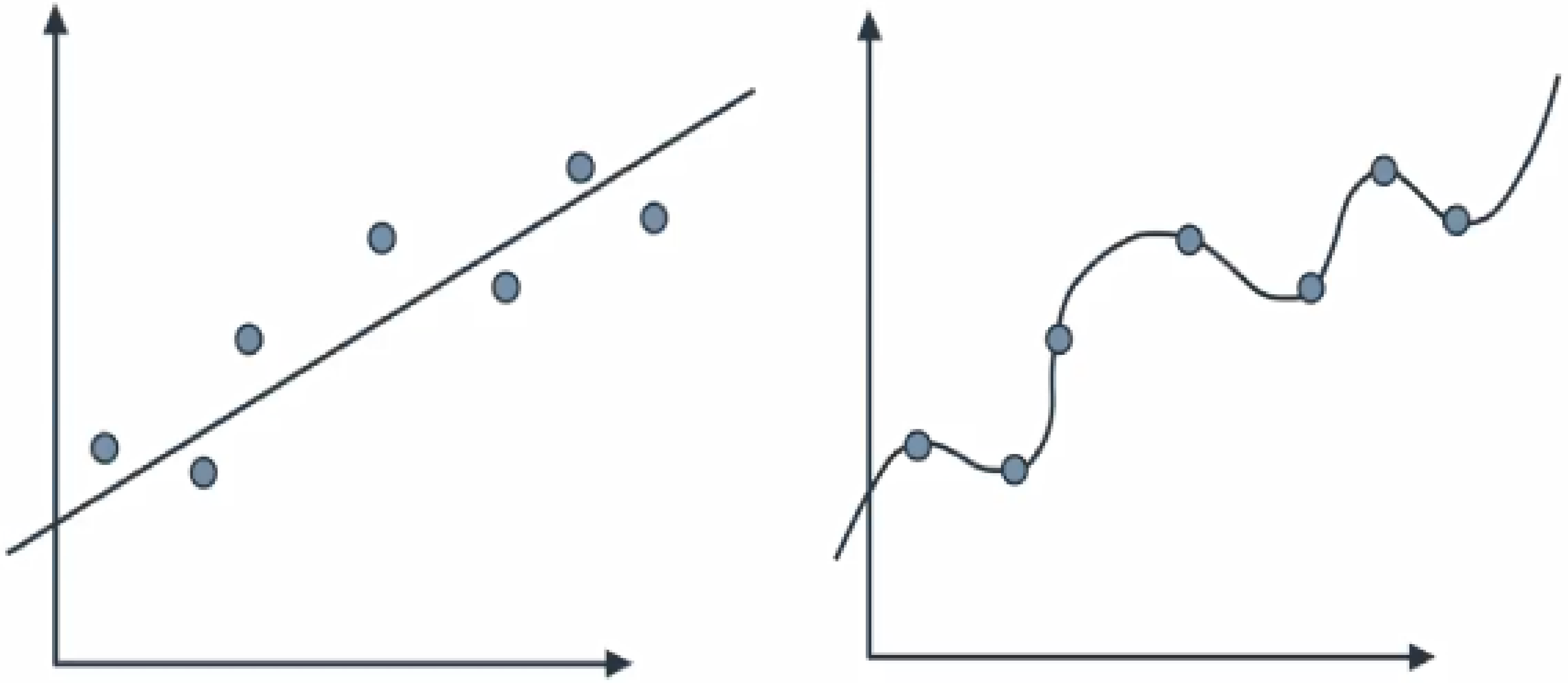
- How well is my model doing?
- is this model good or not?
- How do we improve it based on these metrics?

PROBLEM

TOOLS

MEASUREMENT TOOLS

WHICH MODEL IS BETTER?



Evaluation Metrics

Confusion Matrix

SPAM CLASSIFIER MODEL

NOT SPAM

SPAM

CONFUSION MATRIX

10, 000 PATIENTS

DIAGNOSIS

PATIENTS	Diagnosed Sick	Diagnosed Healthy
Sick	1000 True Positives	200 False Negatives
Healthy	800 False Positives	8000 True Negatives

CONFUSION MATRIX

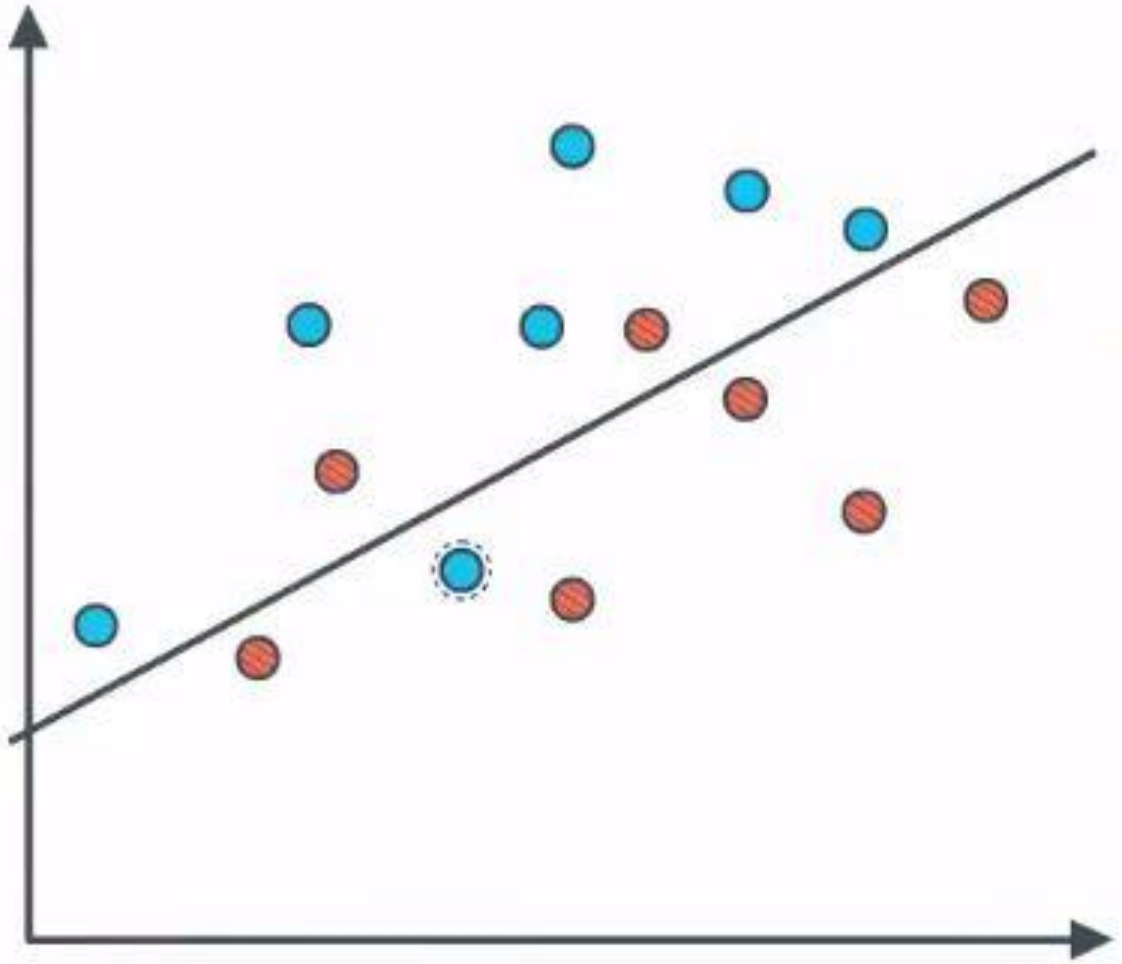
1000 EMAILS

SPAM

EMAIL

	Spam Folder	Inbox
Spam	100 True Positives	170 False Negatives
Not Spam	30 False Positives	700 True Negatives

CONFUSION MATRIX



	Guessed Positive	Guessed Negative
Positive	6 True Positives	1 False Negatives
Negative	2 False Positives	5 True Negatives

ACCURACY

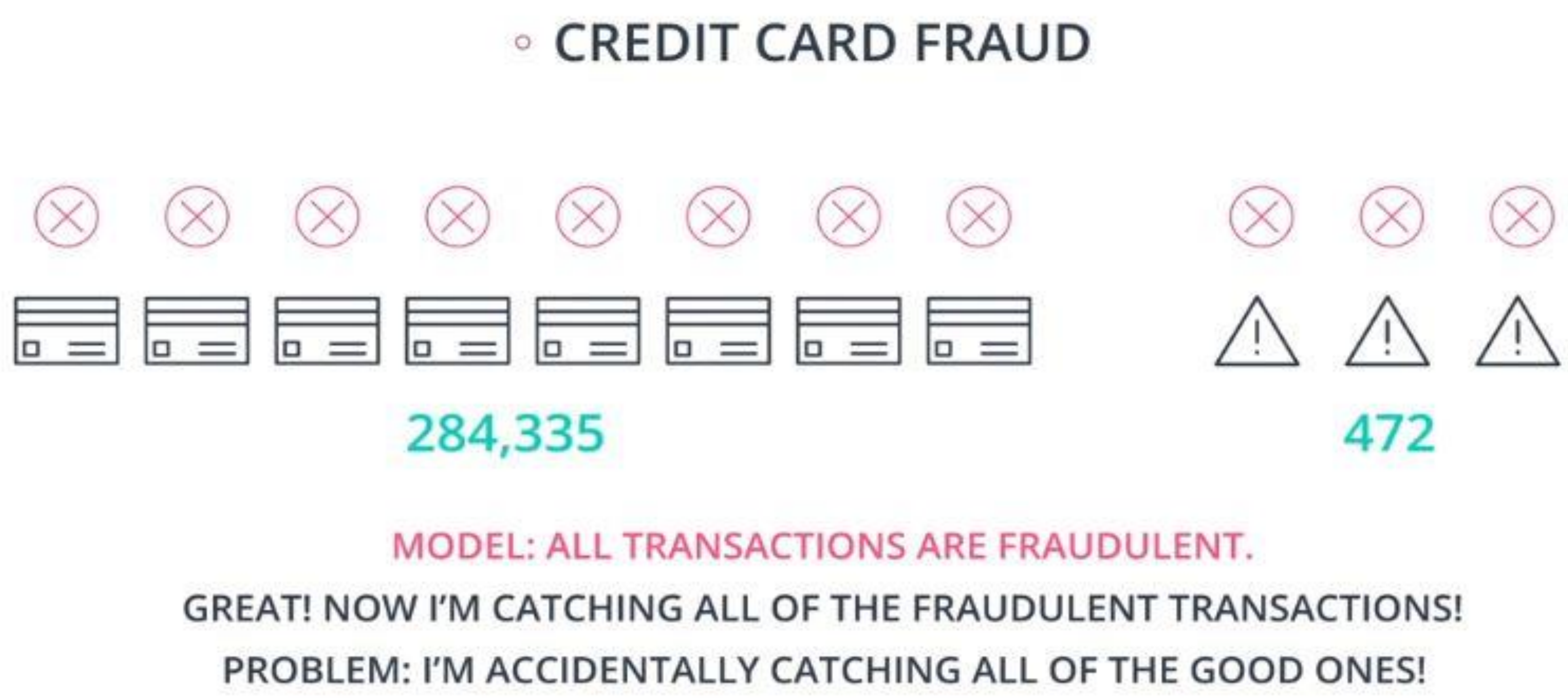
Out of all the patients , how many did we classify correctly ?

