

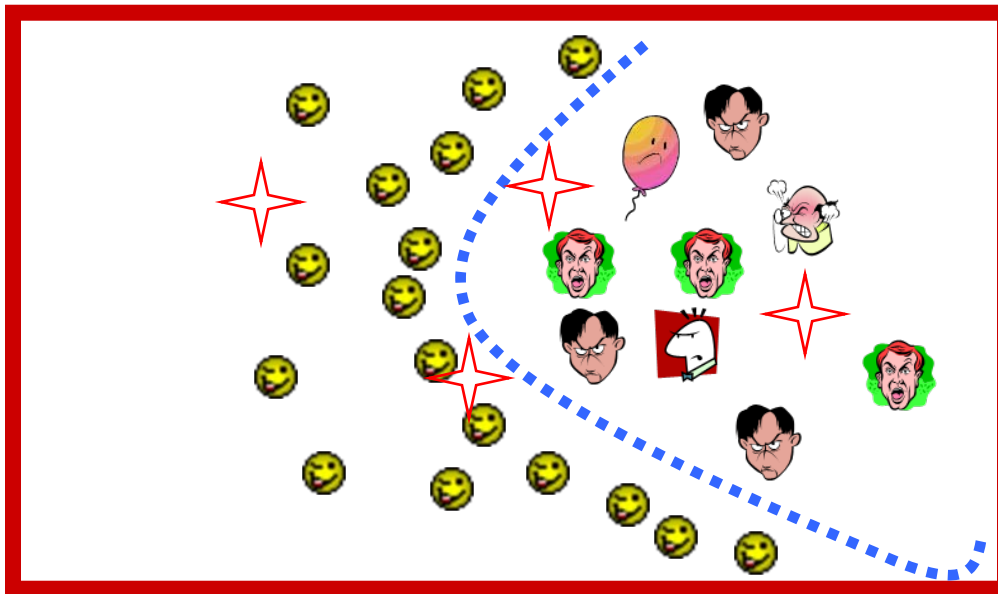
Introduction to Artificial Intelligence  
CSCI 3202  
Fall 2007

# **Introduction to Classification**

Greg Grudic

# This Class: Classification Models

- Collect Training data
- Construct Model: happy =  $F(\text{feature space})$
- **Make a prediction**



