CS 5806 Machine Learning II

Lecture 6 - Learning Theory 2: VC Dimension

September 11th, 2023

Hoda Eldardiry

VC (Vapnik-Chervonenkis) Dimension

Lecture Objectives

- Explore the mathematical definition of the VC dimension
- Show examples of how it can be applied in practice

PAC Learning & Model Complexity

- Model complexity affects performance
- VC Dimension
  — Numeric representation of model complexity
  — Quantitative measure of how limited the concept space is for PAC learning

Learners & Complexity

- Complexity of a learner (its representational power)
  — Its ability to learn a wide variety of input-output relationships
  — At an extreme, its ability to memorize or overfit to the data

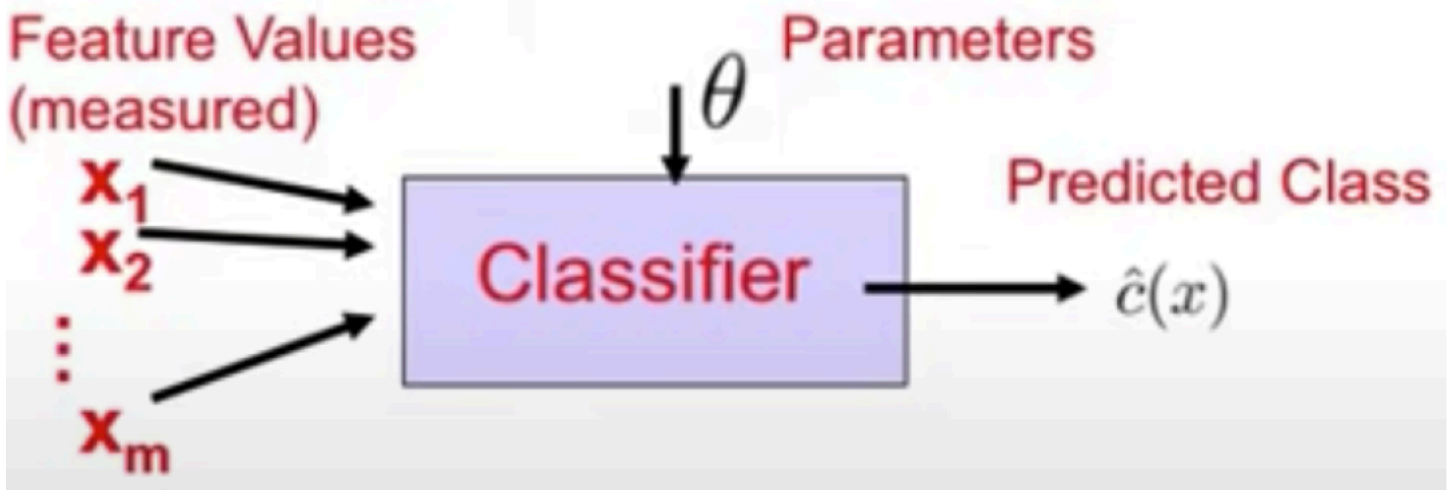
- Different Learners
  — Have different power
  — Define different classes of possible behavior

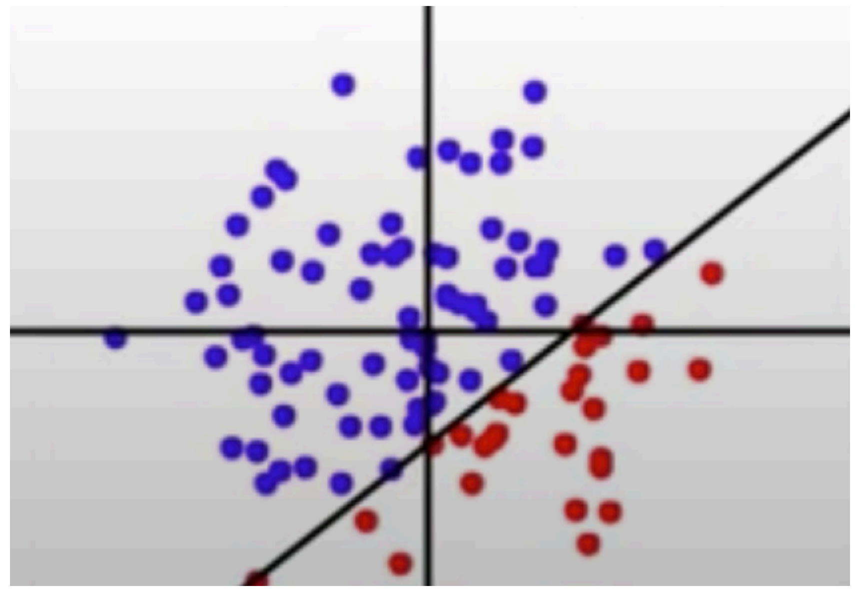
Example: Perceptron Learner

- A classification algorithm that makes predictions using a linear function

- A linear classifier in 2-D features $x_1 \& x_2$ with coefficients $\theta_1 \& \theta_2$ & constant θ_0

- Can learn functions that separate prediction classes (blue & red) by a line

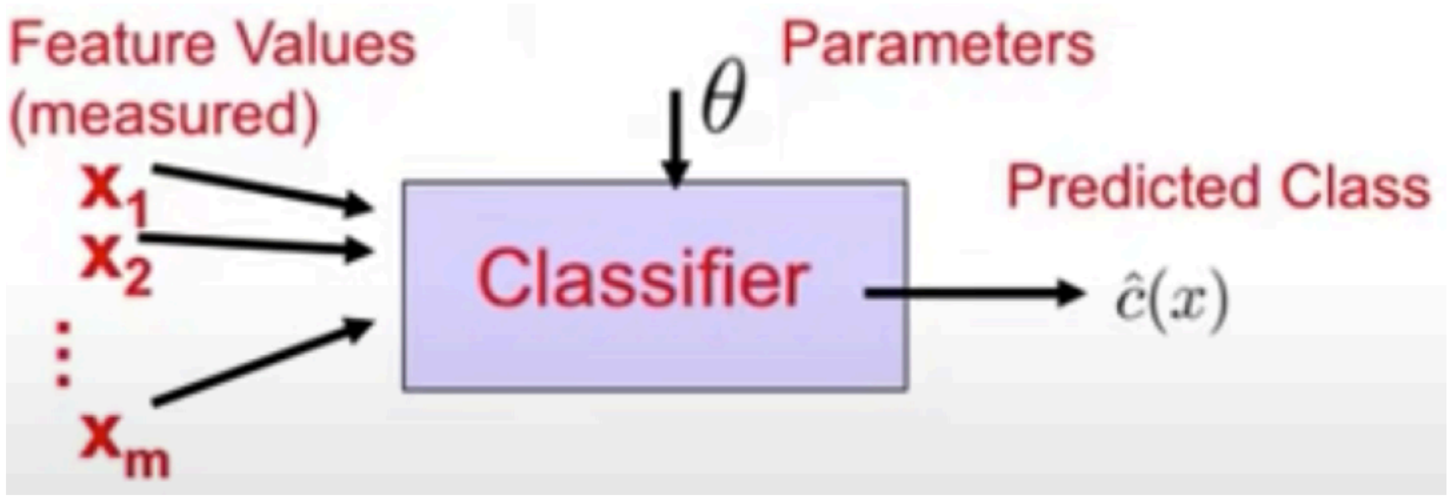




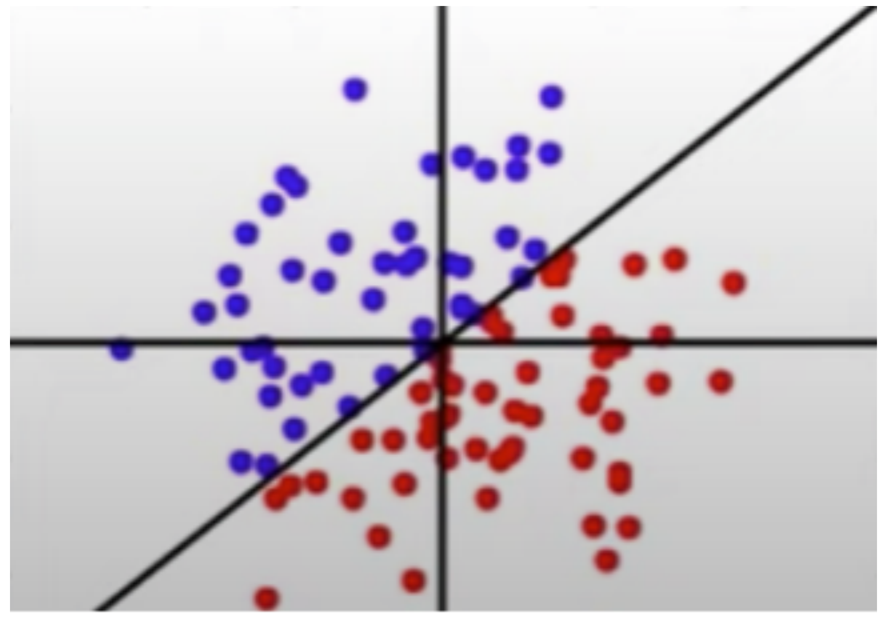
$$\hat{c} = sign(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Another Classifier

- Choosing a **different classifier form** would choose a **different set of possible functions**



- Example:
  — Choose a class of learners where we are thresholding a linear equation
  — No constant feature, only parameters $\theta_1 \& \theta_2$
  — Decision boundary pass through origin
  — Defines a more restrictive class of functions
  — A line divides space into blue & red predictions
  — But the line is restricted



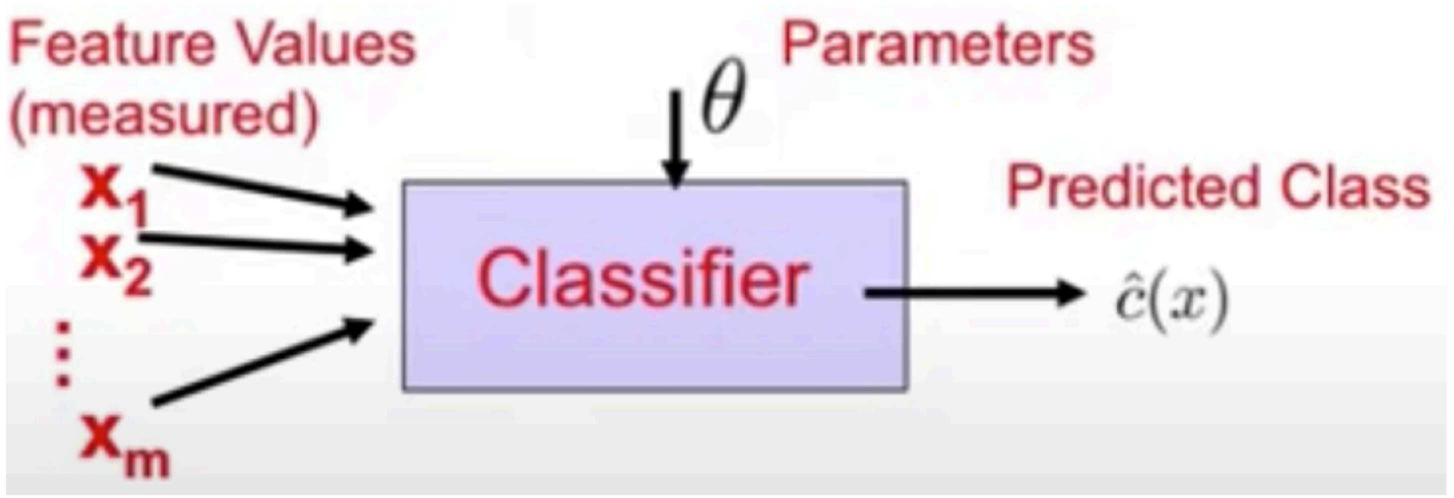
$$\hat{c} = sign(\theta_1 x_1 + \theta_2 x_2)$$

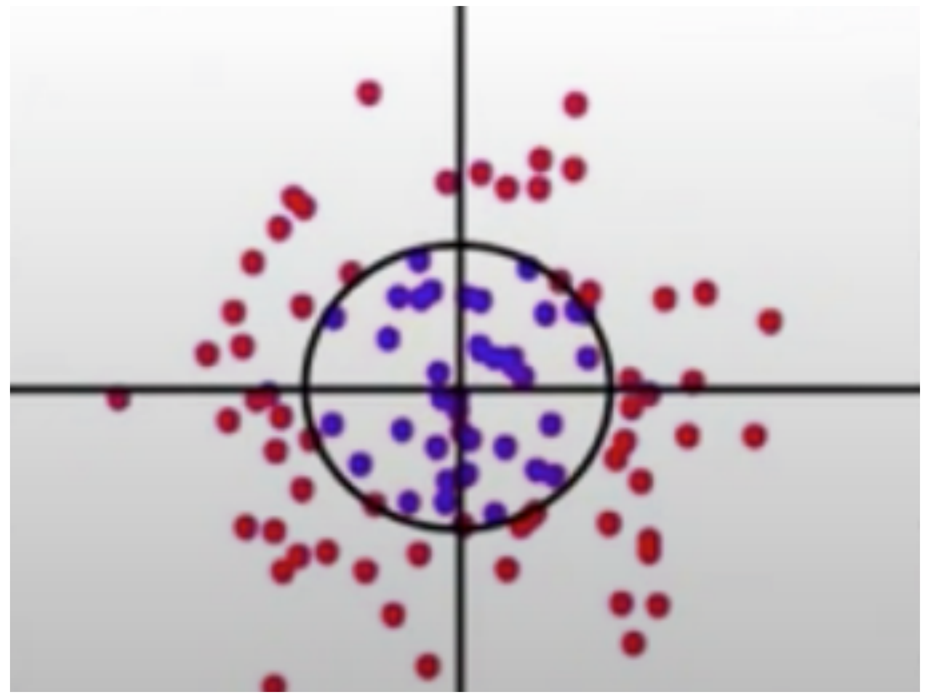
Another Classifier

- Another different functional form
  Take the sign of the norm of x minus some constant θ_0

- The kinds of functions this form can produce:
  — If x is farther from the origin than θ_0
  — Then function predicts class +
  — If x is closer to the origin than θ_0
  — Then function predicts class -1

- The set of all functions learnable by this classifier:
  — Any function where the negative class is in some sphere centered around the origin





$$\hat{c}=sign(x^T x - \theta_0)$$

Learners & Complexity

- We’ve seen many versions of underfit/overfit trade-off
  — Learner complexity
  — “Representational power”