	Spam folder	Inbox
Spam	100	170
Not spam	30	700

$$\text{Accuracy} = \frac{100 + 700}{1{,}000} = 80\%$$

```
from sklearn.metrics import accuracy_score
accuracy_score (y_true, y_pred)
```

When accuracy won't work

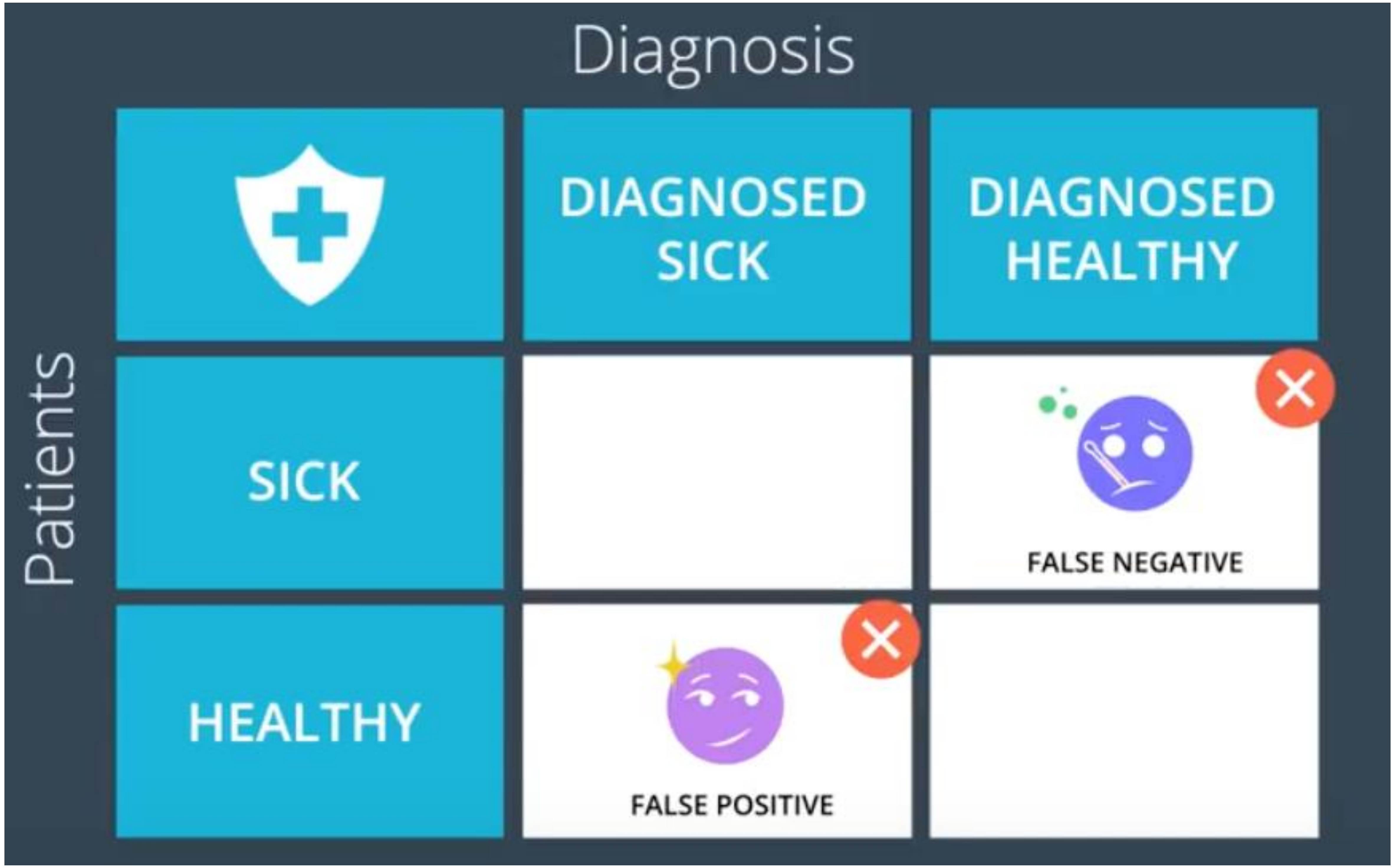


Can you help me think of a model that has over 99 percent accuracy?

QUESTION 1 OF 2

False Negative

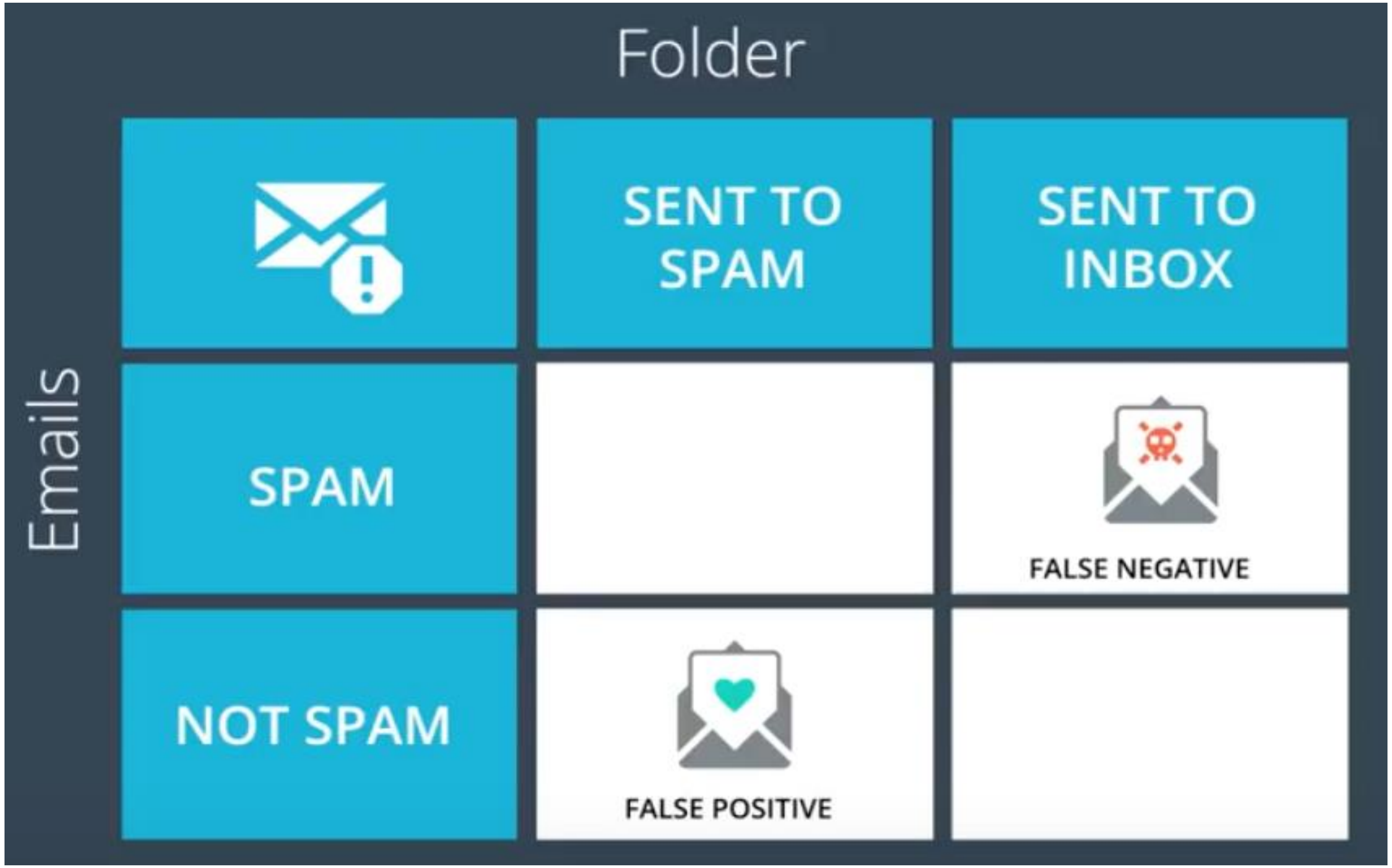
- In the medical example, what is worse, a False Positive, or a False Negative?



QUESTION 2 OF 2

False positive

In the spam detector example, what is worse, a False Positive, or a False Negative?



Medical Model

FALSE POSITIVES OK

FALSE NEGATIVES NOT OK

OK IF NOT ALL ARE SICK
FIND ALL THE SICK PEOPLE

HIGH RECALL

Spam Detector

FALSE POSITIVES NOT OK

FALSE NEGATIVES OK

DON'T NECESSARILY NEED
TO FIND ALL THE SPAM

HIGH PRECISION

PRECISION

FOLDER

EMAIL

	Sent to Spam Folder	Sent to Inbox
Spam	100	170
Not Spam	30	700

OUT OF ALL THE E-MAILS
SENT TO THE SPAM FOLDER,
HOW MANY WERE ACTUALLY SPAM?

$$\text{PRECISION} = \frac{100}{100 + 30} = 76.9\%$$

◦ RECALL

FOLDER

EMAIL

	Sent to Spam Folder	Sent to Inbox
Spam	100	170
Not Spam	30	700

OUT OF ALL THE SPAM E-MAILS, HOW MANY WERE CORRECTLY SENT TO THE SPAM FOLDER?

$$\text{Recall} = \frac{100}{100 + 170} = 37\%$$

PRECISION AND RECALL

ONE SCORE?

MEDICAL MODEL	SPAM DETECTOR
PRECISION: 55.7%	PRECISION: 76.9%
RECALL: 83.3%	RECALL: 37%
AVERAGE = 69.5%	AVERAGE = 56.95%

F1 SCORE

Y

Arithmetic Mean = X+Y /2

Harmonic Mean = 2XY /X+Y

X

Precision = 1
Recall = 0
Average = 0.5
Harmonic Mean = 0

Precision = 0.2
Recall = 0.8
Average = 0.5
Harmonic Mean = 0.32

Arithmetic Mean(Precision , Recall)

F1 score = Harmonic Mean(Precision , Recall)

CREDIT CARD FRAUD

MODEL: ALL TRANSACTIONS ARE GOOD.

PRECISION = 100%

F_1 SCORE = 0

RECALL = 0%

F-BETA SCORE

- F0.5-Measure (beta=0.5): More weight on precision, less weight on recall
- F1-Measure (beta=1.0): Balance the weight on precision and recall.
- F2-Measure (beta=2.0): Less weight on precision, more weight on recall

beta = 1.0

Precision ⟷ **Recal**

beta = 0.5 F1 Score beta = 2 **l**

Boundaries in the F-beta score

Note that in the formula for F_β score, if we set $\beta = 0$, we get

$F_0 = (1 + 0^2) \cdot \frac{\text{Precision}\cdot\text{Recall}}{0\cdot\text{Precision}+\text{Recall}} = \frac{\text{Precision}\cdot\text{Recall}}{\text{Recall}} = \text{Precision}$. Therefore, the minimum value of β is zero, and at this value, we get the precision.

Now, notice that if N is really large, then

$F_\beta = (1 + N^2) \cdot \frac{\text{Precision}\cdot\text{Recall}}{N^2\cdot\text{Precision}+\text{Recall}} = \frac{\text{Precision}\cdot\text{Recall}}{\frac{N^2}{1+N^2}\text{Precision}+\frac{1}{1+N^2}\text{Recall}}$.

As N goes to infinity, we can see that $\frac{1}{1+N^2}$ goes to zero, and $\frac{N^2}{1+N^2}$ goes to 1.

Therefore, if we take the limit, we have

$\lim_{N\to\infty} F_N = \frac{\text{Precision}\cdot\text{Recall}}{1\cdot\text{Precision}+0\cdot\text{Recall}} = \text{Recall}.$

Thus, to conclude, the boundaries of beta are between 0 and ∞.

- If $\beta = 0$, then we get **precision**.
- If $\beta = \infty$, then we get **recall**.
- For other values of β, if they are close to 0, we get something close to precision, if they are large numbers, then we get something close to recall, and if $\beta = 1$, then we get the **harmonic mean** of precision and recall.