*High  
Dimensional  
Feature (input)  
Space*

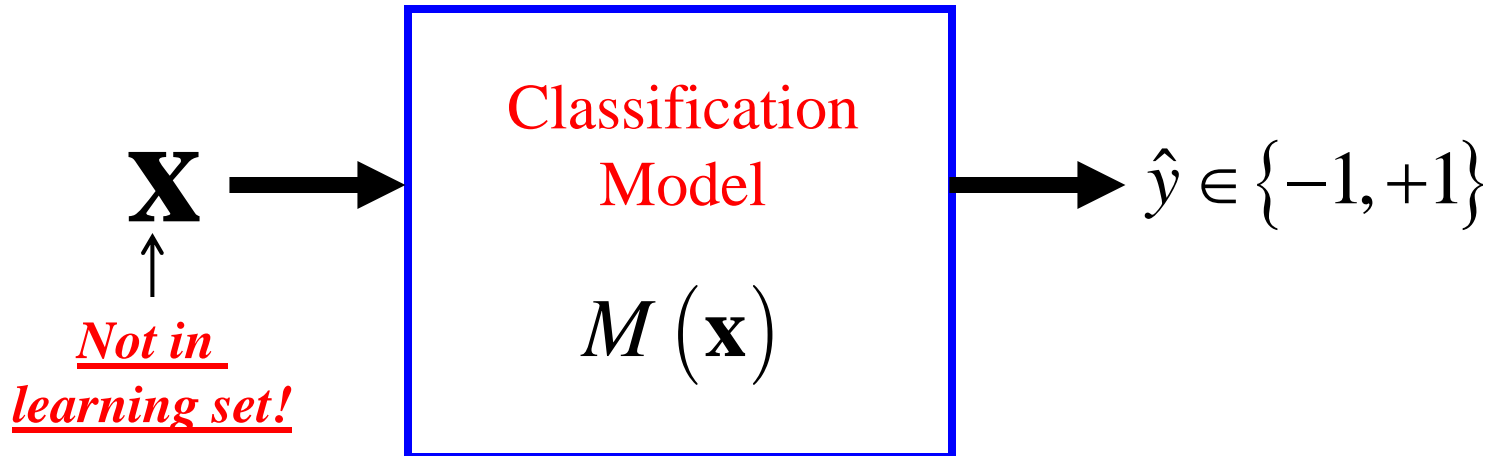
# Binary Classification

- A binary classifier is a mapping from a set of  $d$  inputs to a single output which can take on one of **TWO** values (e.g. path/no path)
- In the most general setting
  - inputs:**  $\mathbf{x} \in \mathbb{R}^d$
  - output:**  $y \in \{-1, +1\}$
- Specifying the output classes as -1 and +1 is arbitrary!
  - Often done as a mathematical convenience

# A Binary Classifier

Given learning data:  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$

A model is constructed:



# Classification Learning Data...

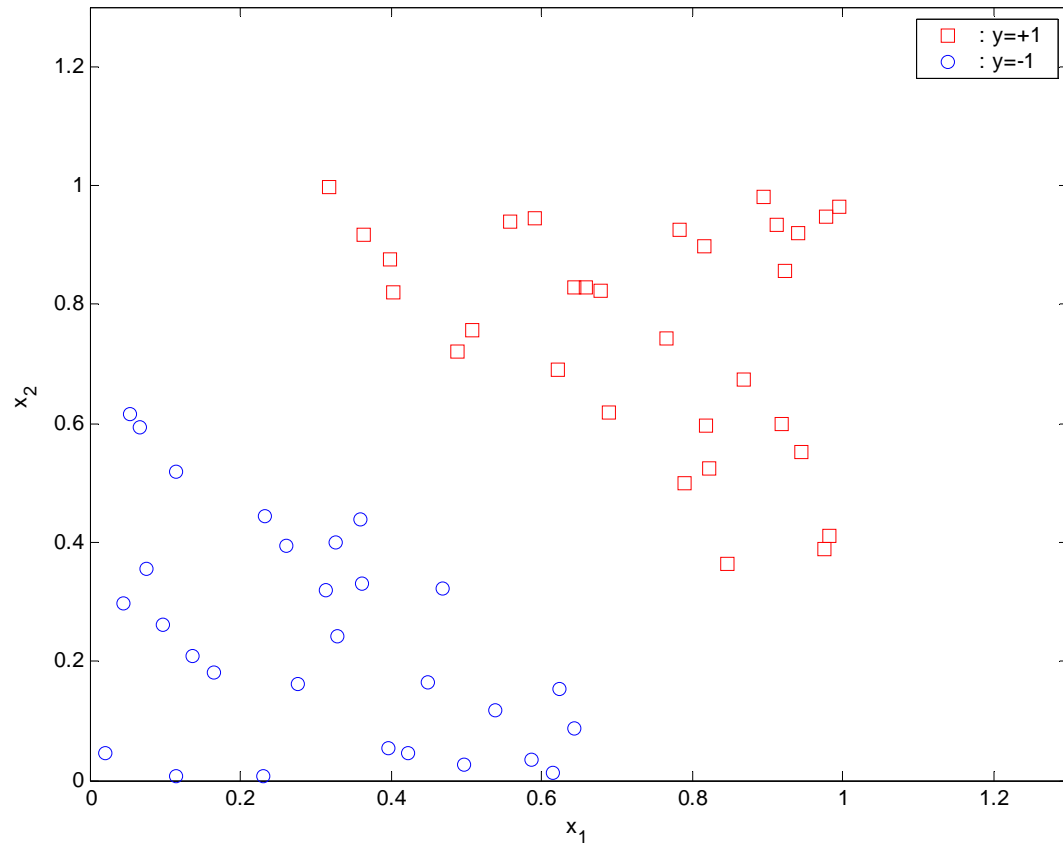
	$x_1$	$x_2$	$y$
<i>Example 1</i>	0.95013	0.58279	1
<i>Example 2</i>	0.23114	0.4235	-1
<i>Example 3</i>	0.8913	0.43291	1
<i>Example 4</i>	0.018504	0.76037	-1
...	...	...	...