- Different learners have different power

- Usual trade-off
  — More power: represent more complex systems, might overfit
  — Less power: won’t overfit, but may not find the “best” learner

- How can we quantify representational power?
  — Not easy
  — One solution is VC dimension

VC Dimension

- A key concept in machine learning theory that measures the complexity of a learning algorithm
- A measure of the capacity of a class of functions that can be learned by a learning algorithm
- A way to understand how a model can generalize to new data
- Can help to identify when a model is overfitting or underfitting

VC Dimension

- VC-dim measures the “power” of a learner
- Does *not* necessarily equal the number of parameters!
- Number of parameters does not necessarily equal complexity
  — Can define a classifier with many parameters but not much power
  — Can define a classifier with 1 parameter but lots of power
- Lots of work to determine what the VC dimension of various learners is

Binary Classification of Images

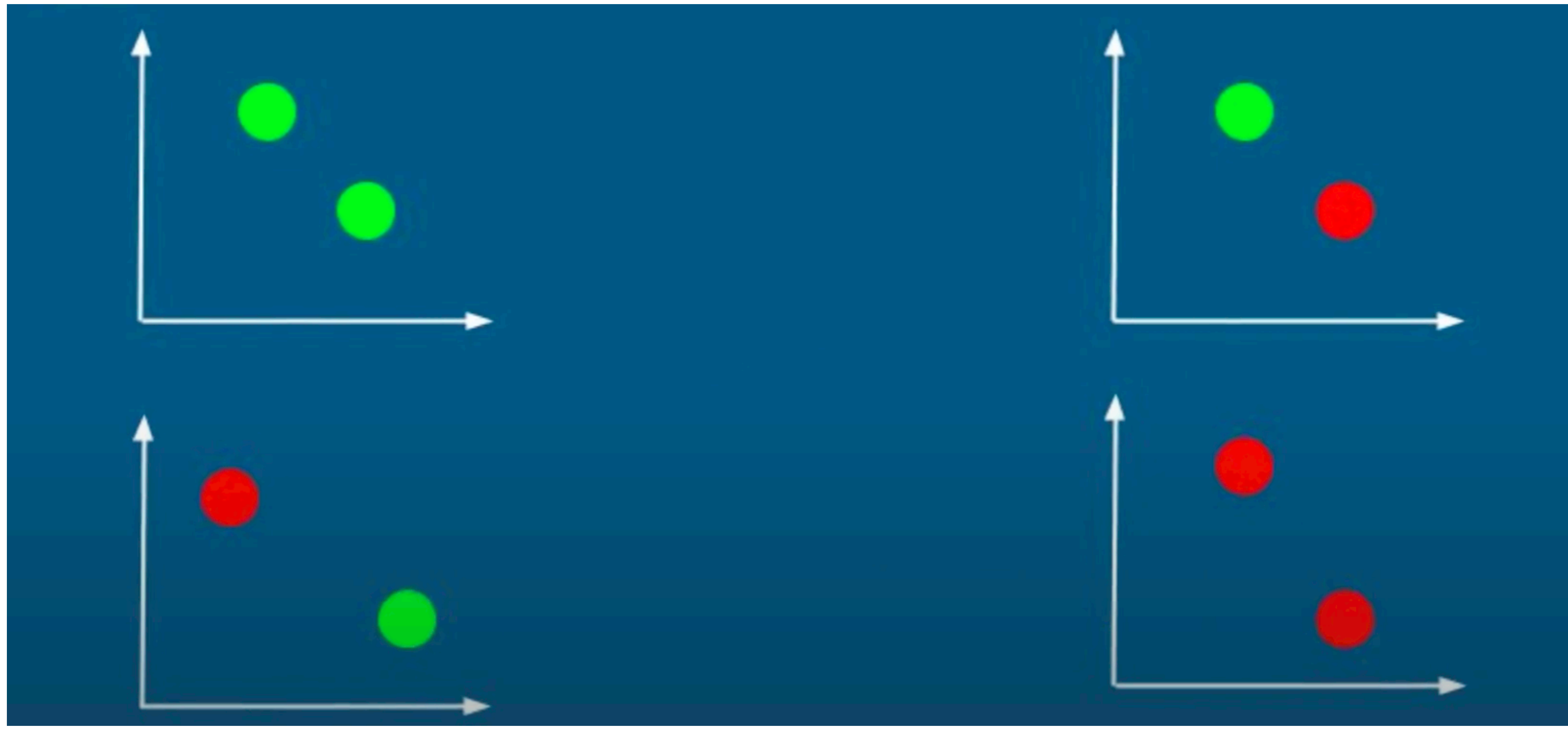
- Classify an image as cat or dog
- Assume we have only 2 instances
- 2 instances can be classified in 4 different ways:
  — Both cats
  — Image 1 cat, image 2 dog
  — Image 1 dog, image 2 cat
  — Both dogs

Binary Classification in 2-dim plane

- Consider positive/negative class instead of cat/dog
- Use green/red dots to represent positive/negative
- Draw a scatter plot