RECEIVER OPERATING CHARACTERISTIC CURVE (ROC)

is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

- True Positive Rate
- False Positive Rate

RECEIVER OPERATING CHARACTERISTIC CURVE (ROC)

True Positive Rate (TPR) is a synonym for recall and is therefore defined as follows:

$$TPR = \frac{TP}{TP + FN}$$

Sensitivity

False Positive Rate (FPR) is defined as follows:

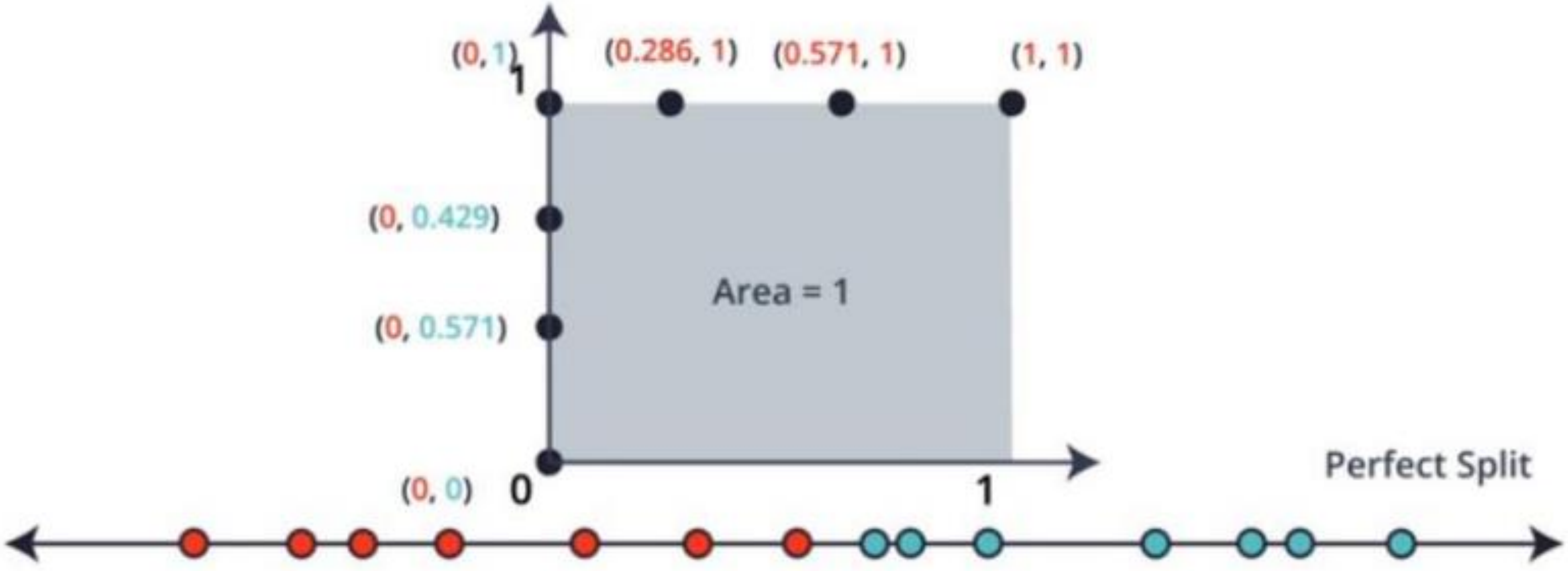
$$FPR = \frac{FP}{FP + TN}$$

Specificity

RECEIVER OPERATING CHARACTERISTIC CURVE (ROC)

$$\text{True Positive Rate} = \frac{\text{TRUE POSITIVES}}{\text{ALL POSITIVES}}$$

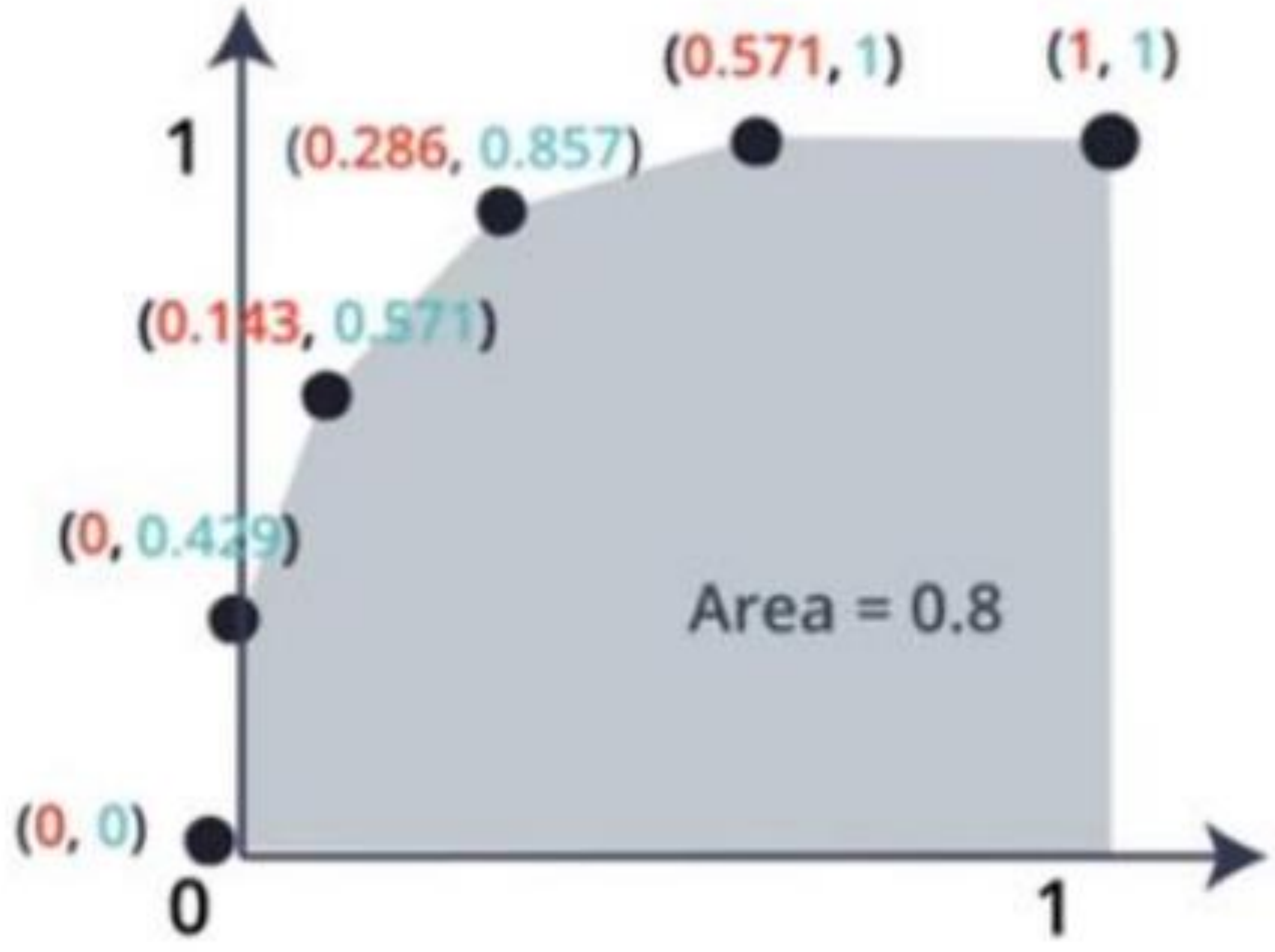
$$\text{False Positive Rate} = \frac{\text{FALSE POSITIVES}}{\text{ALL NEGATIVES}}$$



RECEIVER OPERATING CHARACTERISTIC CURVE (ROC)

$$\text{True Positive Rate} = \frac{\text{TRUE POSITIVES}}{\text{ALL POSITIVES}}$$

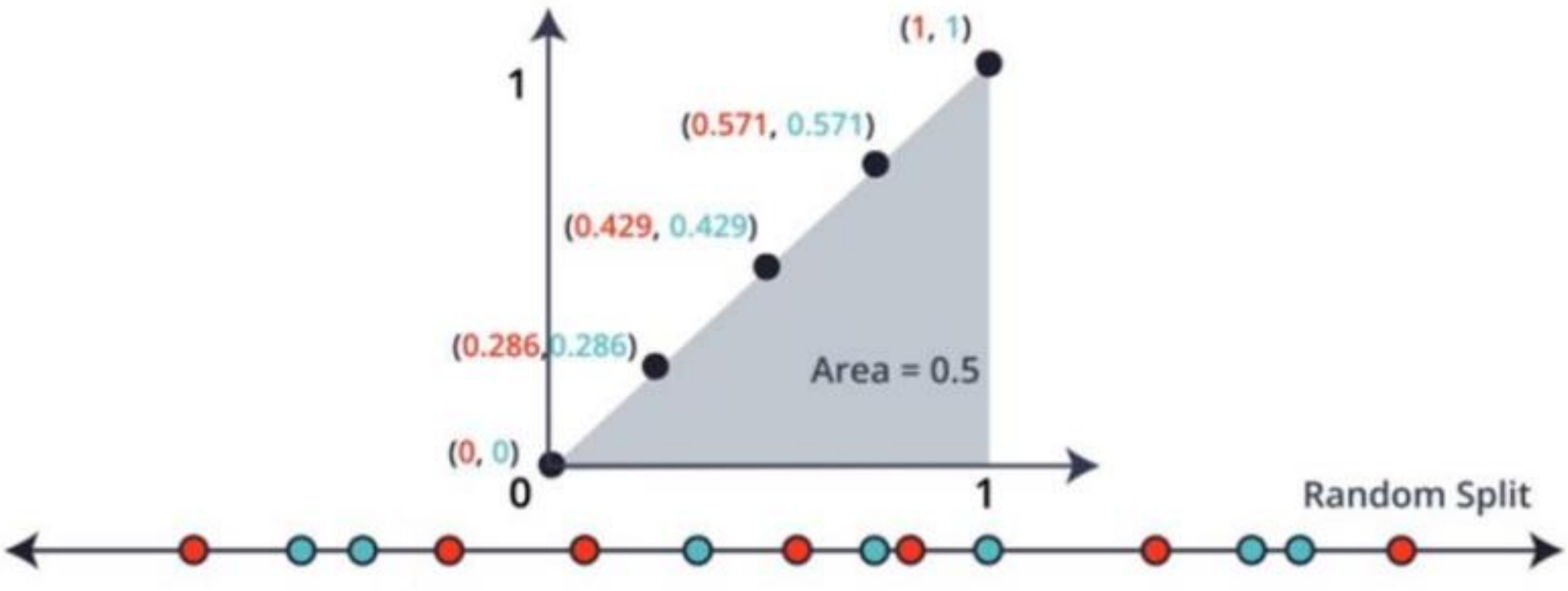
$$\text{False Positive Rate} = \frac{\text{FALSE POSITIVES}}{\text{ALL NEGATIVES}}$$



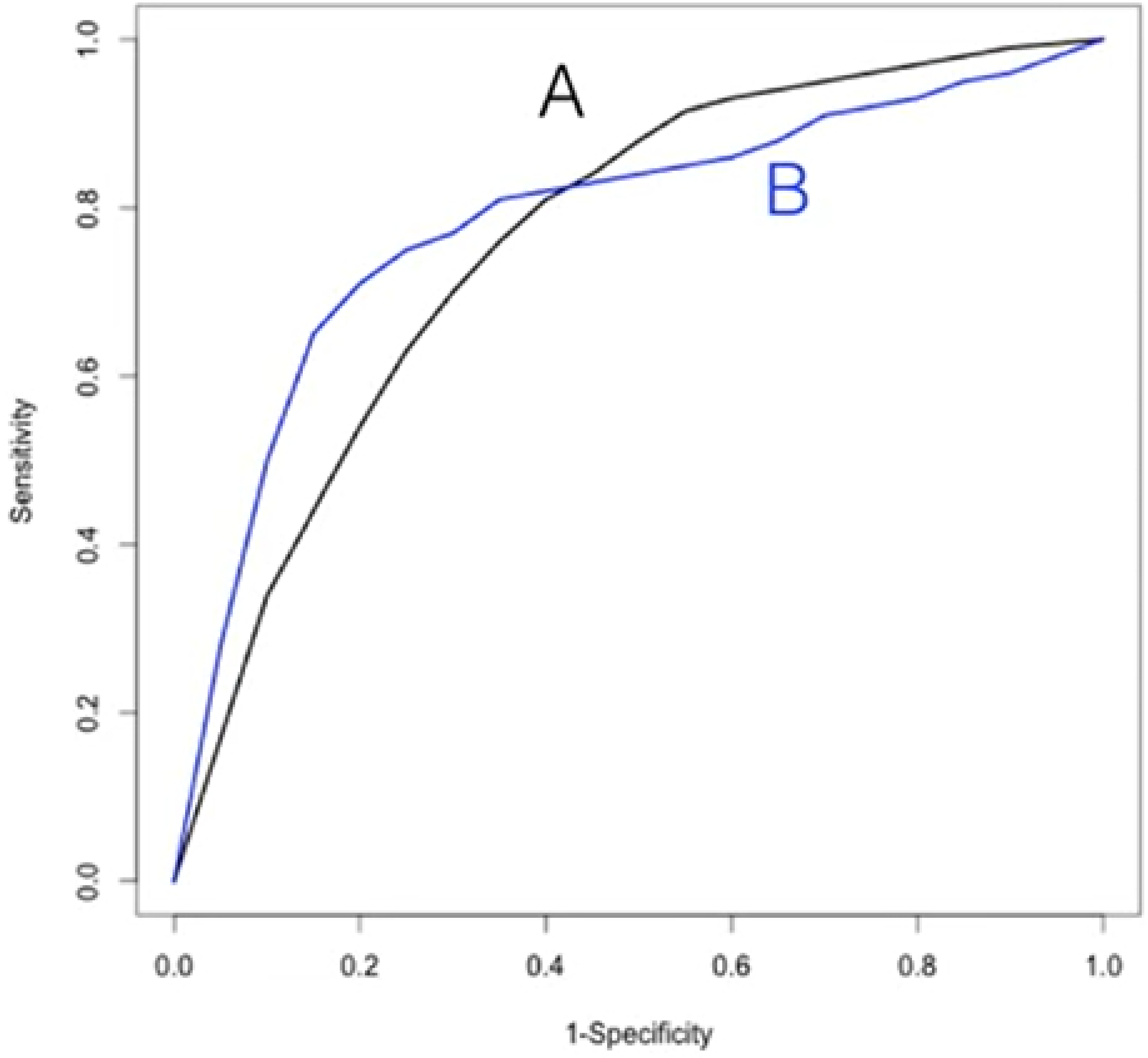
RECEIVER OPERATING CHARACTERISTIC CURVE (ROC)

$$\text{True Positive Rate} = \frac{\text{TRUE POSITIVES}}{\text{ALL POSITIVES}}$$

$$\text{False Positive Rate} = \frac{\text{FALSE POSITIVES}}{\text{ALL NEGATIVES}}$$