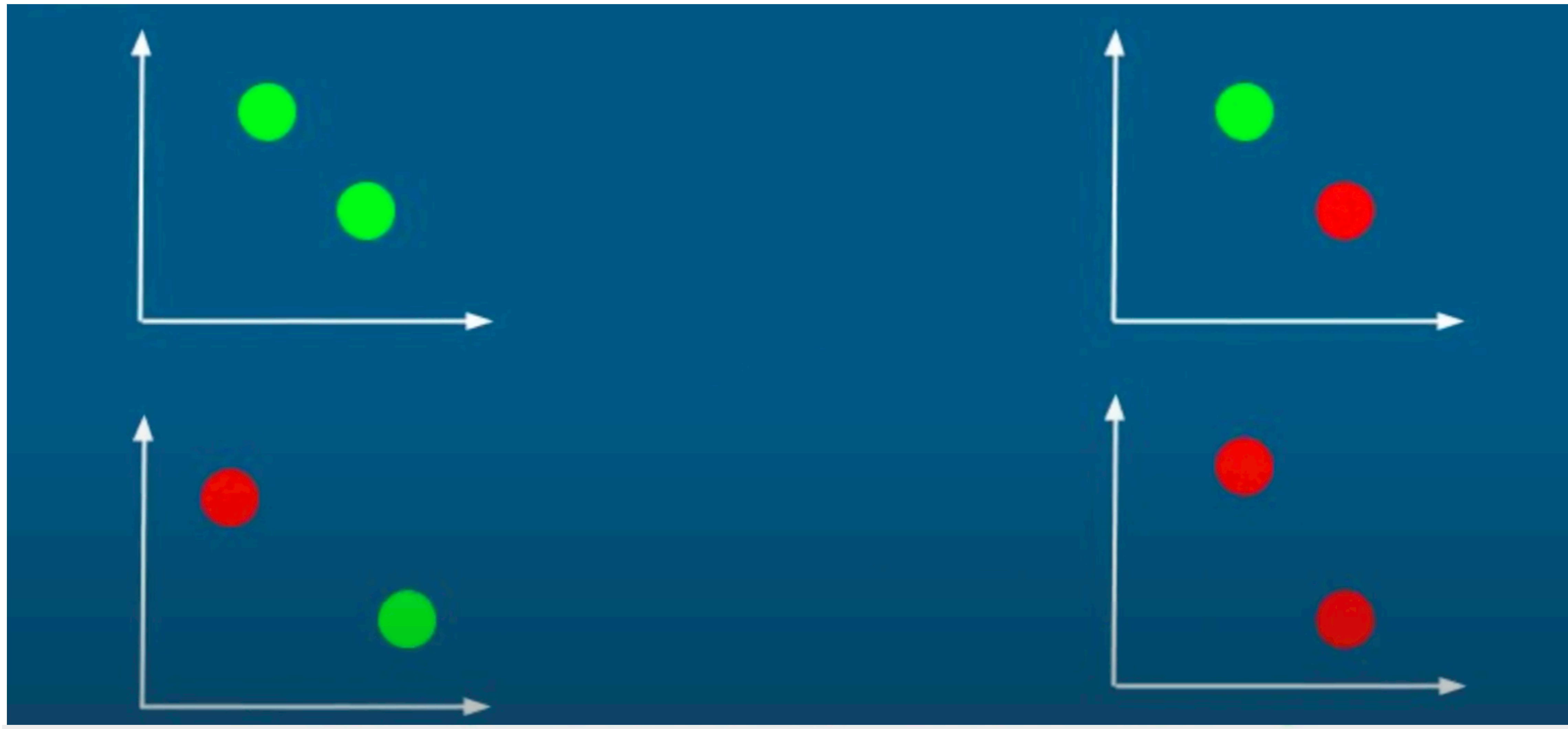
Binary Classification in 2-dim plane

- **2 data points** can be classified in 4 different ways

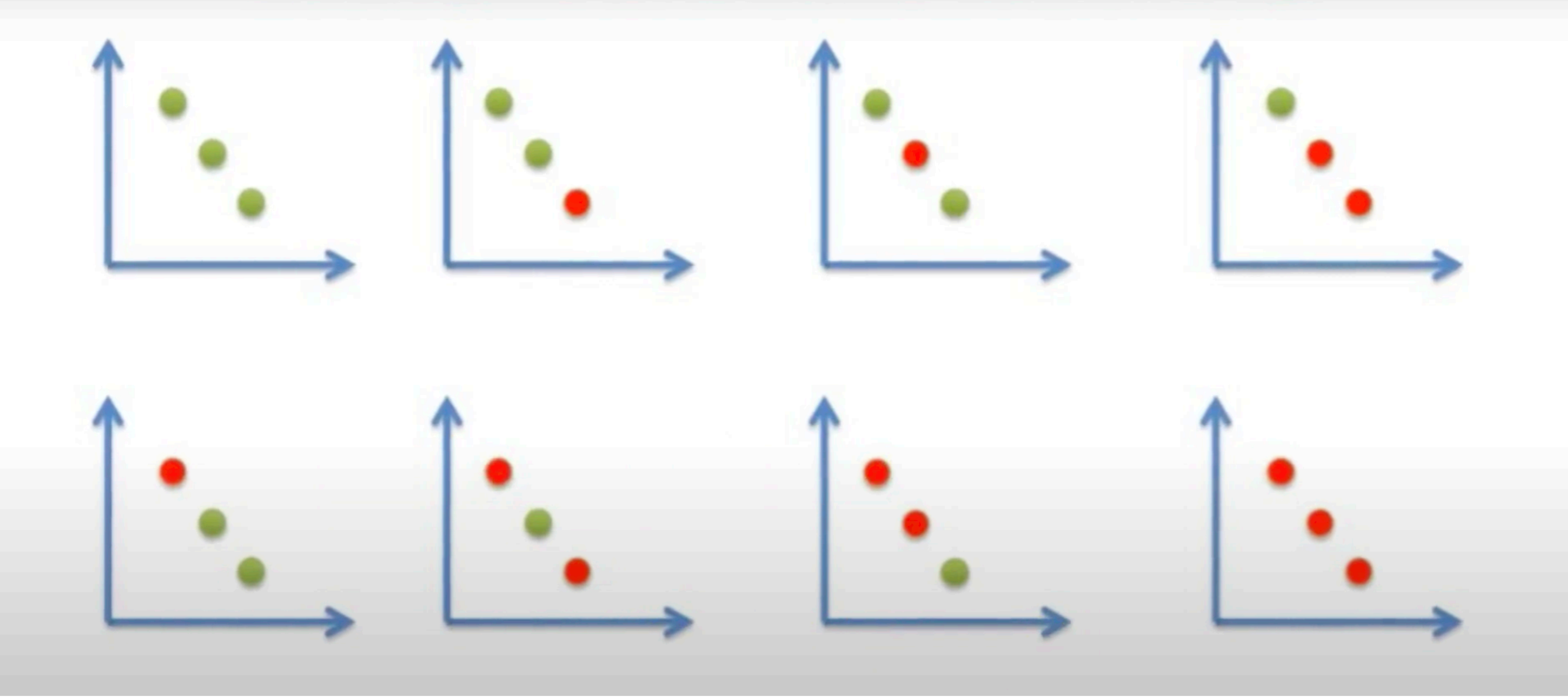


Binary Classification in 2-dim plane

- **2 data points** can be classified in 4 different ways

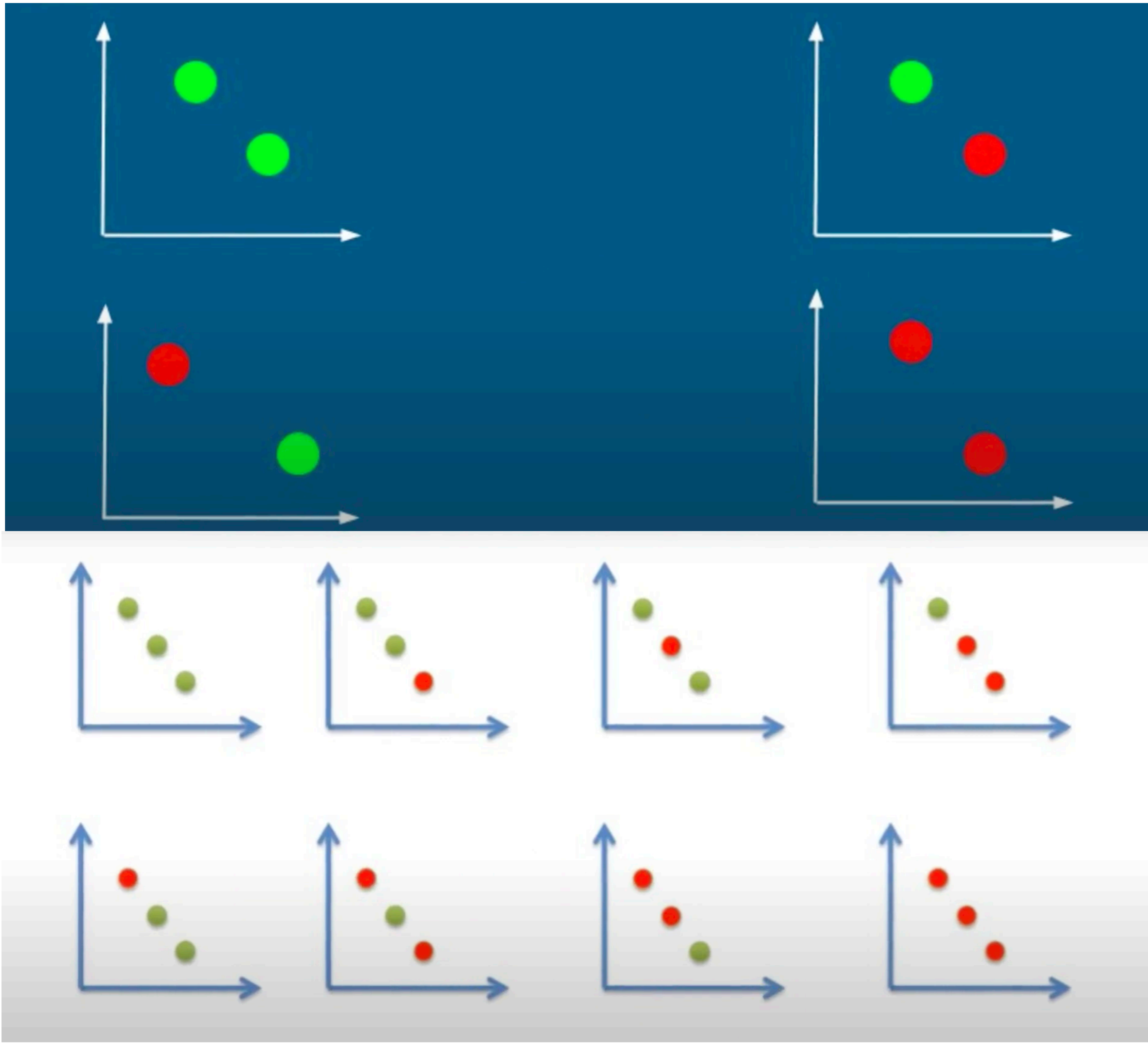


- **3 data points** can be classified in 8 different ways



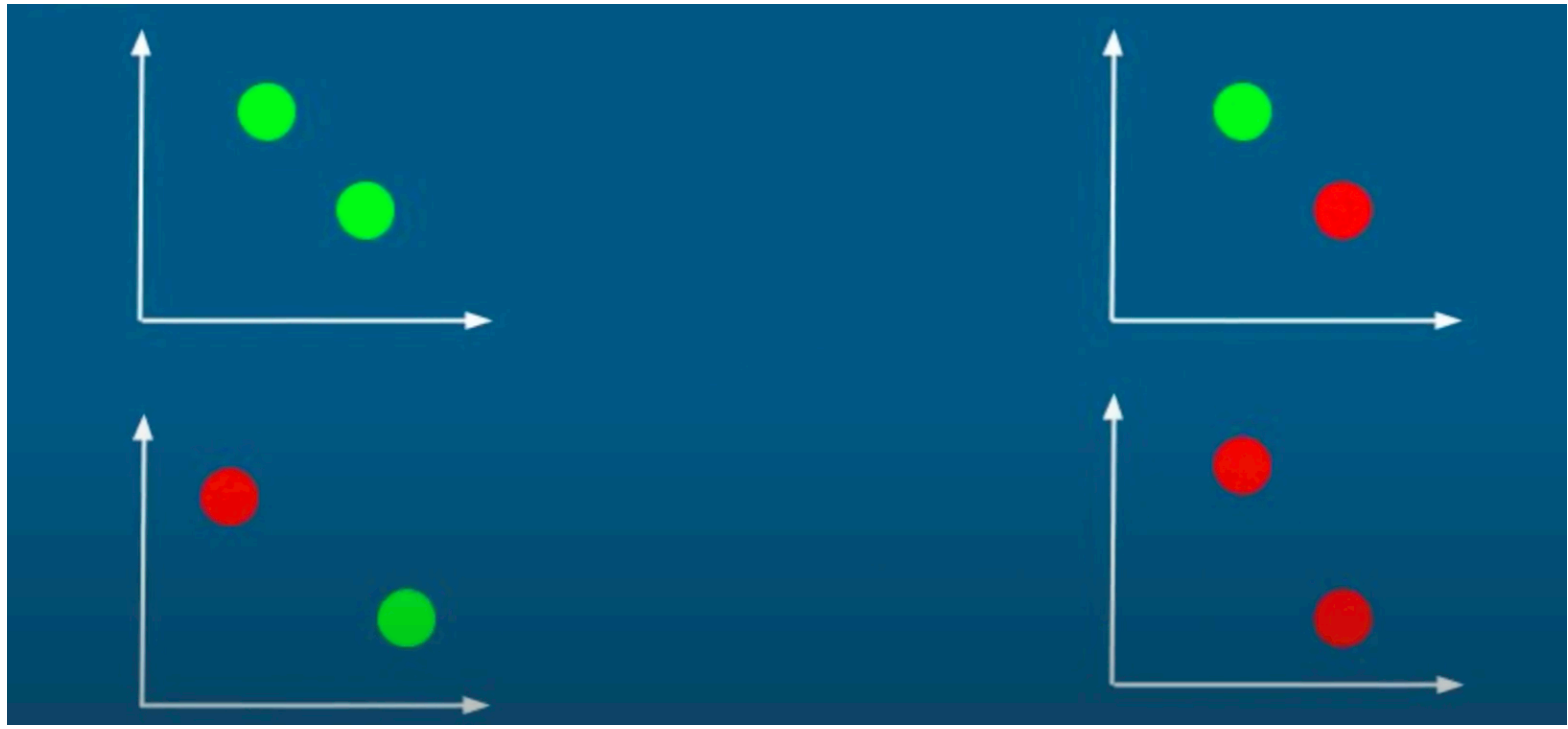
Binary Classification in 2-dim plane

- **2 data points** can be classified in 4 different ways
- **3 data points** can be classified in 8 different ways
- In general, **N points** can be classified in 2^N different ways



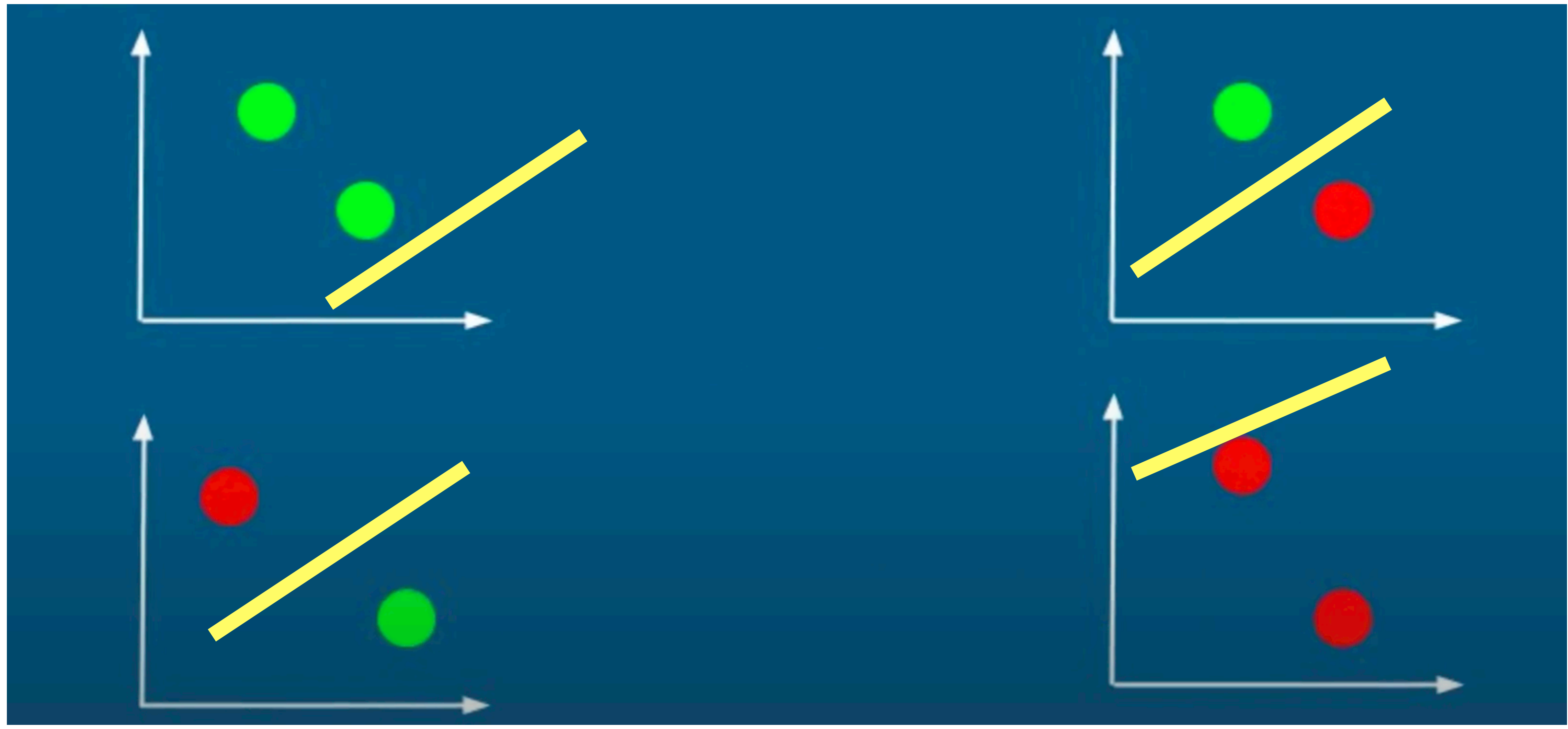
Linear Classifier in 2-D Plane: 2 points

- For the scatter plot with 2 data points
- Assume classifier is a straight line
- Can the 2 points be correctly classified using a straight line?



Linear Classifier in 2-D Plane: 2 points

- Case 1 (2 green dots)
  — Draw straight line below the 2 points
  — Model classifies all points above line as positive & all below line as negative
  — 2 points are correctly classified



- Same applies for other 3 cases
  — Possible to draw line that separates positive & negative in all 4 cases

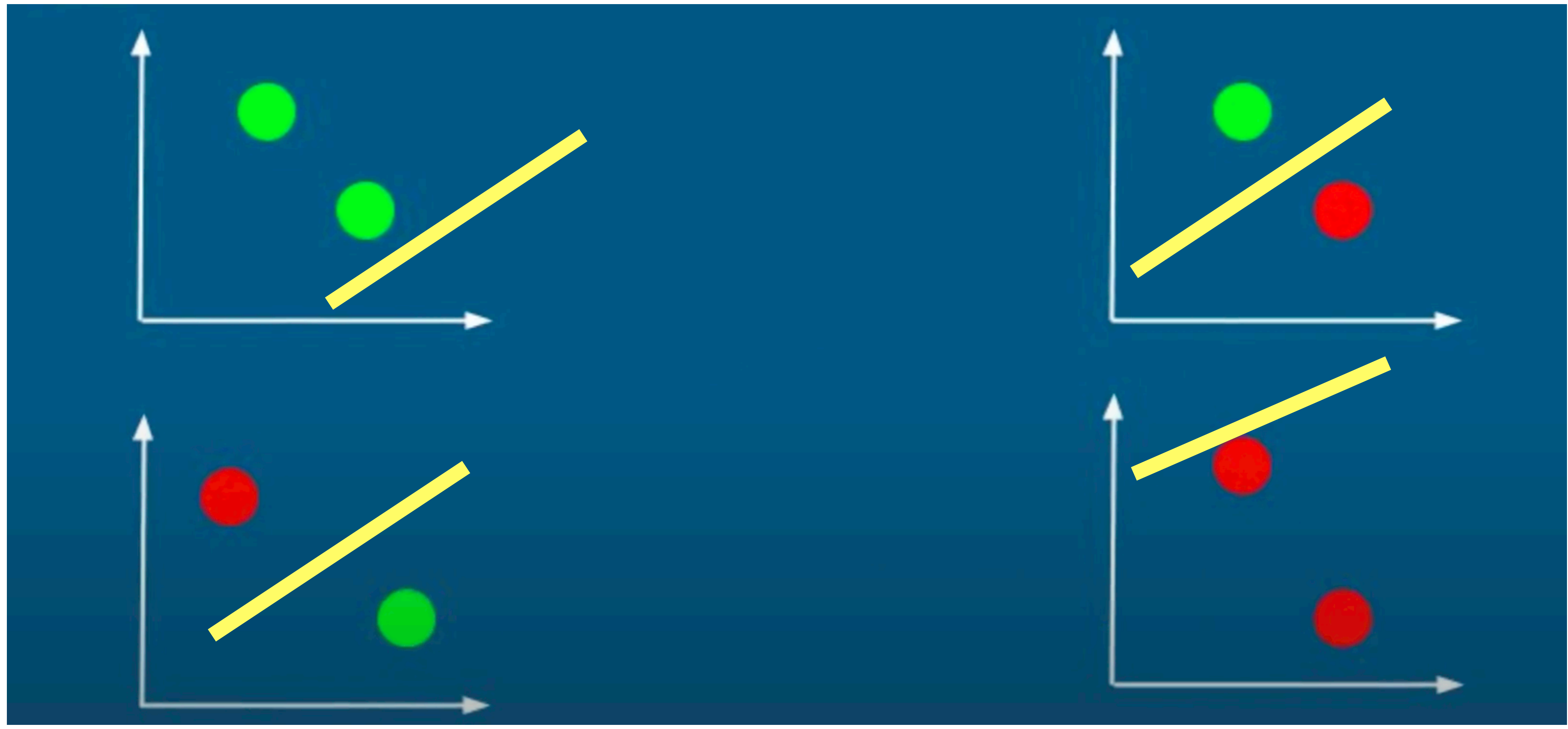
- We say:
  <u>Straight line (or linear function) is expressive enough to **shatter** 2 points on a 2-dim plane</u>

Shattering

- A set of N points is shattered by hypothesis space H
  if there is a hypothesis in H
  that separates positive examples from negative examples
  in all of the 2^N possible ways

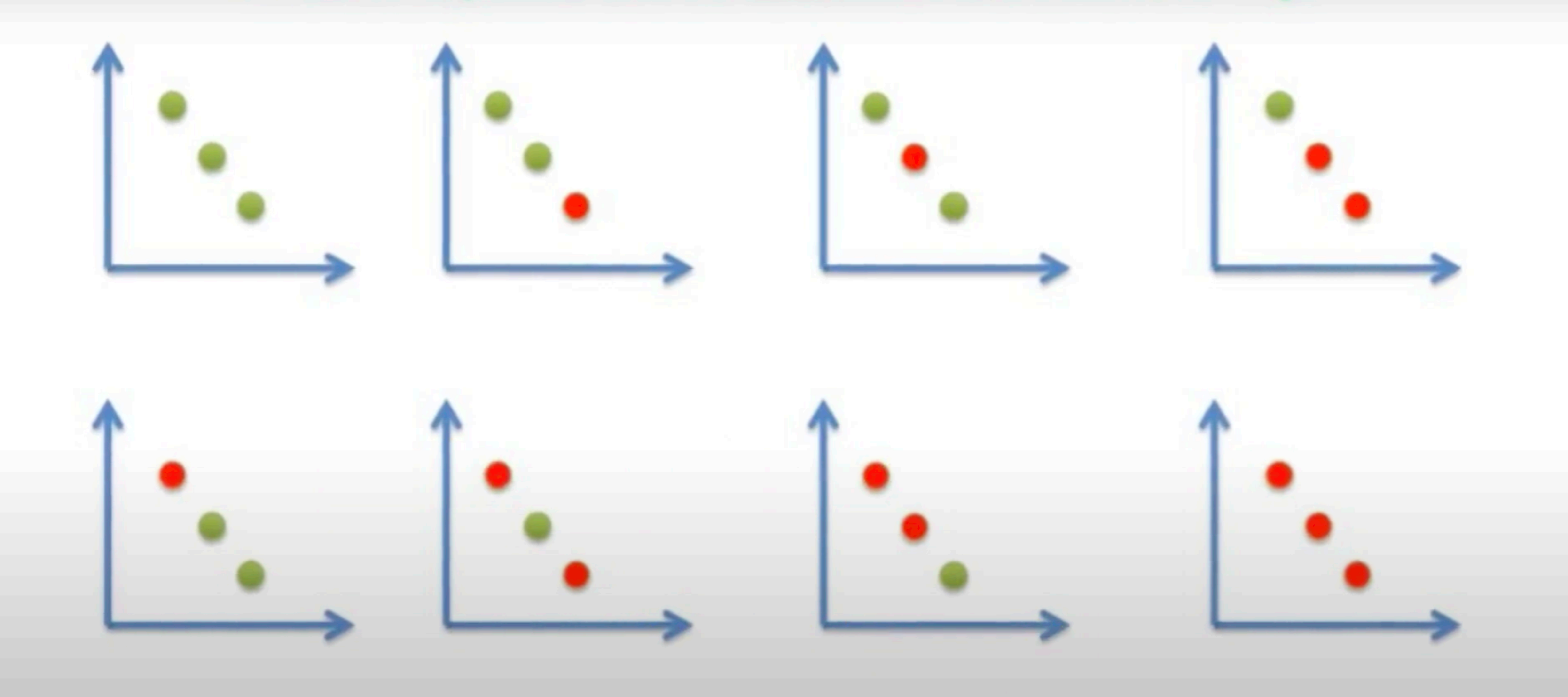
Linear Classifier in 2-D Plane: 2 points

- Hypothesis space: the set of all straight lines in the 2-dimensional plane
- If there is hypothesis H
  (if there is any straight line in the set of consecutive lines)
  that correctly classifies the data points in all 4 possible ways
- Hence a **line** can **shatter** 2 points on a 2-dim plane



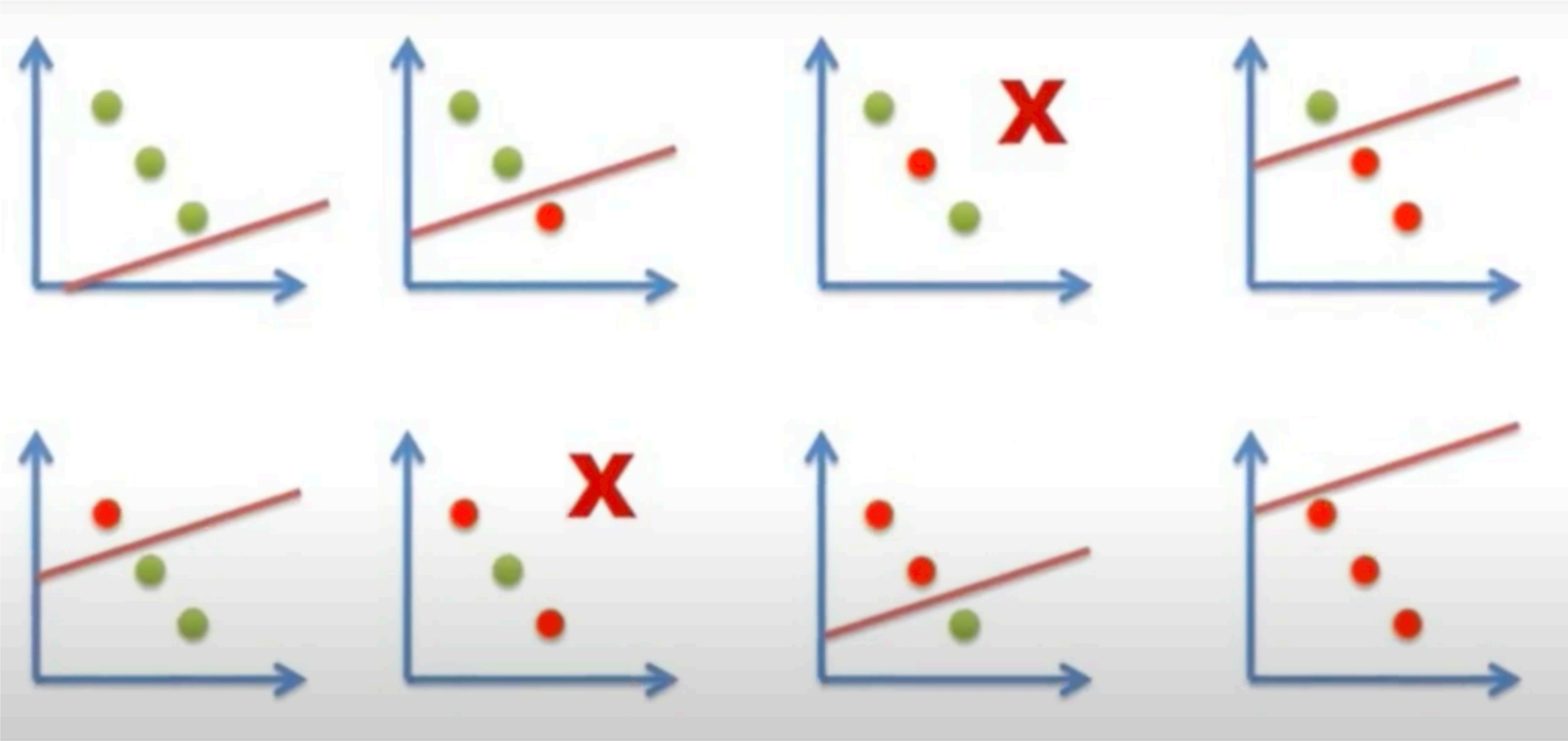
Linear Classifier in 2-D Plane: 3 points

- Consider the case of 3 data points
- Assume classifier is a straight line
- Can a straight line correctly classify all 8 possible combinations?

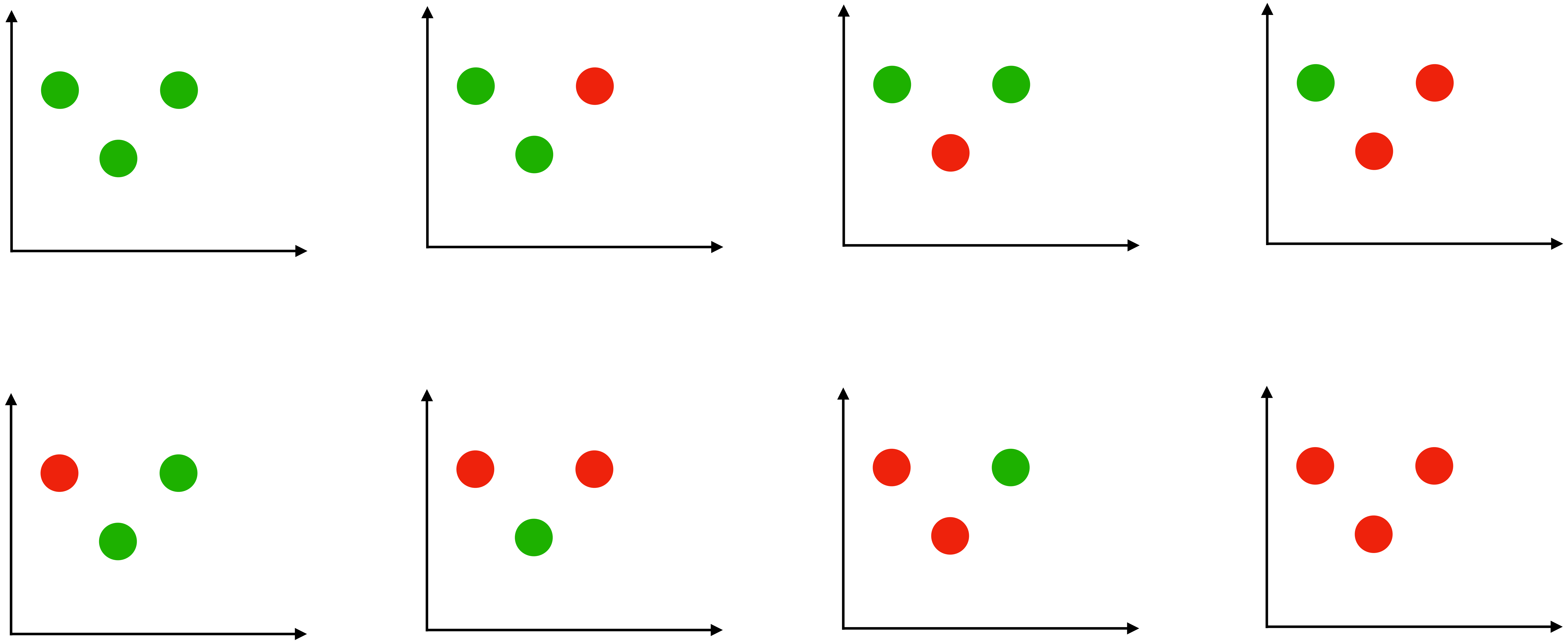


Linear Classifier in 2-D Plane: 3 points

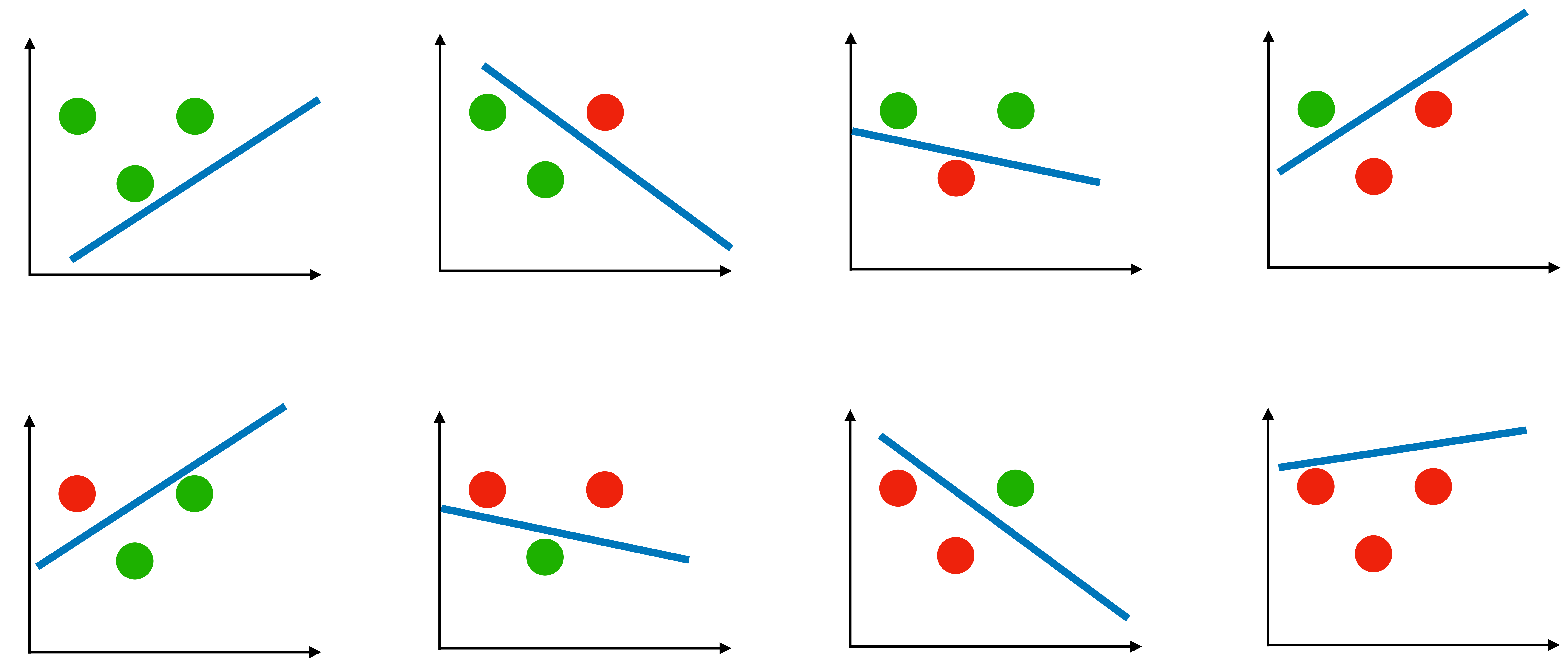
- 2 out of 8 cases cannot be correctly classified using a straight line
- Can we find another set of 3 points that can be correctly classified using a straight line?



How about this set of 3 points?