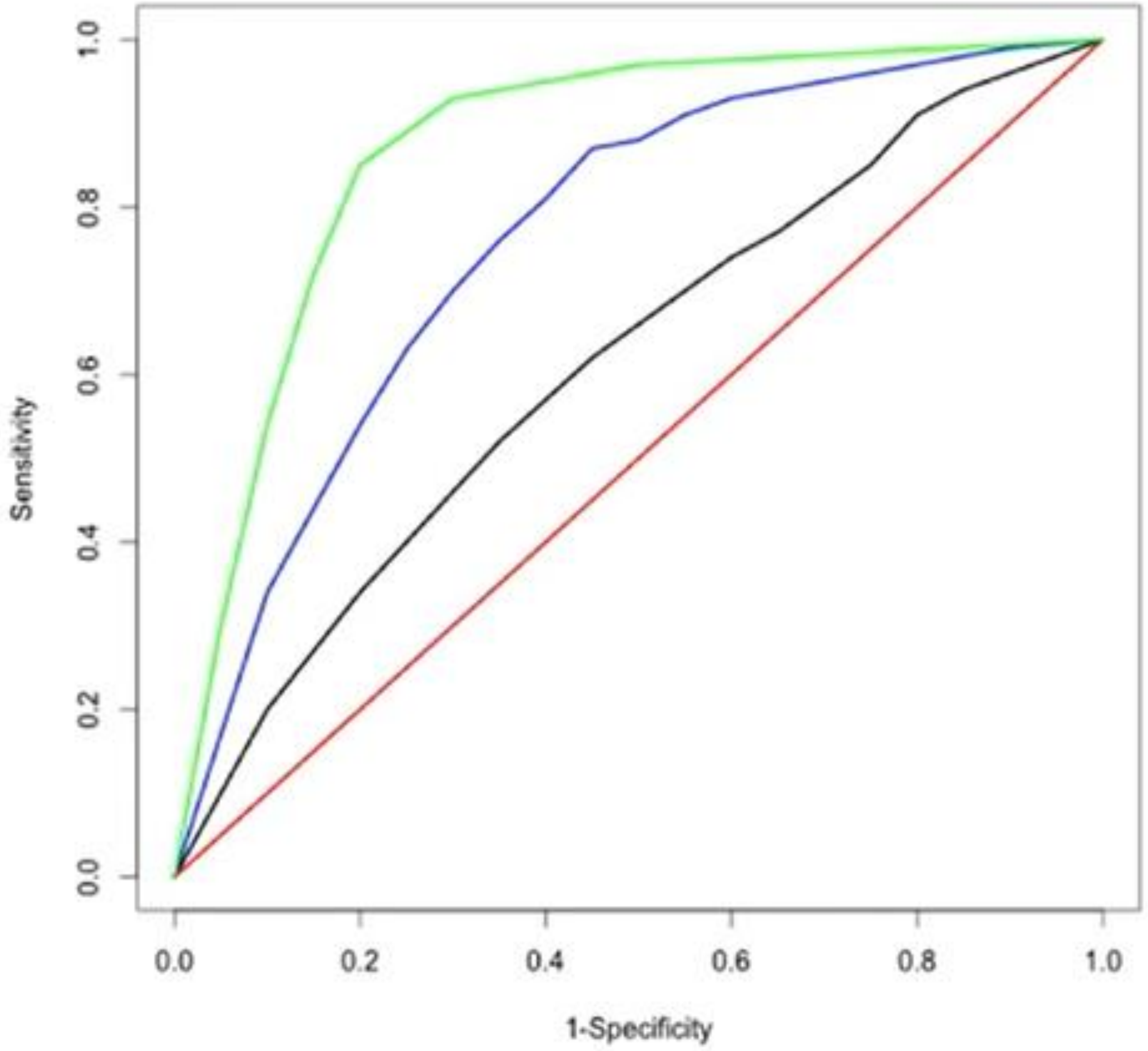


Which one is better?

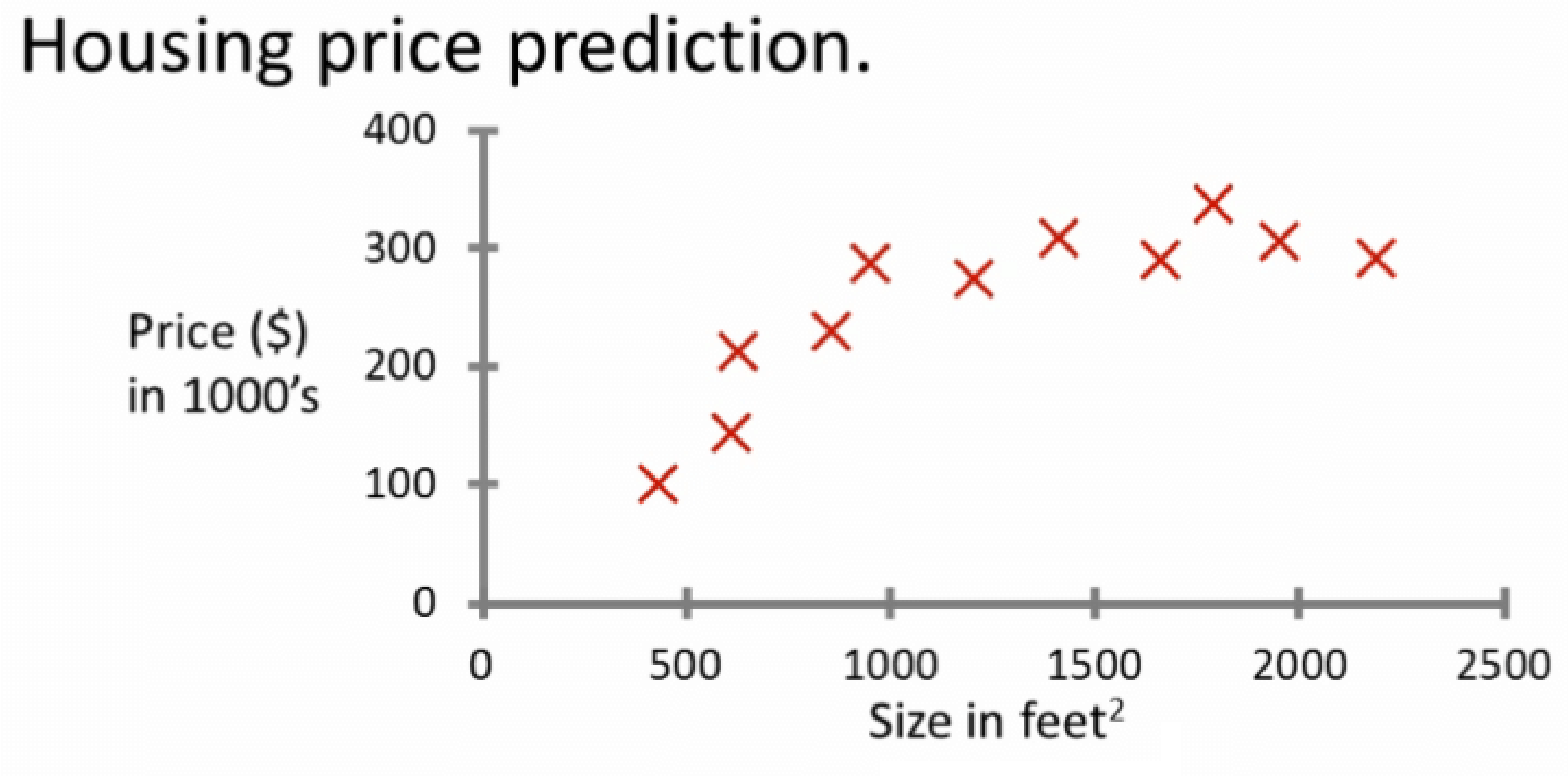


AUC ROC-curve **A** = 0.75

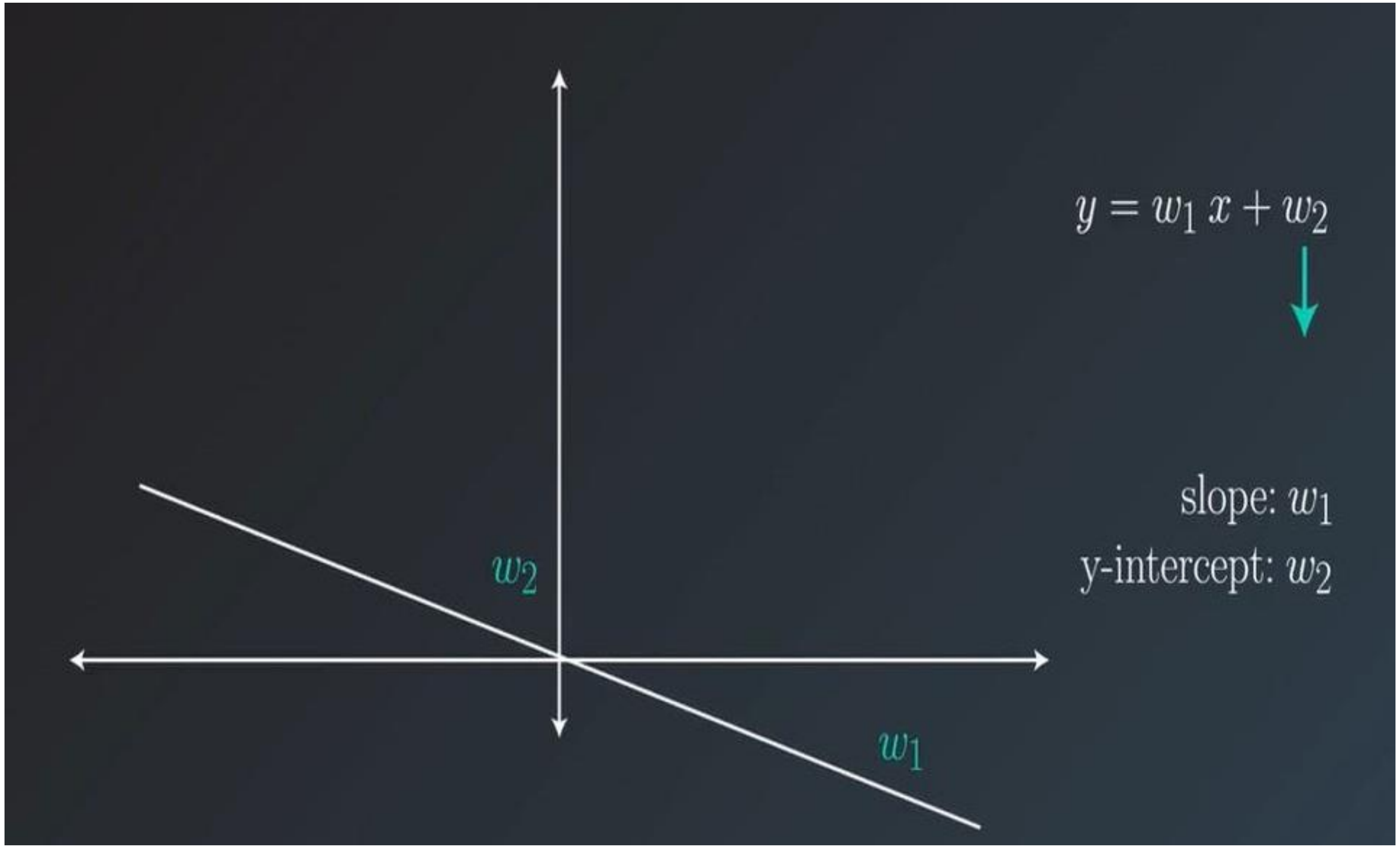
AUC ROC-curve **B** = 0.78



WHAT IS LINEAR REGRESSION?