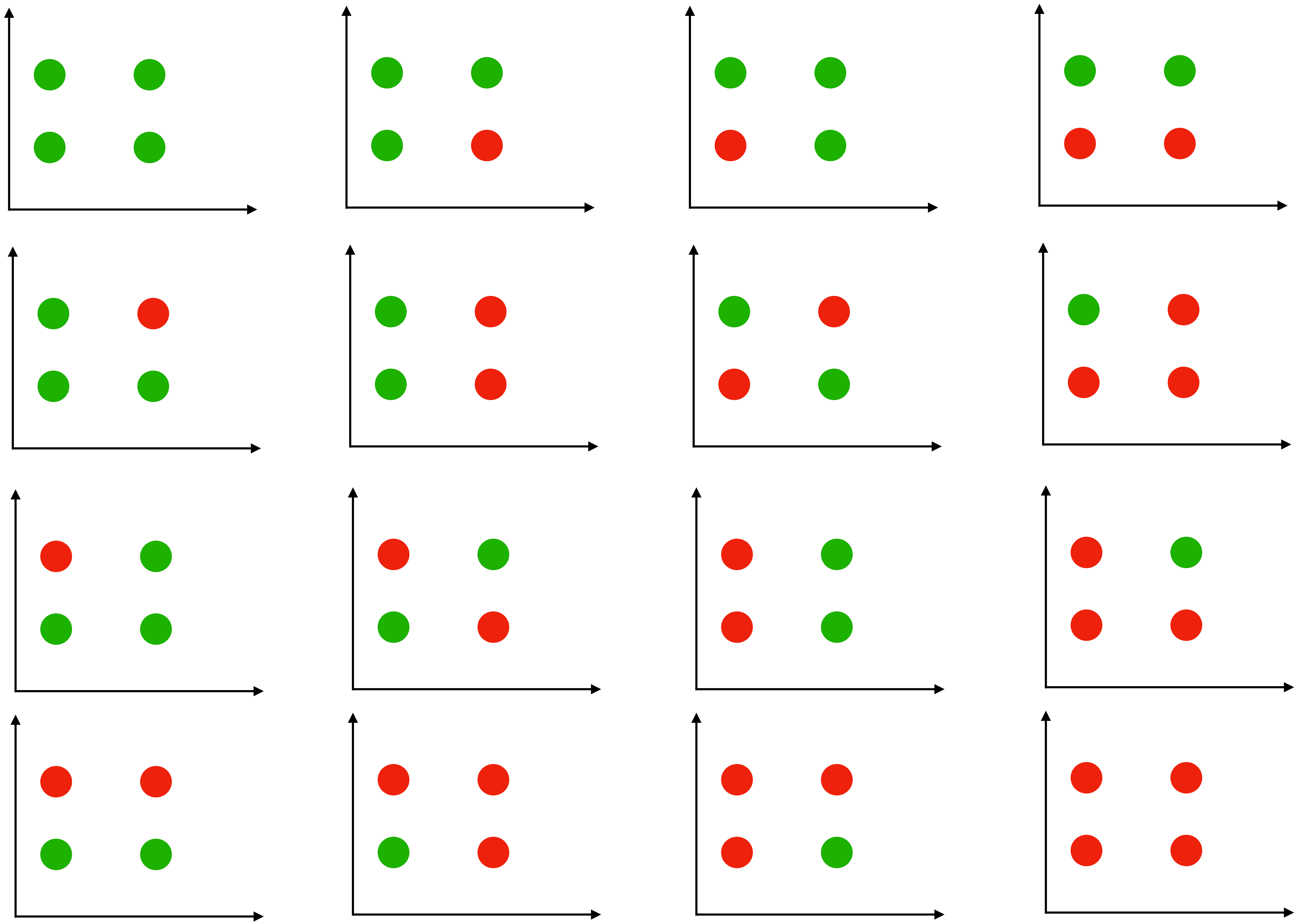
How about this set of 3 points?



Possible to classify all 8 cases using a straight line
“A straight line can shatter 3 data points on a 2-dimensional plane”

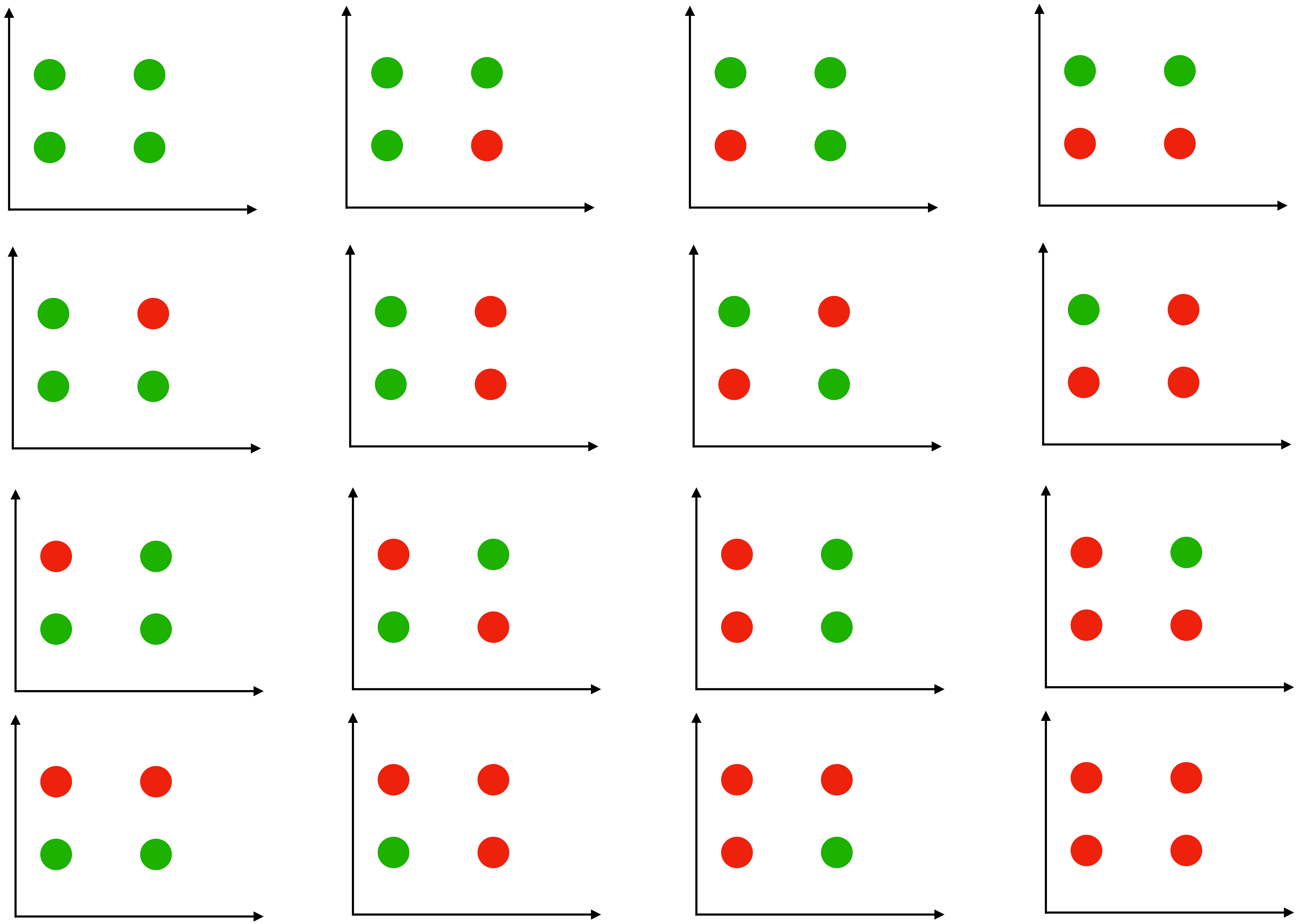
May not be able to shatter every possible set of 3 points in 2 dimensions
There may be some set of points that may not be shattered
But it is enough to find 1 set of 3 points that can be shattered

Linear Classifier in 2-D Plane: 4 points



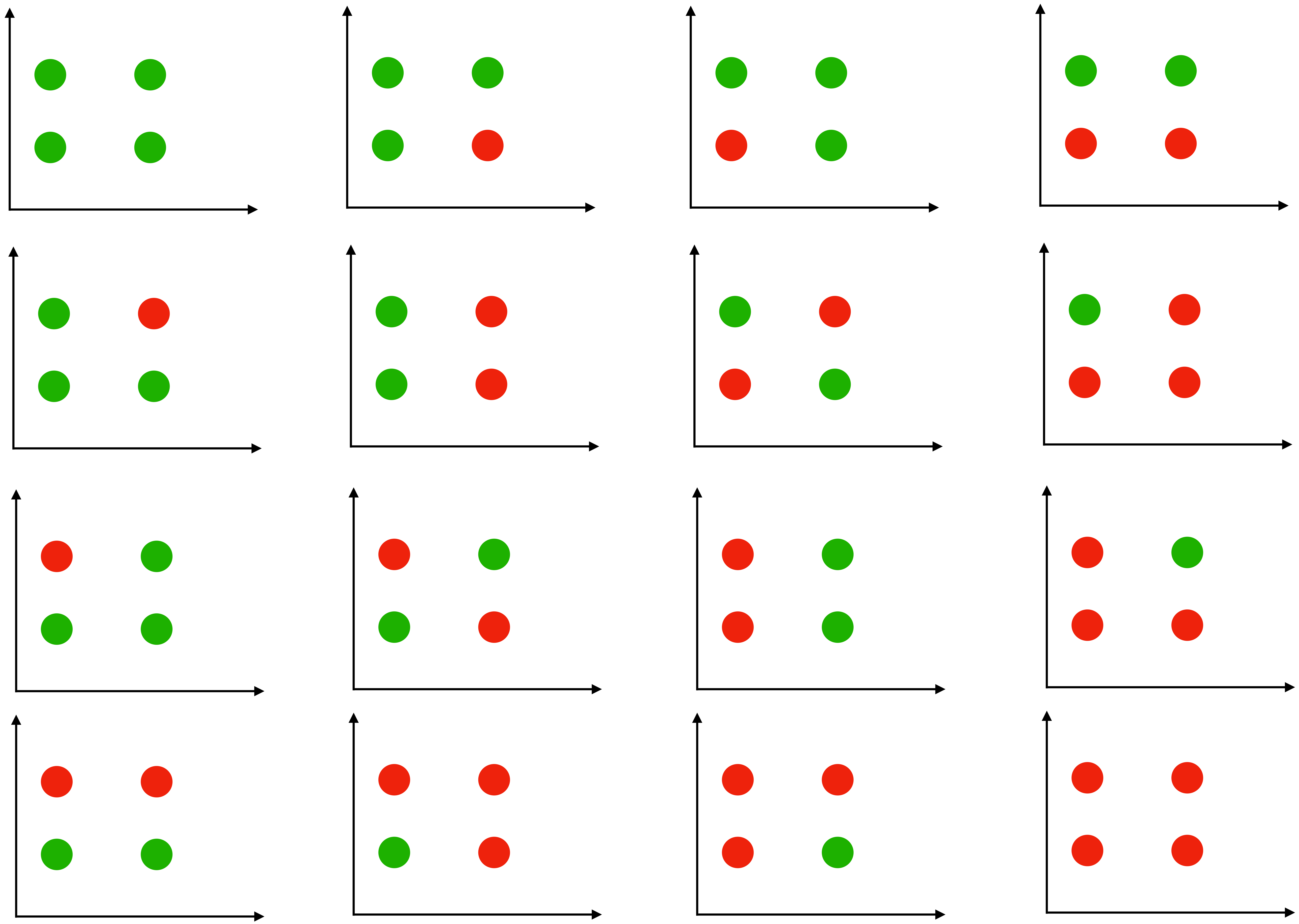
- 16 different ways of classification
- Can this set of 4 points be shattered by a straight line?

Linear Classifier in 2-D Plane: 4 points



- 16 different ways of classification
- Can this set of 4 points be shattered by a straight line?
  — No!
- Can we find another set of 4 points that can be shattered by a straight line?

Linear Classifier in 2-D Plane: 4 points



- 16 different ways of classification
- Can this set of 4 points be shattered by a straight line?
  — No!
- Can we find another set of 4 points that can be shattered by a straight line?
  — No!

Straight Line

- The maximum number of points in a 2-dimensional plane (R^2) that can be shattered by a straight line is 3

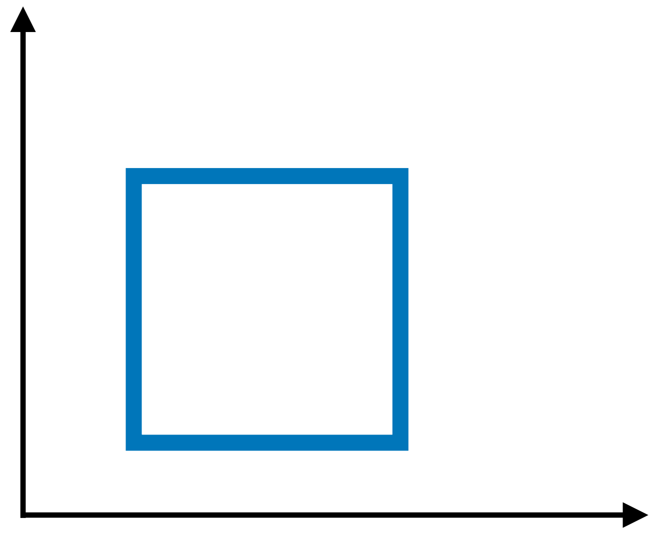
- The VC dimension of a straight line in a 2-dimensional plane is 3

- VC(straight line in R^2) = 3

- VC(H) is the maximum number of points that can be shattered by H

- VC dimension: a measure that characterizes the expressive power or capacity of a hypothesis class

Axis-aligned rectangle



- Let’s take another class of hypotheses
- Consider the axis-aligned rectangle
  (rectangle edges parallel to coordinate axis)
- Let’s find the VC-Dim(axis-aligned rectangle)

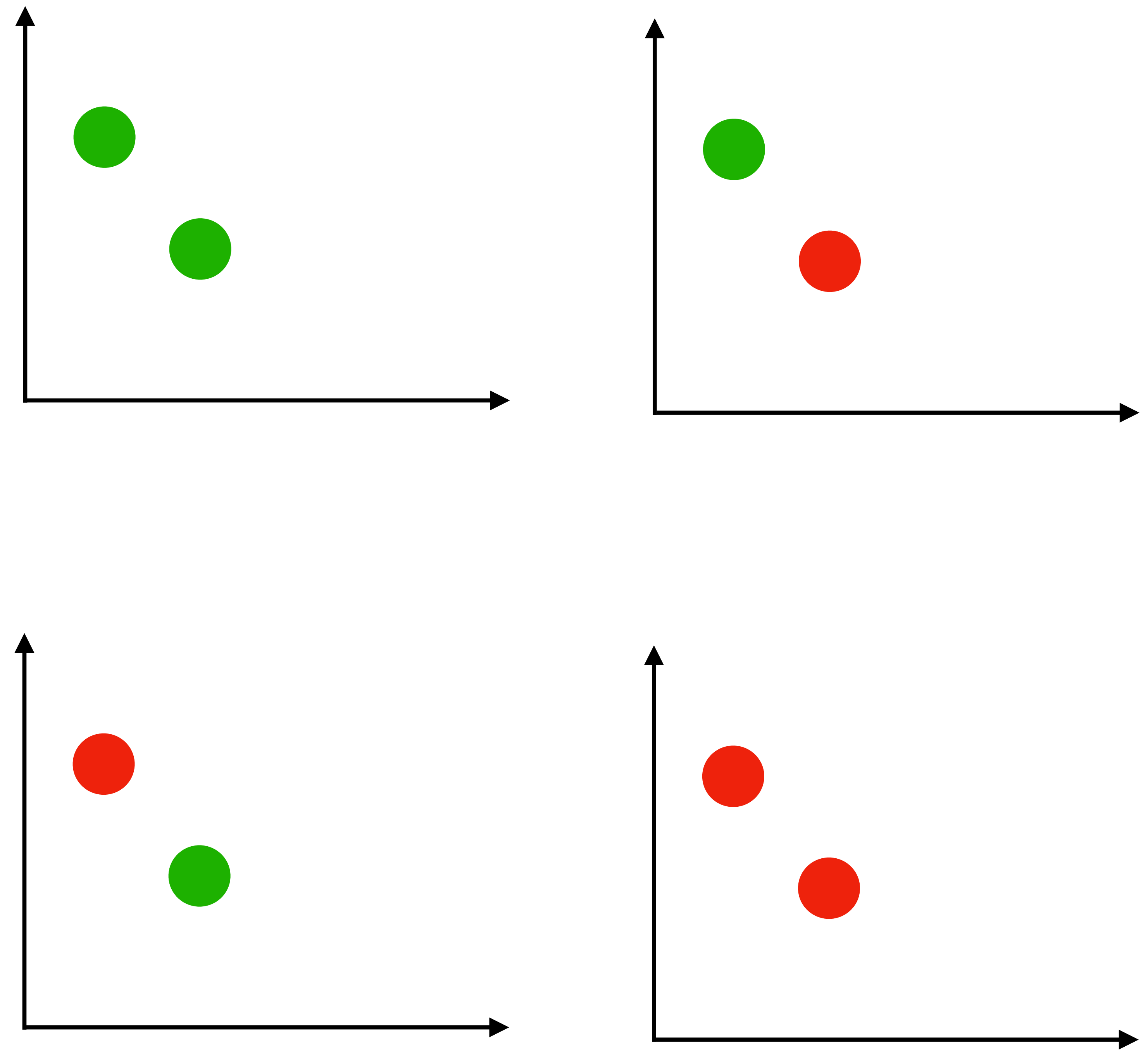
Axis-aligned rectangle: 2 points

- Take 2 data points
- How many ways can 2 points be classified?

Axis-aligned rectangle: 2 points

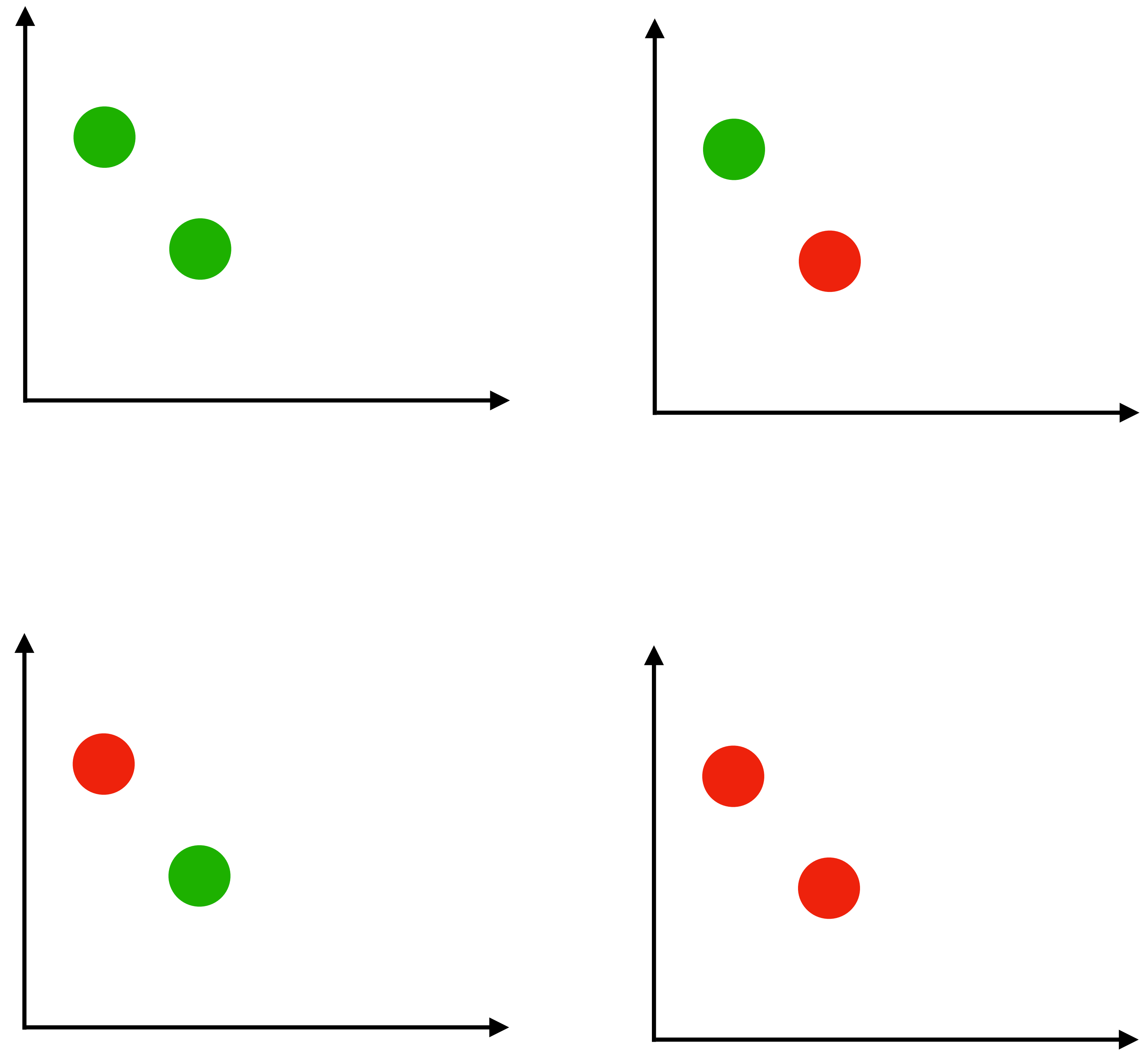
- How many ways can 2 points be classified?

- 4 ways



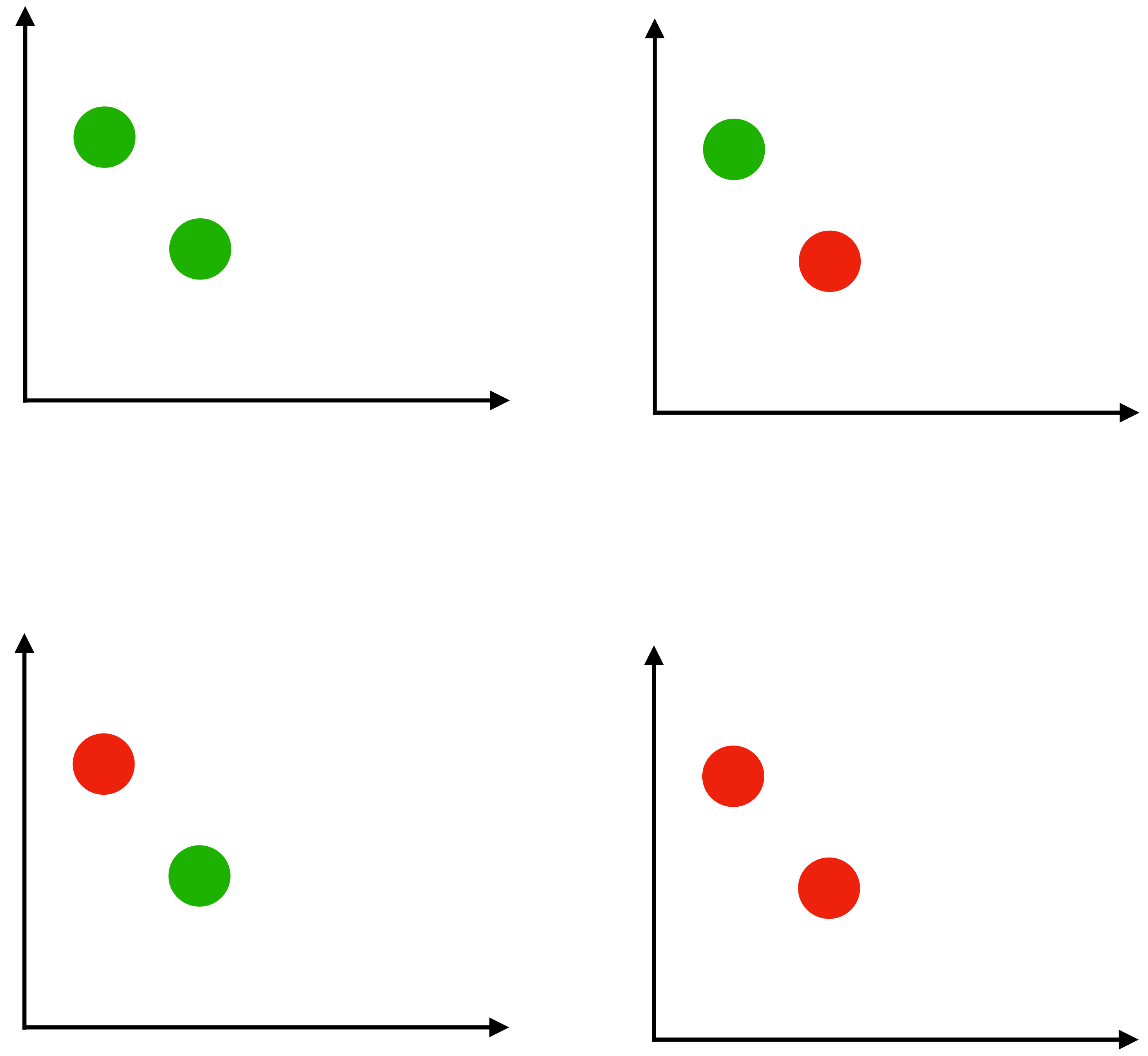
Axis-aligned rectangle: 2 points

- How many cases out of the 4 cases can be correctly classified using an axis-aligned rectangle?
- Can we correctly classify all 4 cases using an axis-aligned rectangle?