WHAT IS LINEAR REGRESSION?

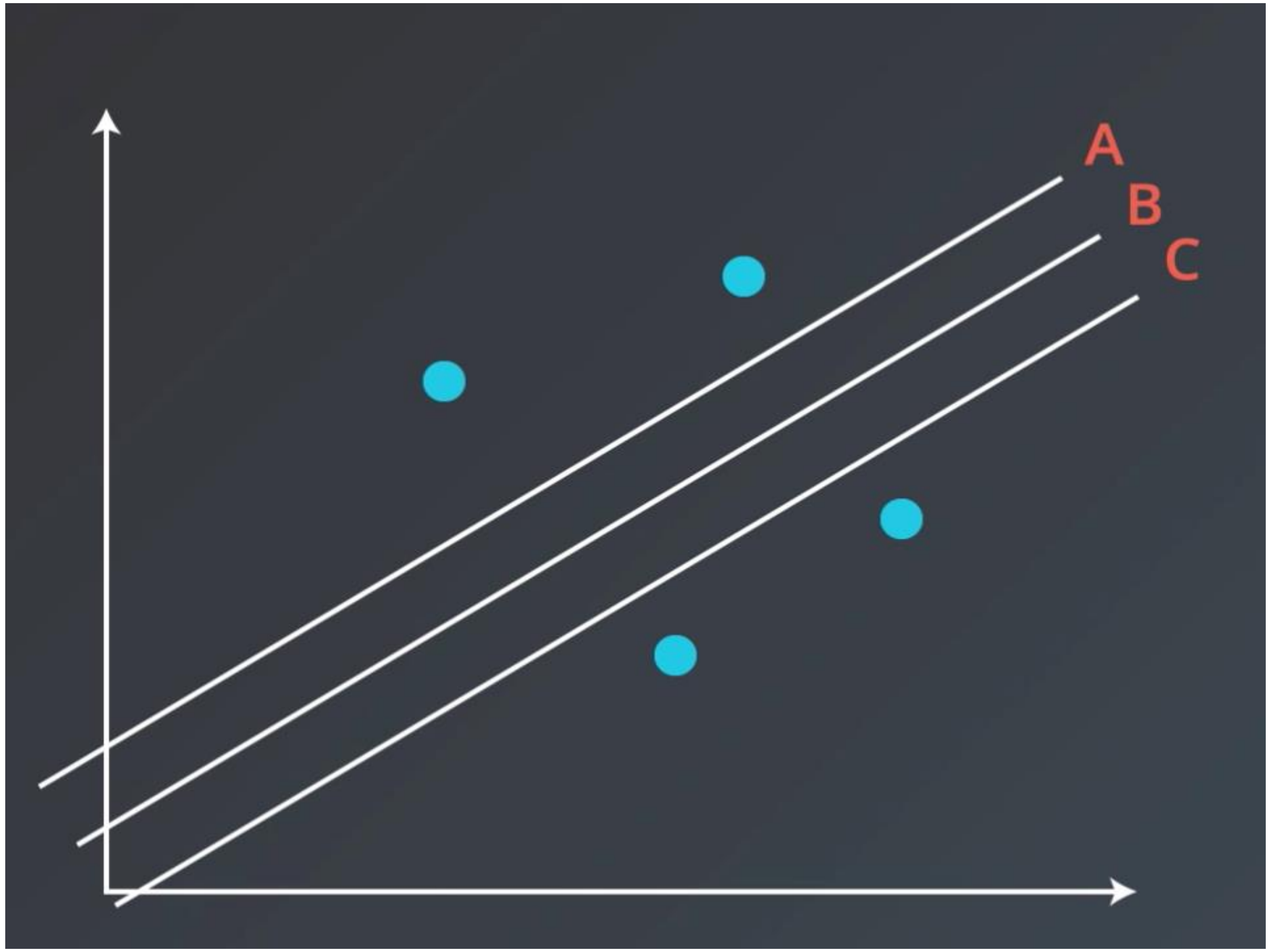


What is Gradient Descent?

It is an Optimization Algorithm to find the Minimum of a Function

Start with a random point on the function and move in the **negative direction** of the **gradient of the function** to reach the **local/global minima**.

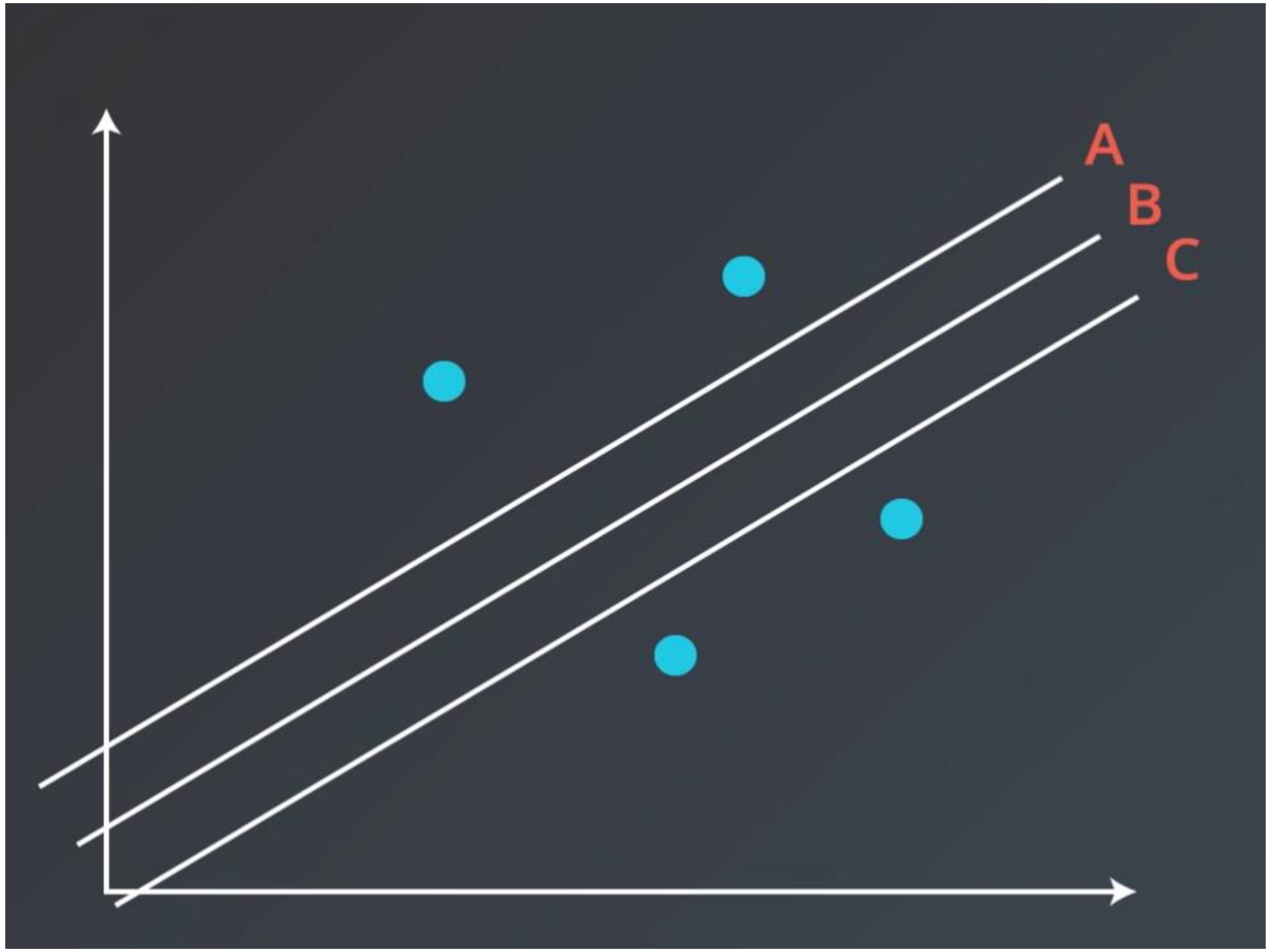
Which of the three lines gives you a smaller Mean Absolute Error?



SOLUTION:

They all give the same error

Which of the three lines gives you a smaller **Mean Squared Error**?

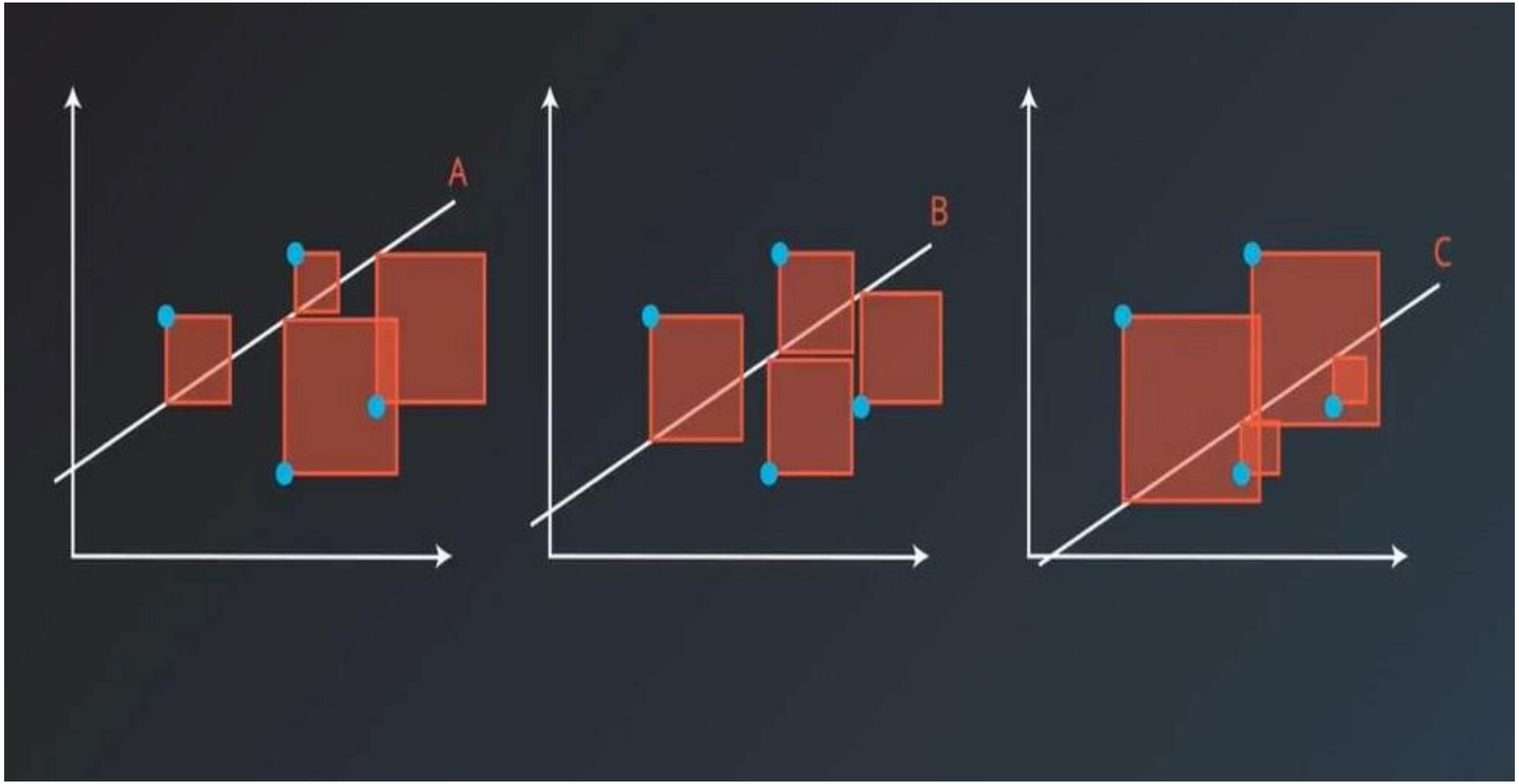


SOLUTION:

B

SOLUTION:

B



R2 SCORE

BAD MODEL

The errors should be similar.