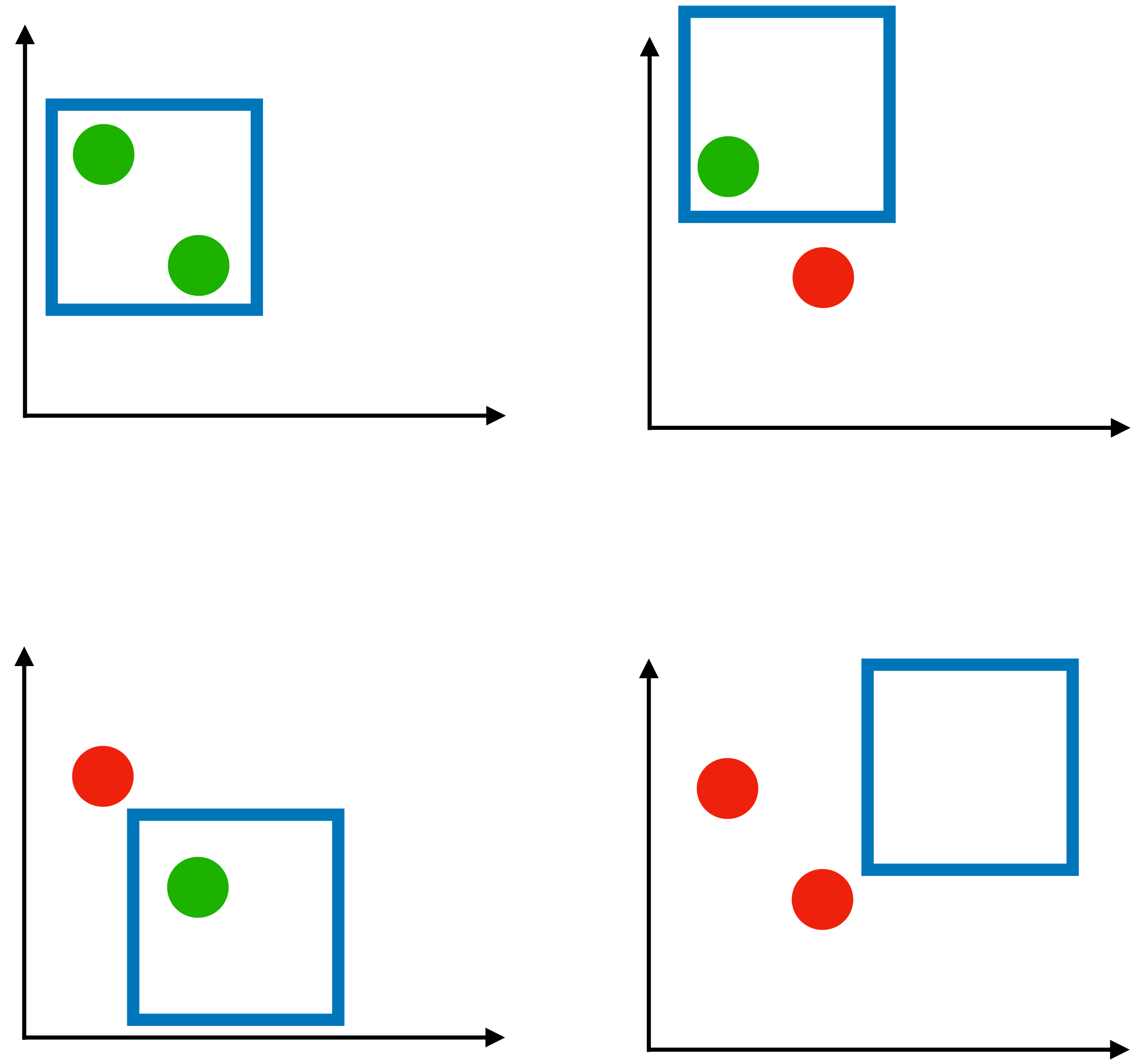
Axis-aligned rectangle: 2 points

- Assume all points inside rectangle as positive & outside rectangle as negative



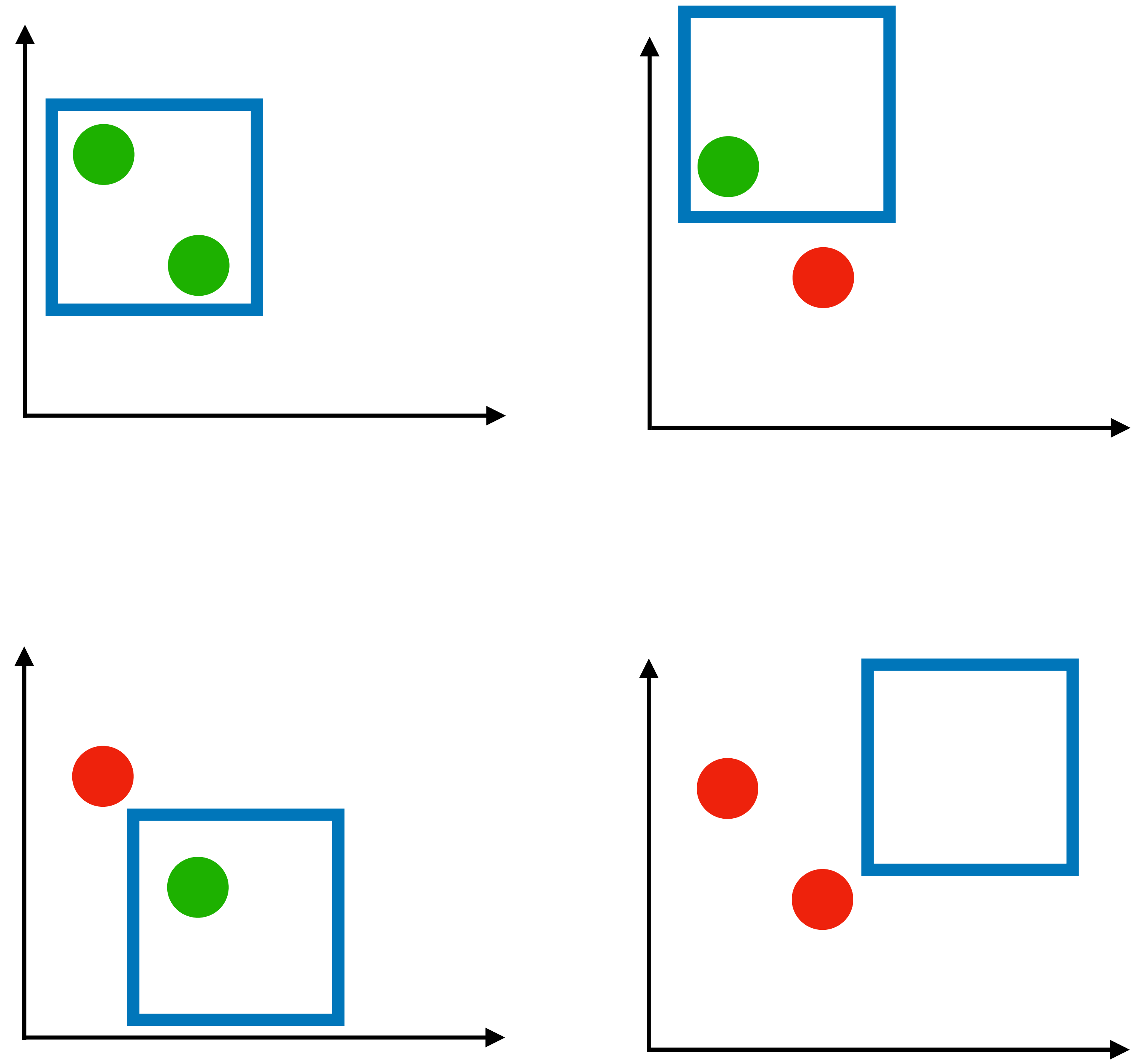
Axis-aligned rectangle

- Assume all points inside rectangle as positive & outside rectangle as negative
- We can correctly classify all 4 cases
- What is the conclusion?



Axis-aligned rectangle: 2 points

- Assume all points inside rectangle as positive & outside rectangle as negative
- We can correctly classify all 4 cases
- **The axis-aligned rectangle can shatter 2 points in a 2-dimensional plane (R^2)**

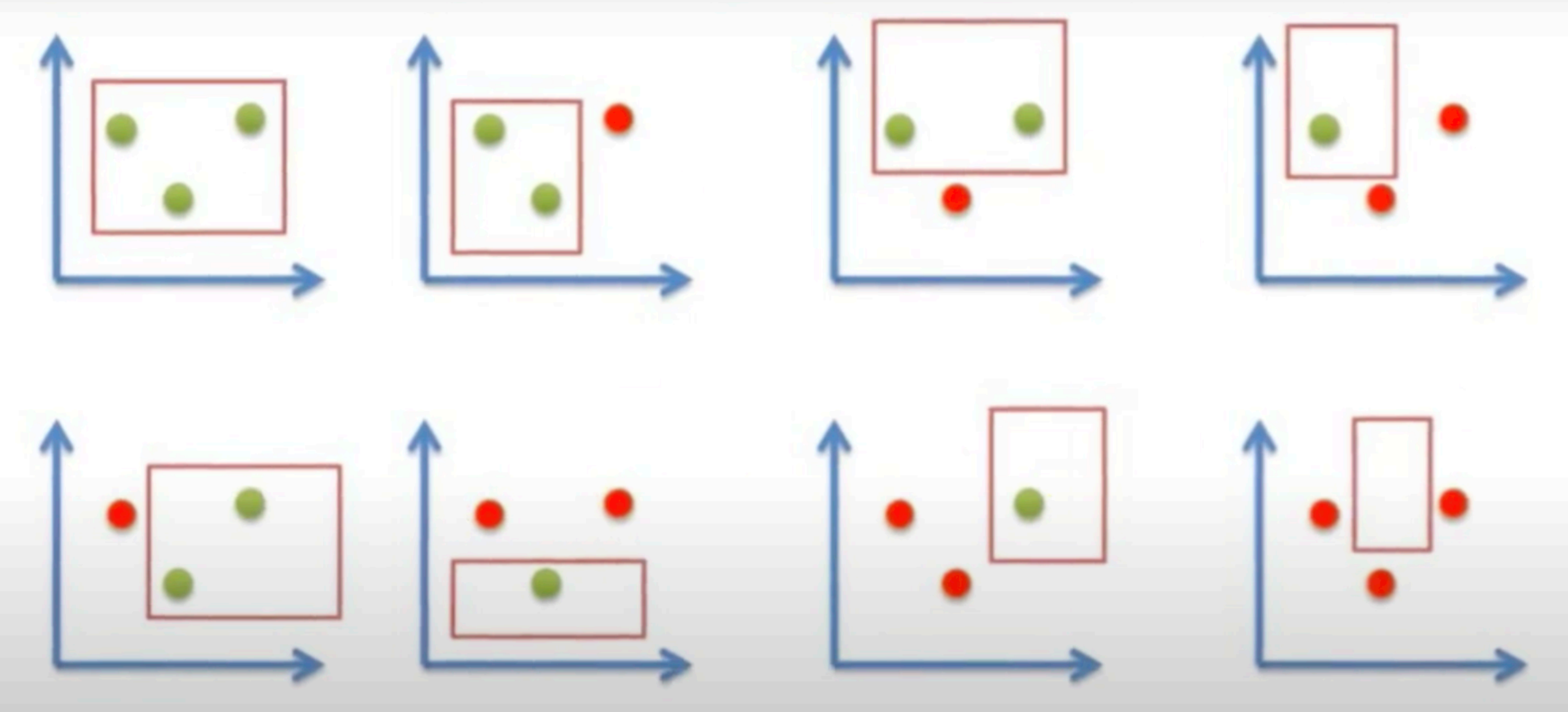


Axis-aligned rectangle: 3 points

- Consider 3 points
- How many possible ways can we classify 3 points?
- Can an axis-aligned rectangle correctly classify all possible ways?

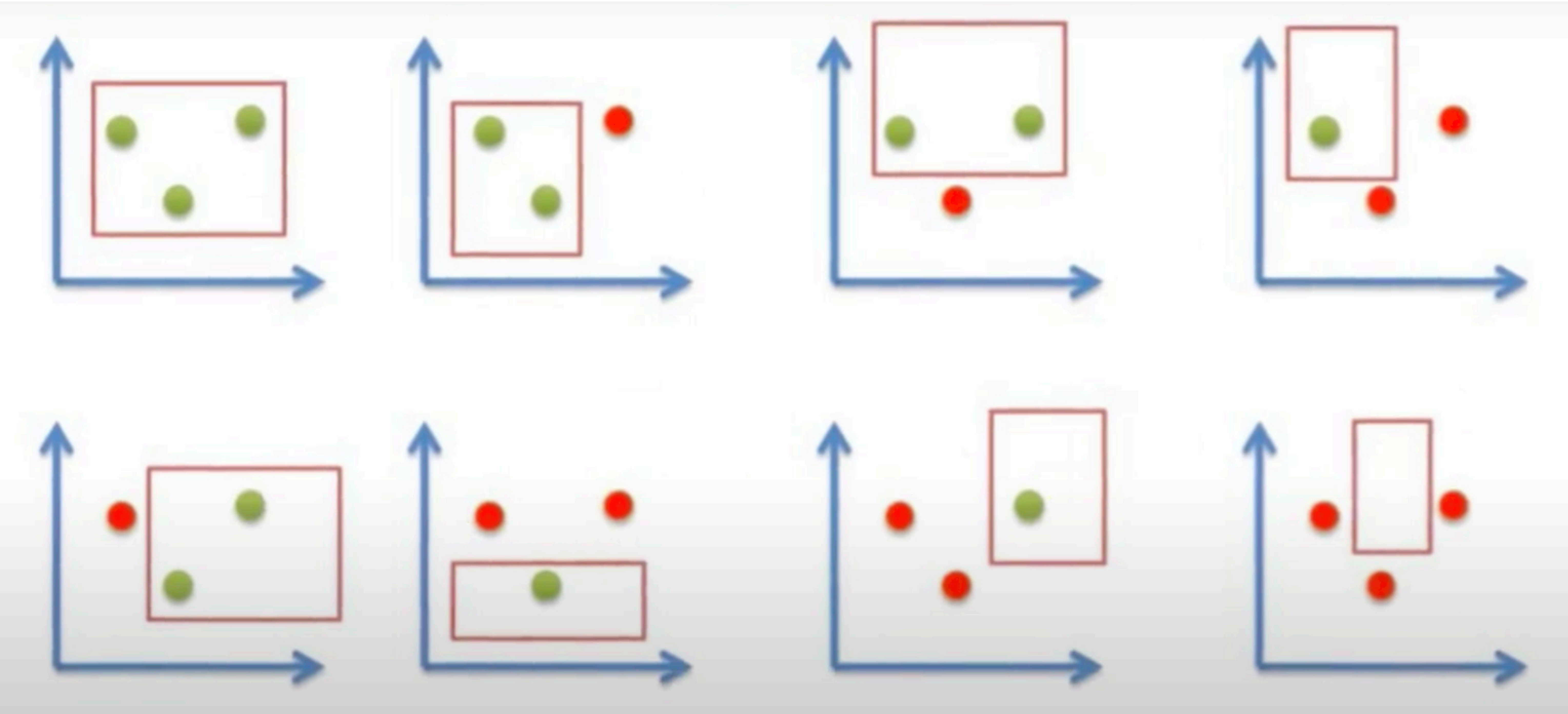
Axis-aligned rectangle: 3 points

- 3 points can be classified in 8 different ways
- It is possible for an axis-aligned rectangle to correctly classify all 8 cases
- Conclusion?



Axis-aligned rectangle: 3 points

- 3 points can be classified in 8 different ways
- It is possible for an axis-aligned rectangle to correctly classify all 8 cases
- **The axis-aligned rectangle can shatter 3 points on a 2-dim plane (R^2)**

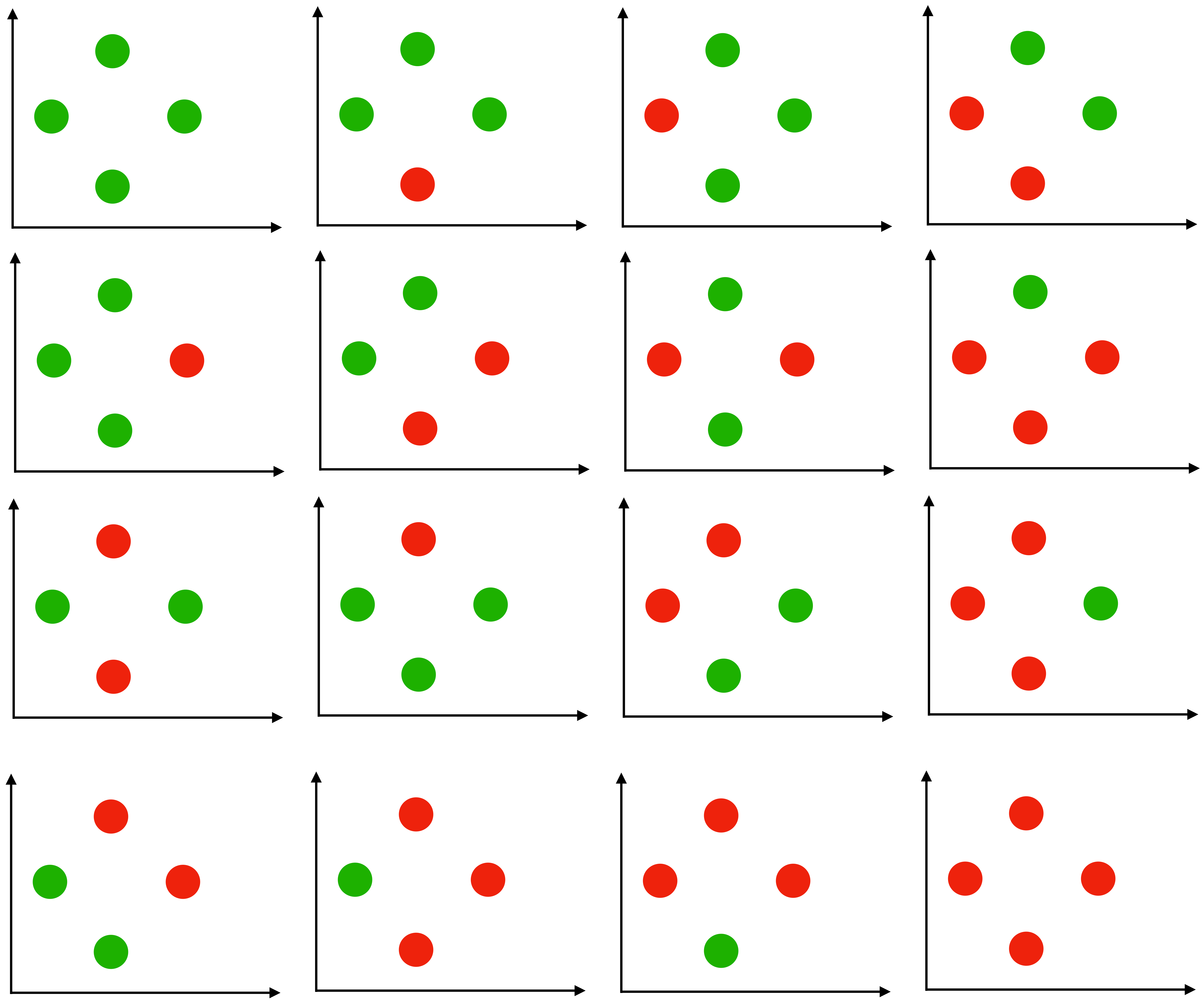


Axis-aligned rectangle: 4 points

- Consider 4 points

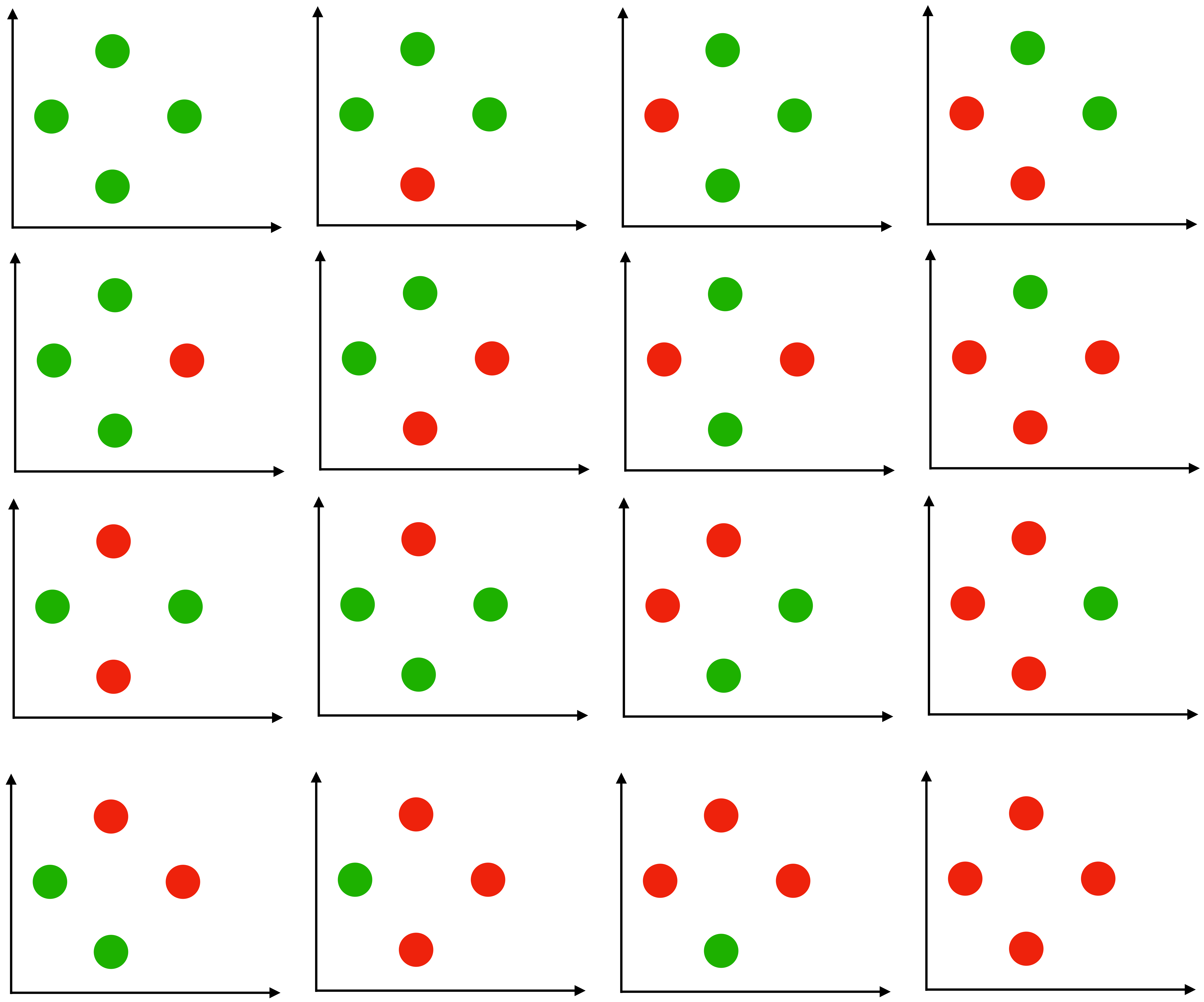
- How many ways can we classify 4 points?

Axis-aligned rectangle: 4 points



4 points can be classified in 16 different ways

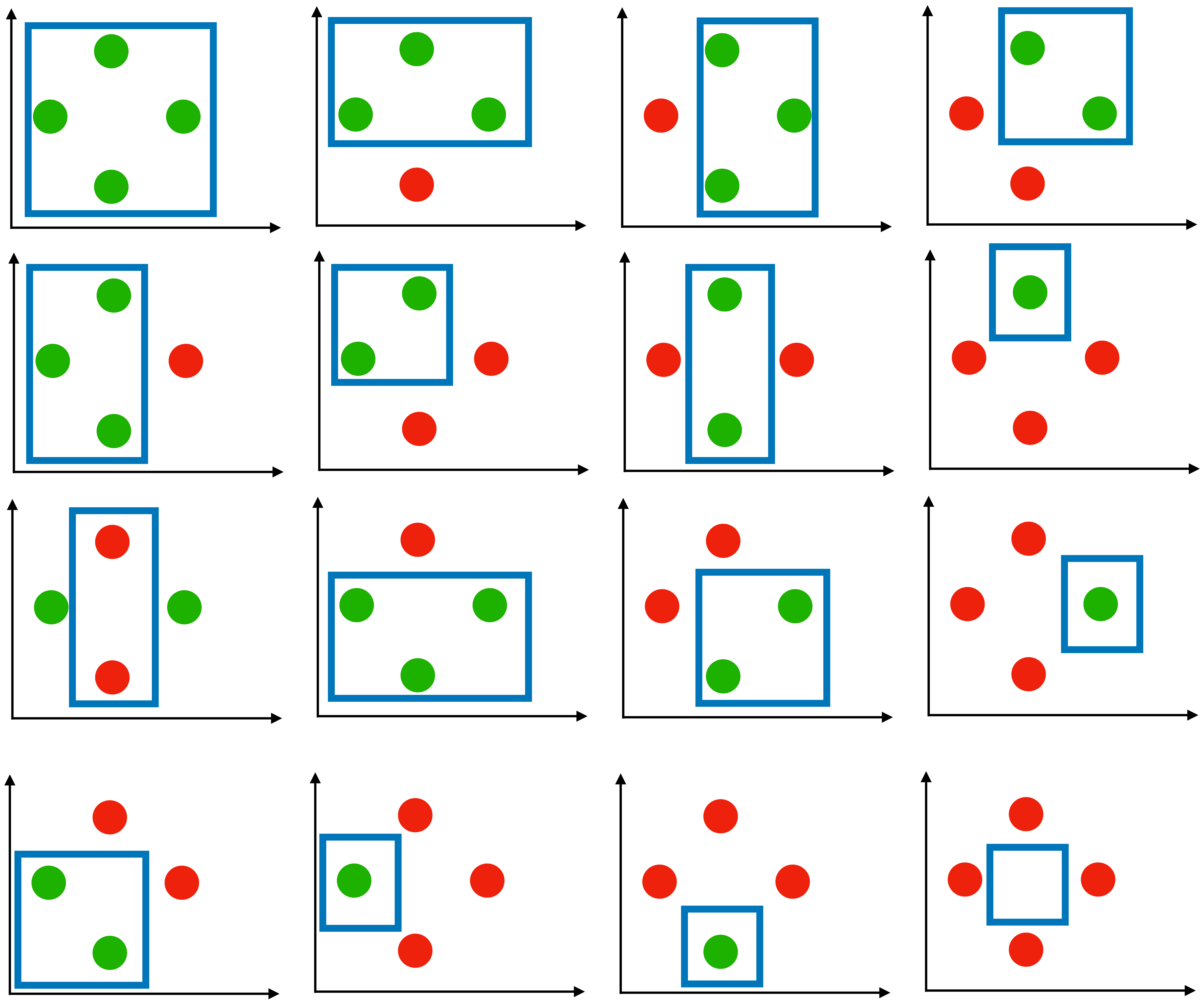
Axis-aligned rectangle: 4 points



4 points can be classified in 16 different ways

Can they all be correctly classified using an axis-aligned rectangle?

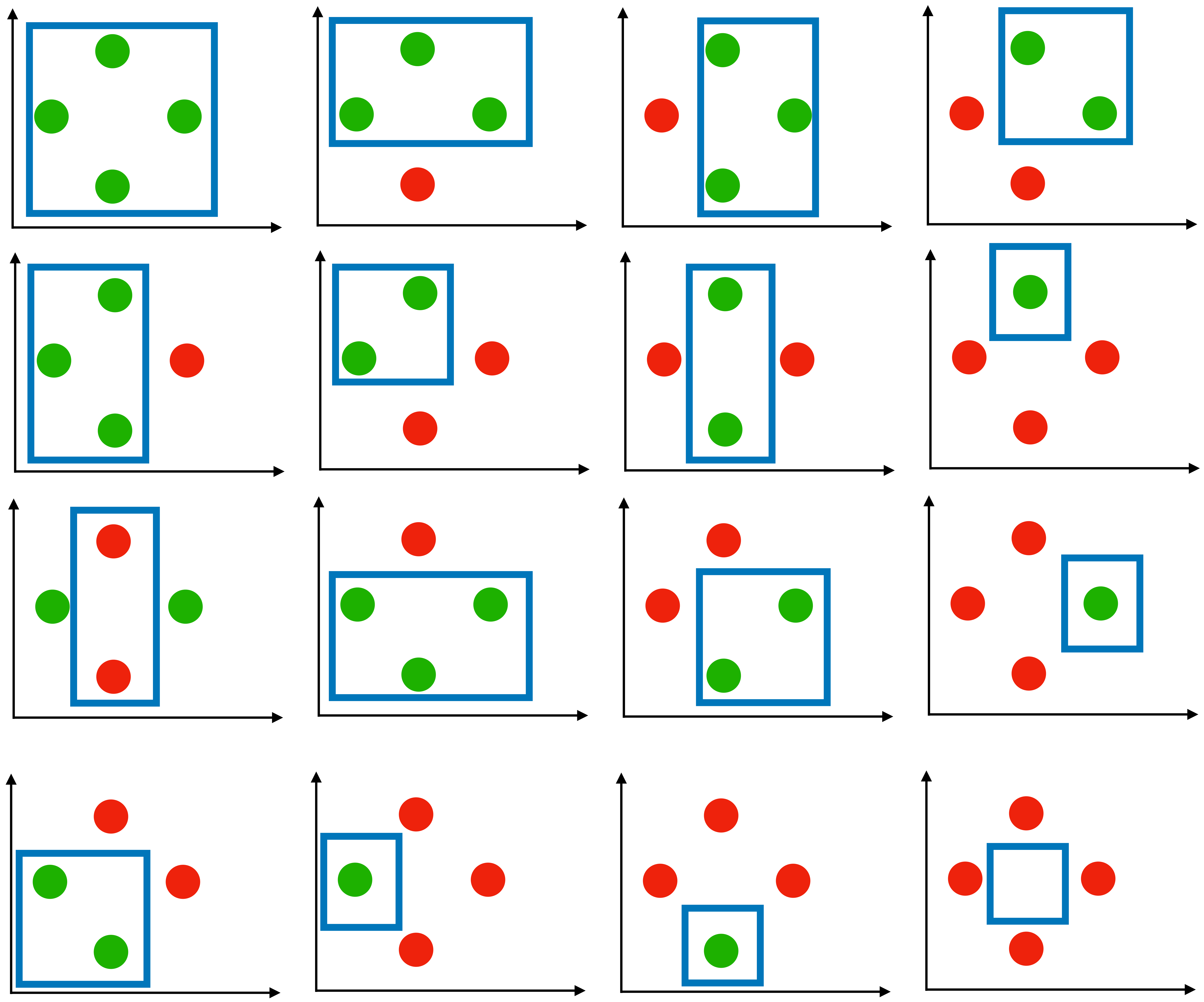
Axis-aligned rectangle: 4 points



- It is possible for an axis-aligned rectangle to correctly classify all 16 cases

- Conclusion?

Axis-aligned rectangle: 4 points



- It is possible for an axis-aligned rectangle to correctly classify all 16 cases
- **The axis-aligned rectangle can shatter 4 points on a 2-dimensional plane (R^2)**

Axis-aligned rectangle: 5 points

- Can we find a set of 5 points that can be shattered by an axis-aligned rectangle?

Axis-aligned rectangle: 5 points

- Can we find a set of 5 points that can be shattered by an axis-aligned rectangle?
- No!

Axis-aligned rectangle

- The maximum number of data points that can be shattered by an axis-aligned rectangle is 4
- VC-dimension(axis-aligned rectangle in 2-dimensional plane R^2) = 4