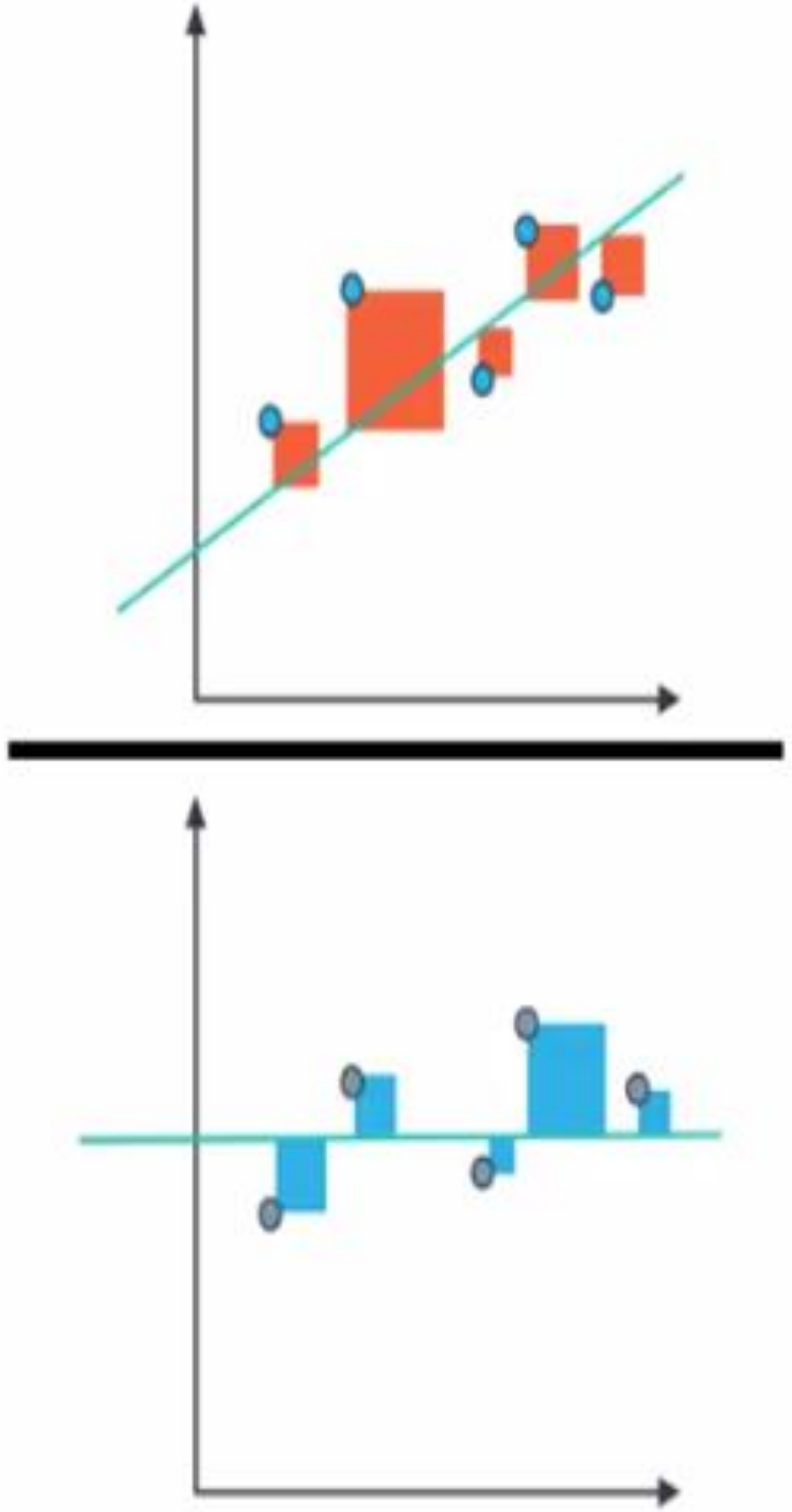
R2 score should be close to 0.

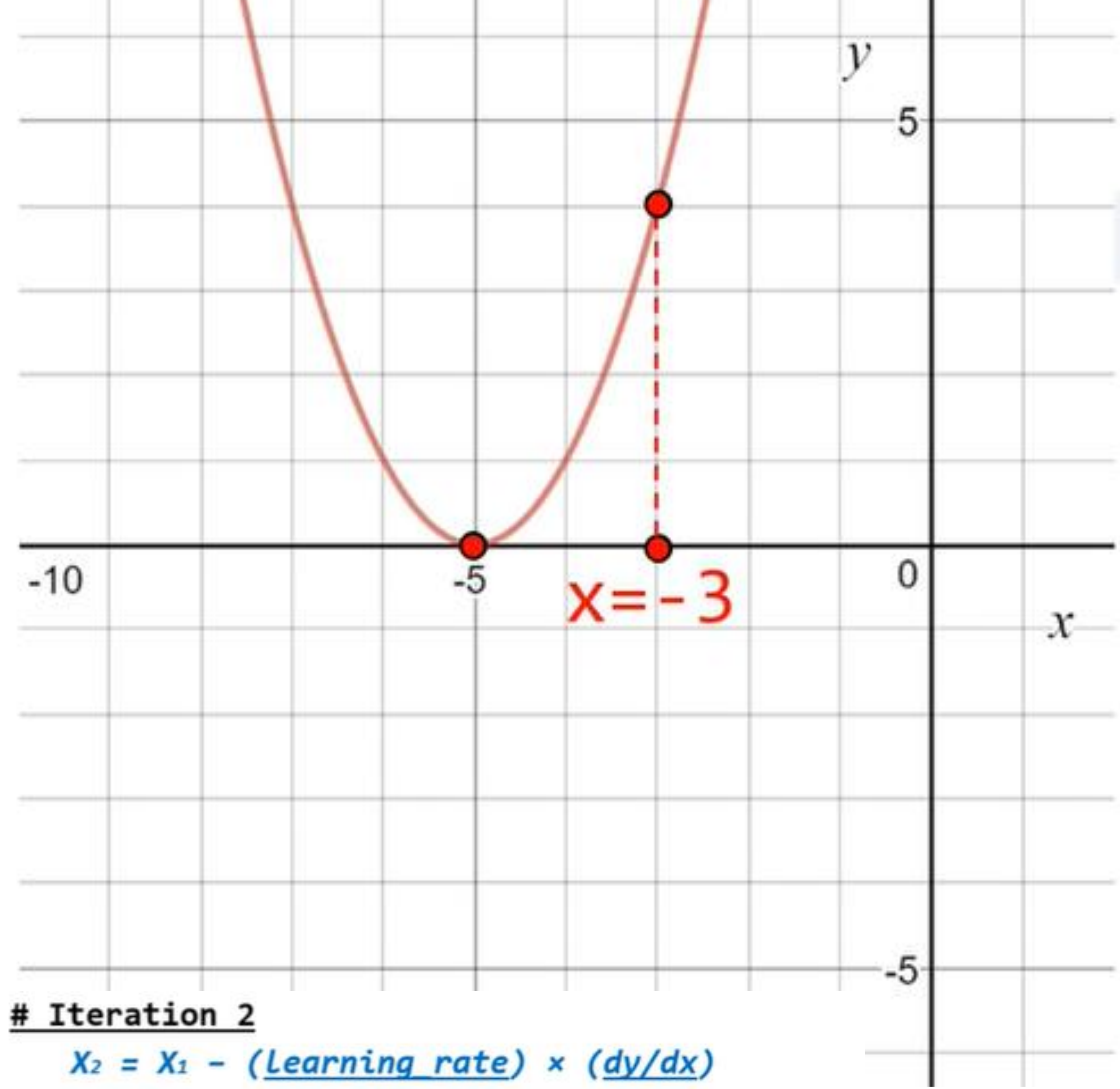
GOOD MODEL

The mean squared error for the linear regression model should be a lot smaller than the mean squared error for the simple model.

R2 score should be close to 1.

R2 = 1 -





Iteration 2

$X_2 = X_1 - (\underline{\text{Learning\_rate}}) \times (\underline{dy/dx})$

$X_1 = (-3.04) - (0.01) \times (2 \times ((-3.04)+5) =$

$y=(x+5)^2$

STEP 1

let's start from random point **X = -3**

then find the gradient of the function,

dy/dx = 2x(X + 5)

STEP 2

move in the direction of the **negative of the gradient**.

But: How much to move**? learning_rate = 0.01**

STEP 3

Perform 2 iterations of gradient descent

initialize parameters

X = -3   *learning_rate = 0.01*   *dy/dx = 2x(X + 5)*

Iteration 1

$X_1 = X_0 - (\underline{\text{Learning\_rate}}) \times (\underline{dy/dx})$

$X_1 = (-3) - (0.01) \times (2 \times ((-3)+5)$

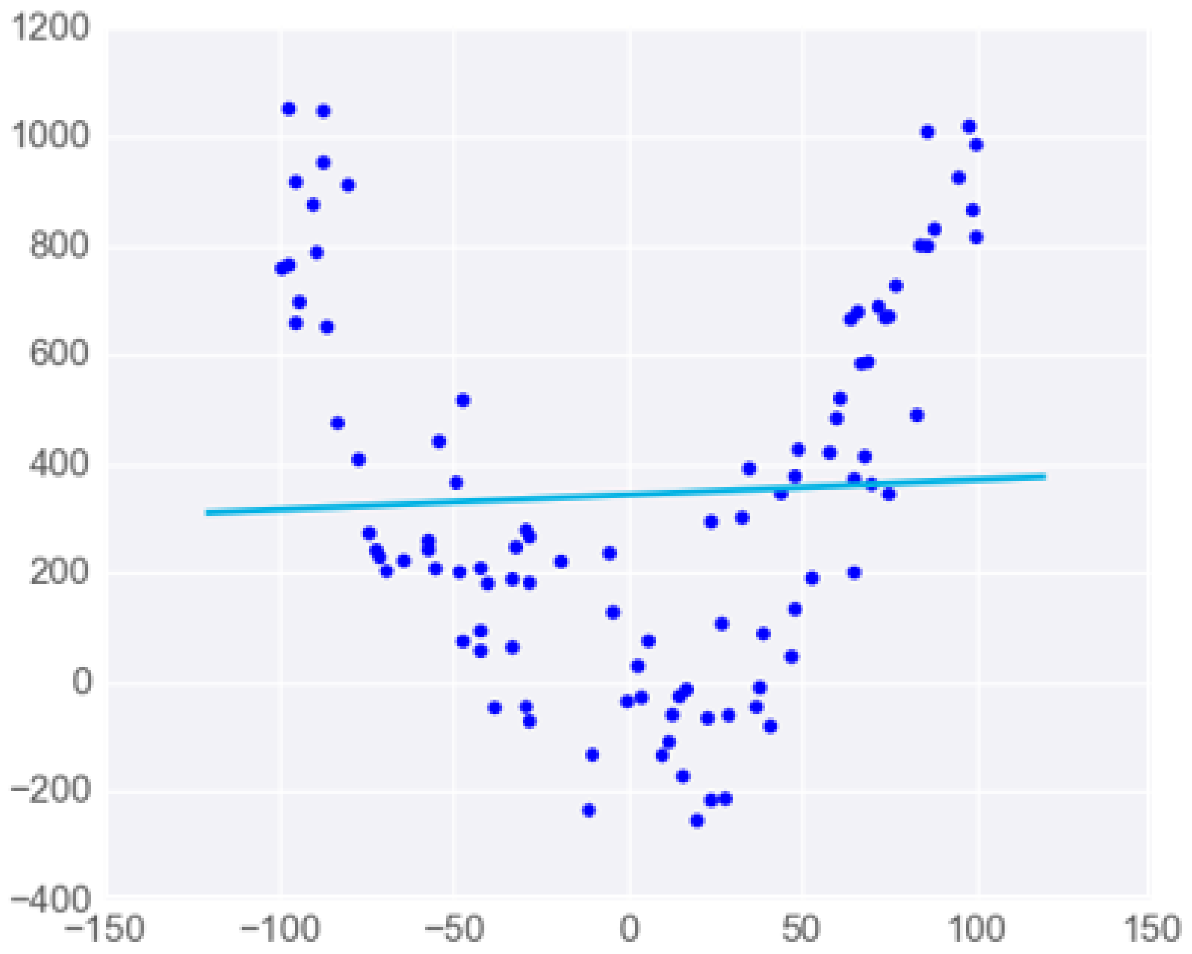
random_point   learning_rate   dy/dx

Linear Regression Warnings!!

Linear Regression Works Best When the Data is Linear

Linear regression produces a straight line model from the training data.

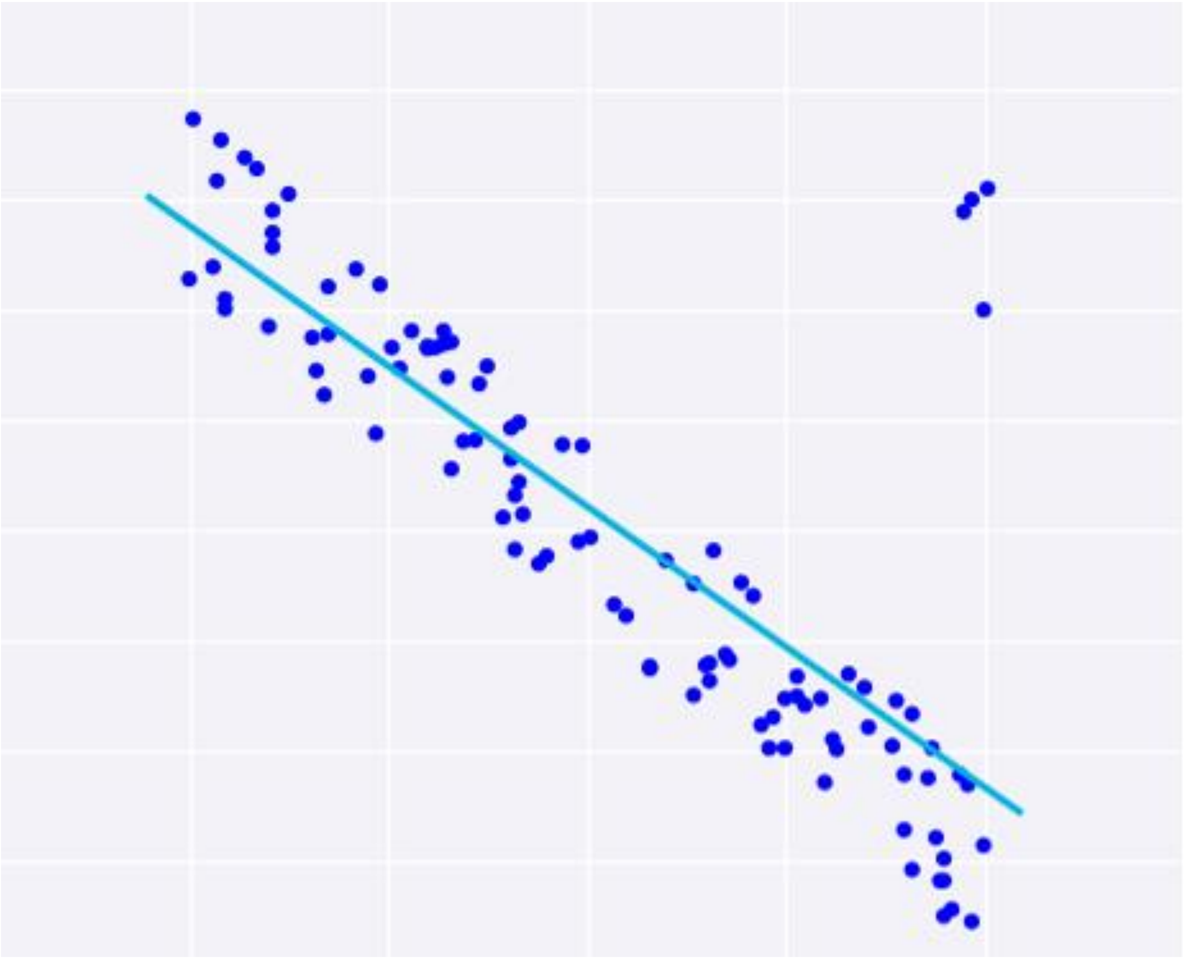
If the relationship in the training data is not really linear, you'll need to either make adjustments (transform your training data), add features, or use another kind of model.



Linear Regression is Sensitive to Outliers

If your dataset has some outlying extreme values that don't fit a general pattern, they can have a surprisingly large effect.

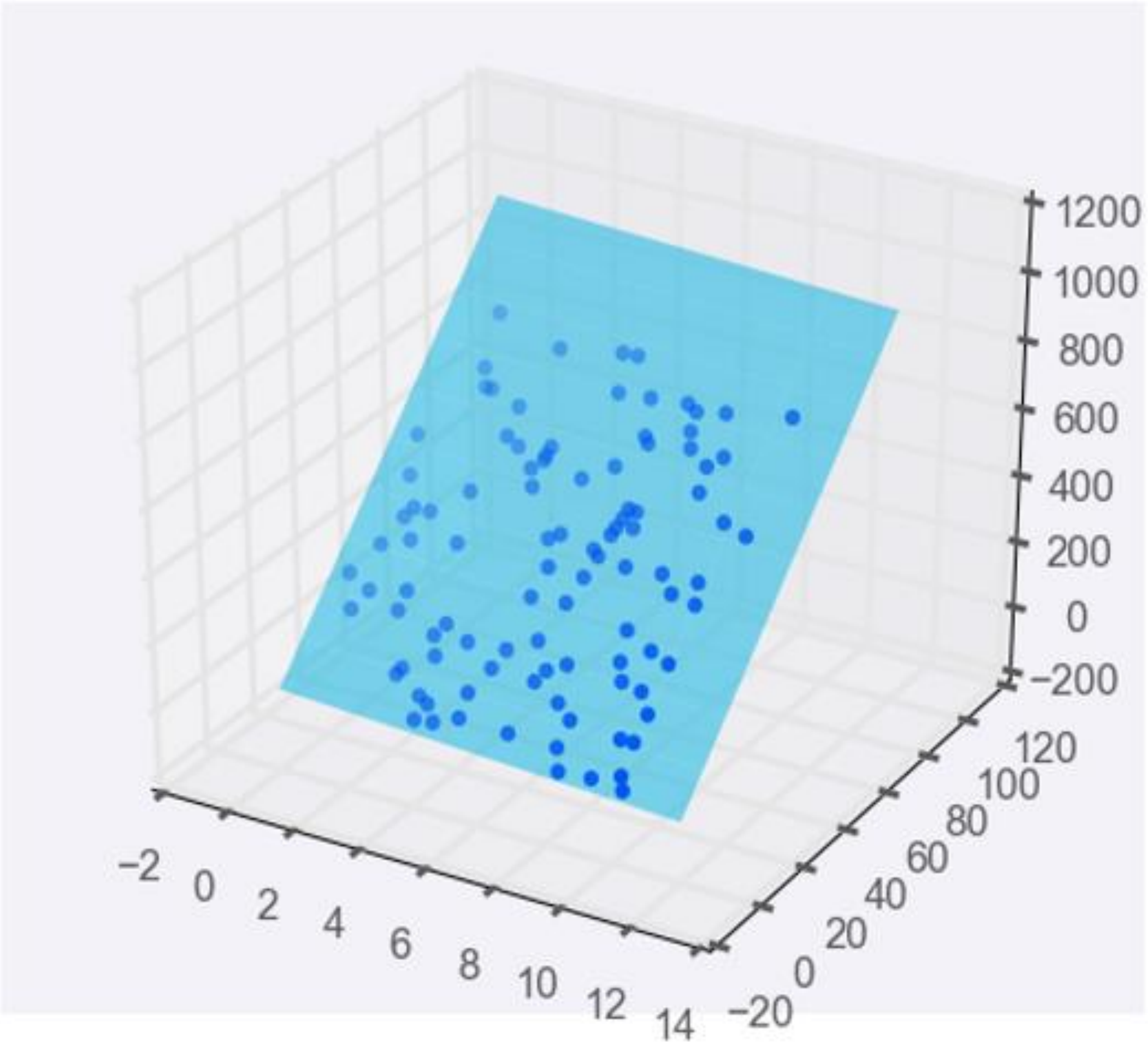
adding a few points that are outliers and don't fit the pattern really changes the way the model predicts.



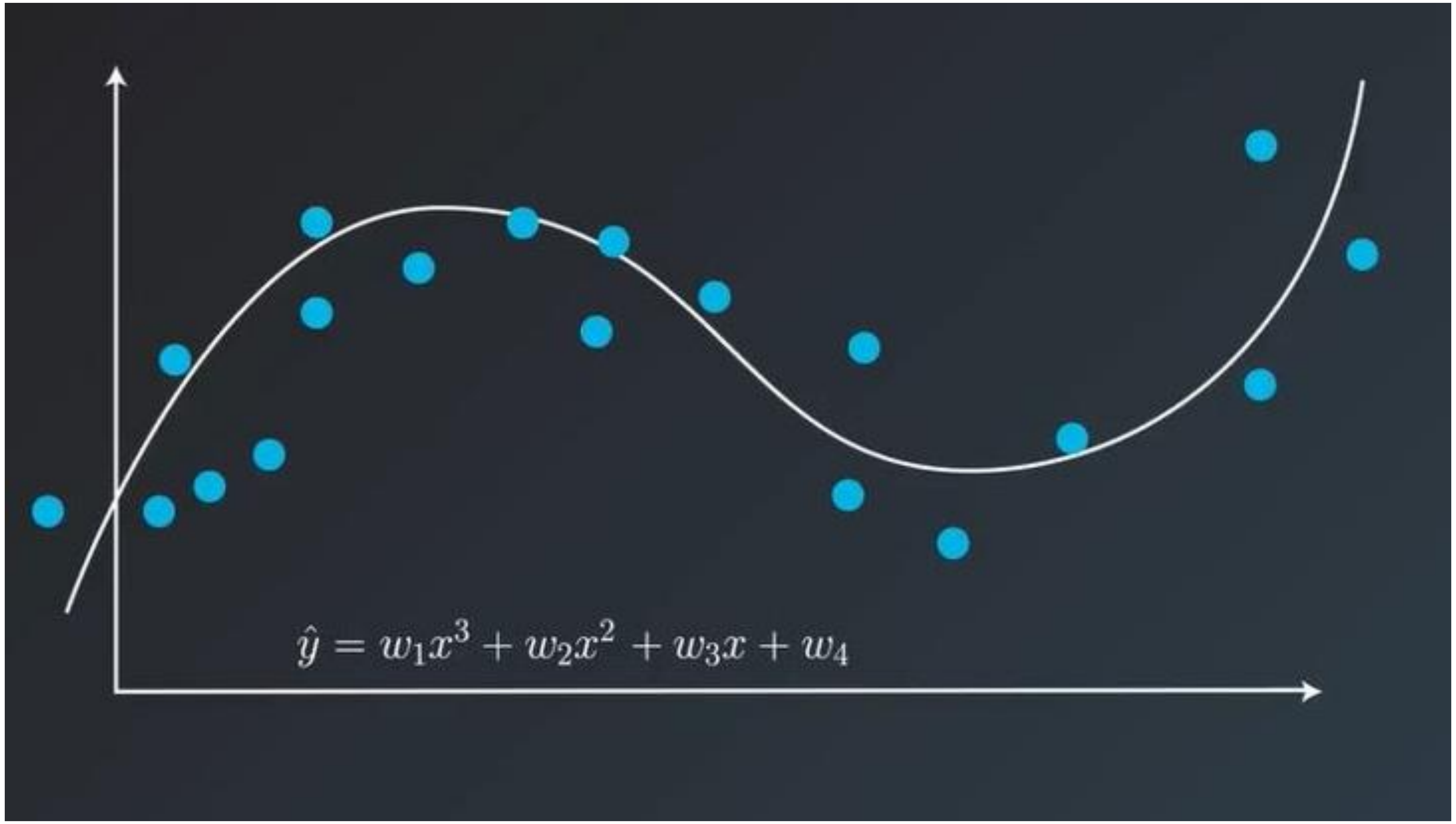
Multiple Linear Regression

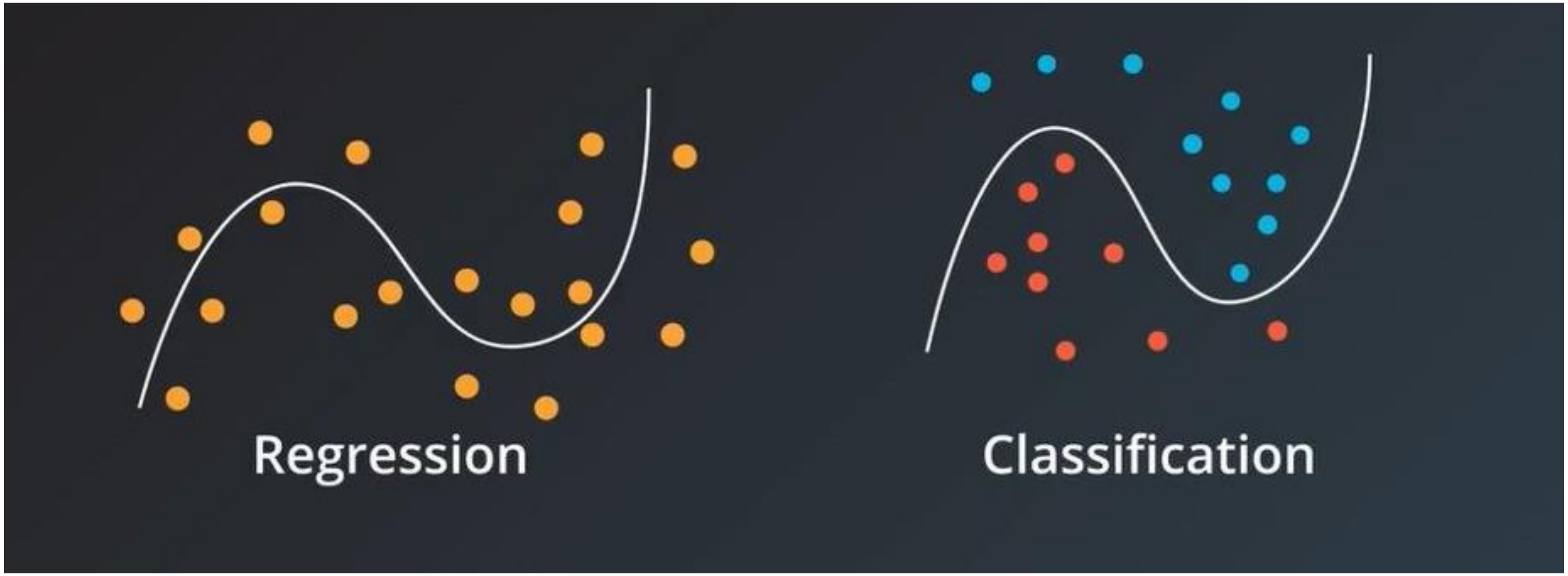
$$y = m_1 x_1 + m_2 x_2 + b$$

To represent this graphically, we'll need a three-dimensional plot, with the linear regression model represented as a plane:

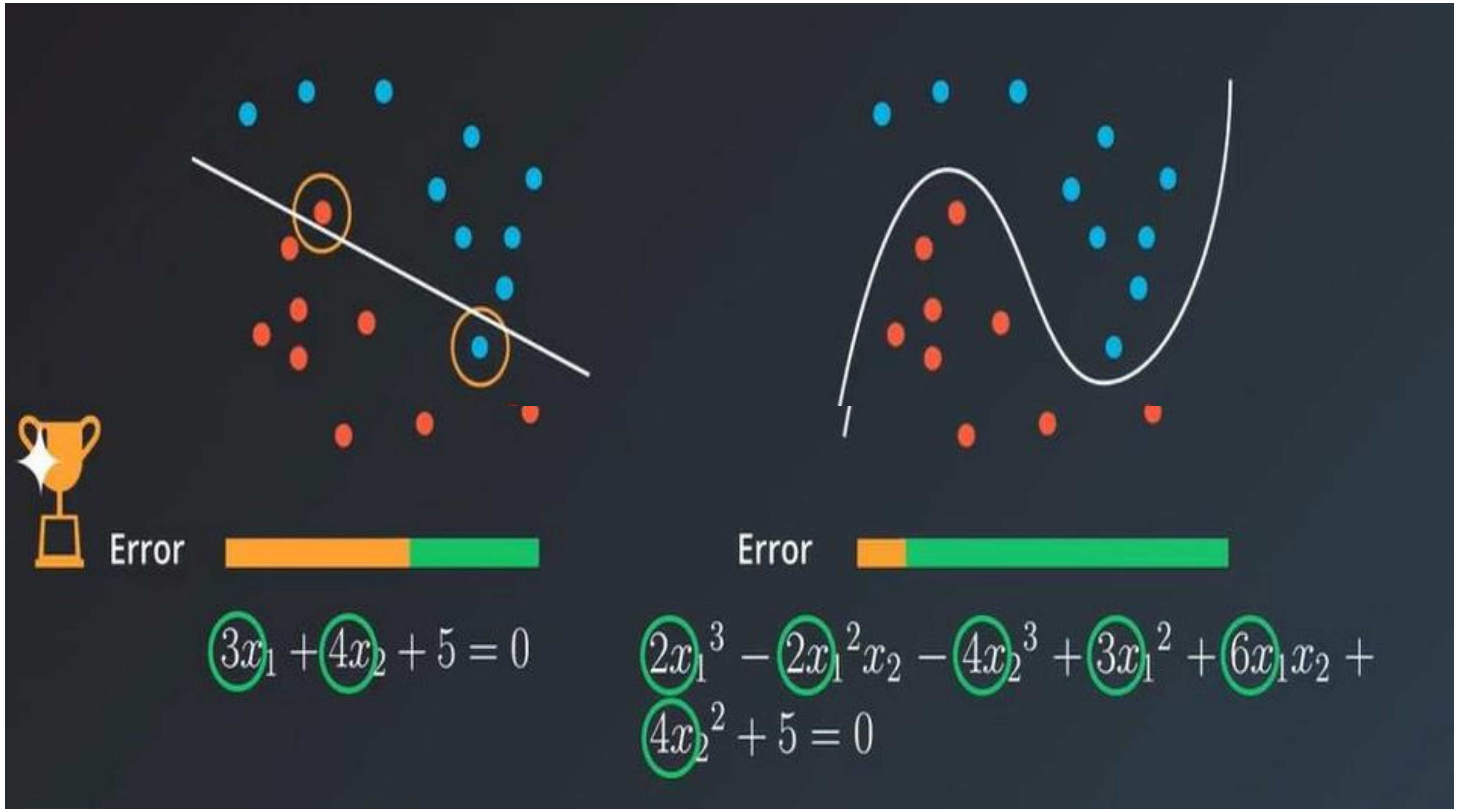


Polynomial Regression

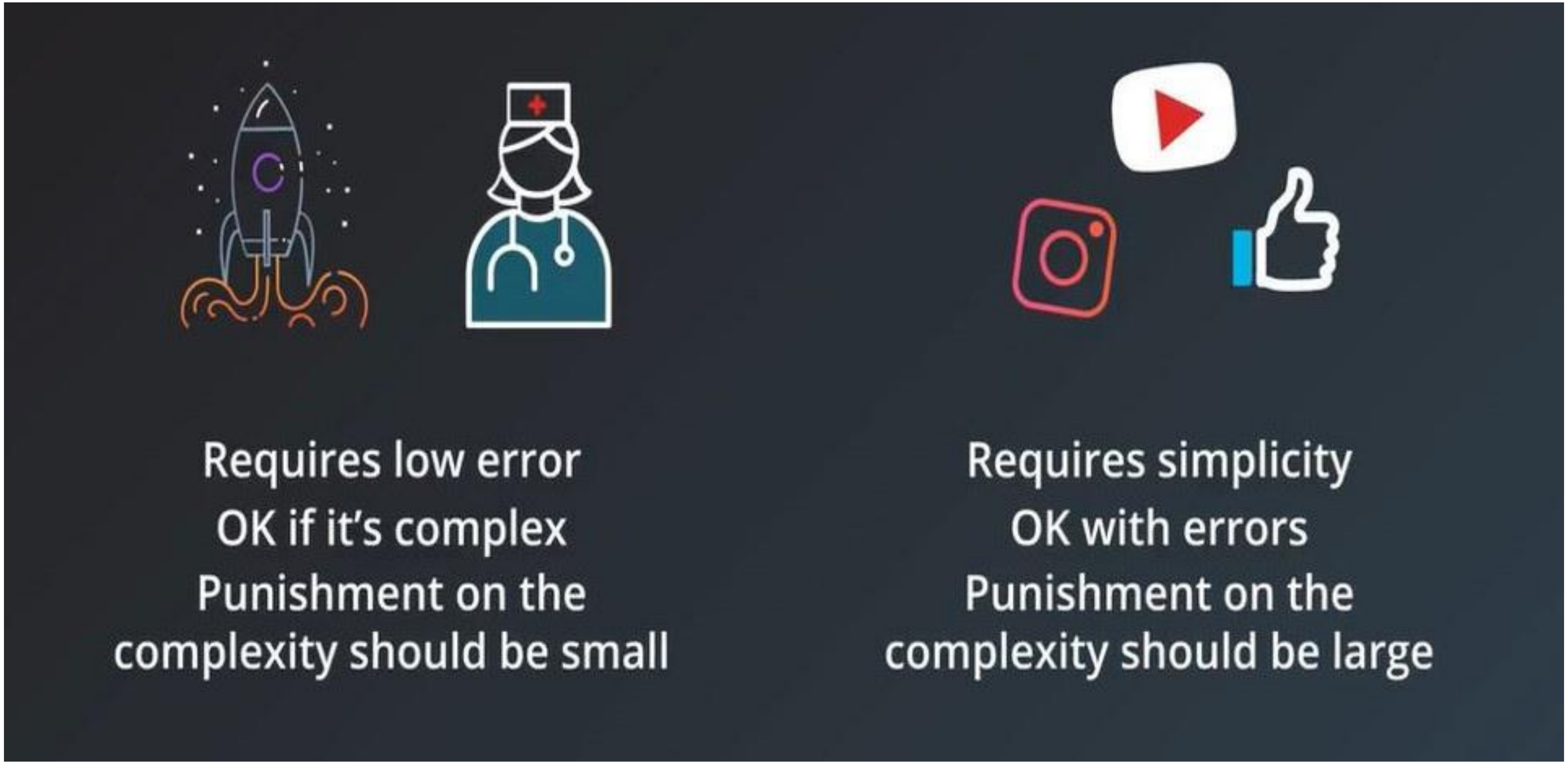




regression and classification


Error
$3x_1 + 4x_2 + 5 = 0$
Error
$2x_1{}^3 - 2x_1{}^2 x_2 - 4x_2{}^3 + 3x_1{}^2 + 6x_1 x_2 + 4x_2{}^2 + 5 = 0$

Simple VS Complex Models



WHAT IS SKLEARN?

SKLEARN FEATURES

- **Supervised Learning algorithms** – Linear Regression, (SVM), Decision Tree etc
- **Unsupervised Learning algorithms** – from clustering, factor analysis, PCA to unsupervised neural networks
- **Clustering** – This model is used for grouping unlabeled data.
- **Ensemble methods** – As name suggest, it is used for combining the predictions of multiple supervised models.

scikit learn

Let's Code!

Quiz Time!

Any Questions

Thank You