# The Learning Data

- Matrix Representation of  $N$  learning examples of  $d$  dimensional inputs

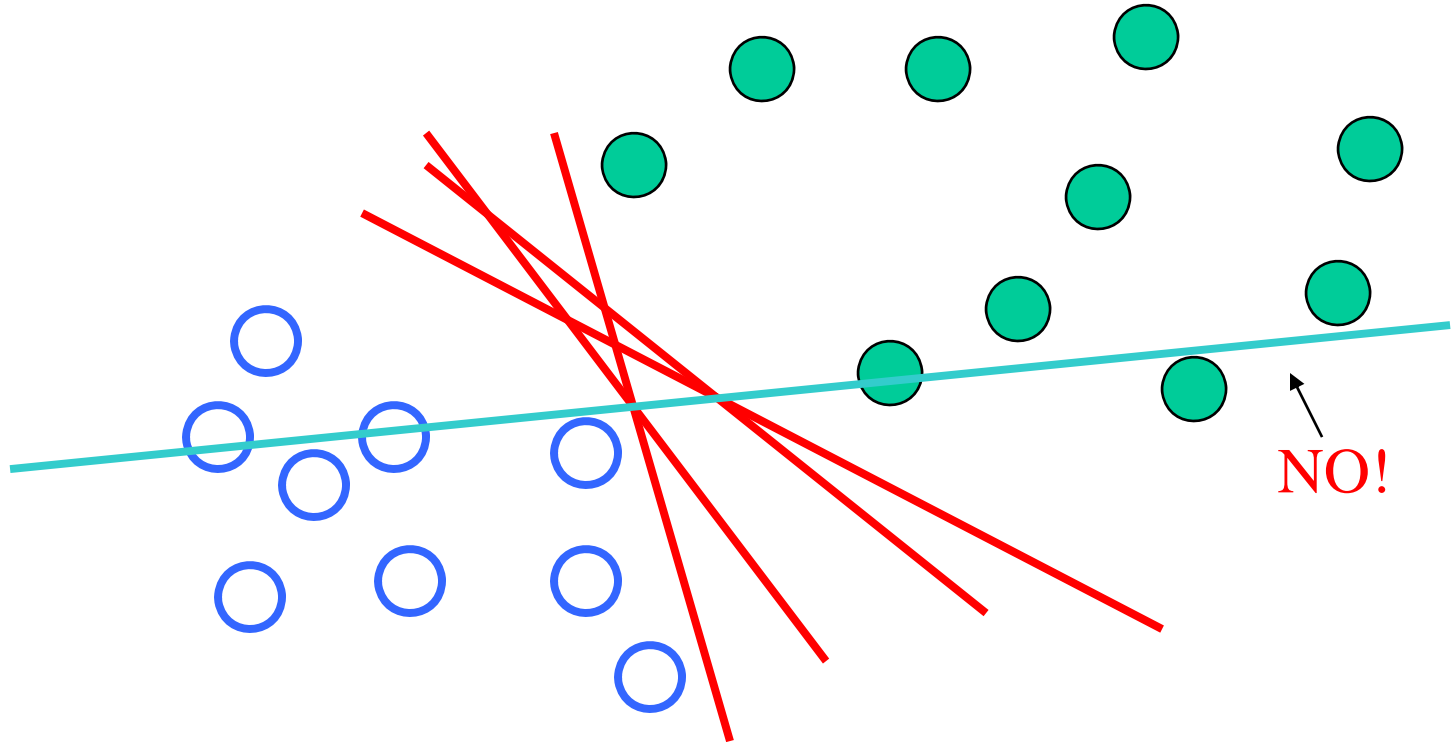
$$\begin{pmatrix} x_{11} & \cdots & x_{1d} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{N1} & \cdots & x_{Nd} & y_N \end{pmatrix}$$

# Graphical Representation of 2D Classification Training Data

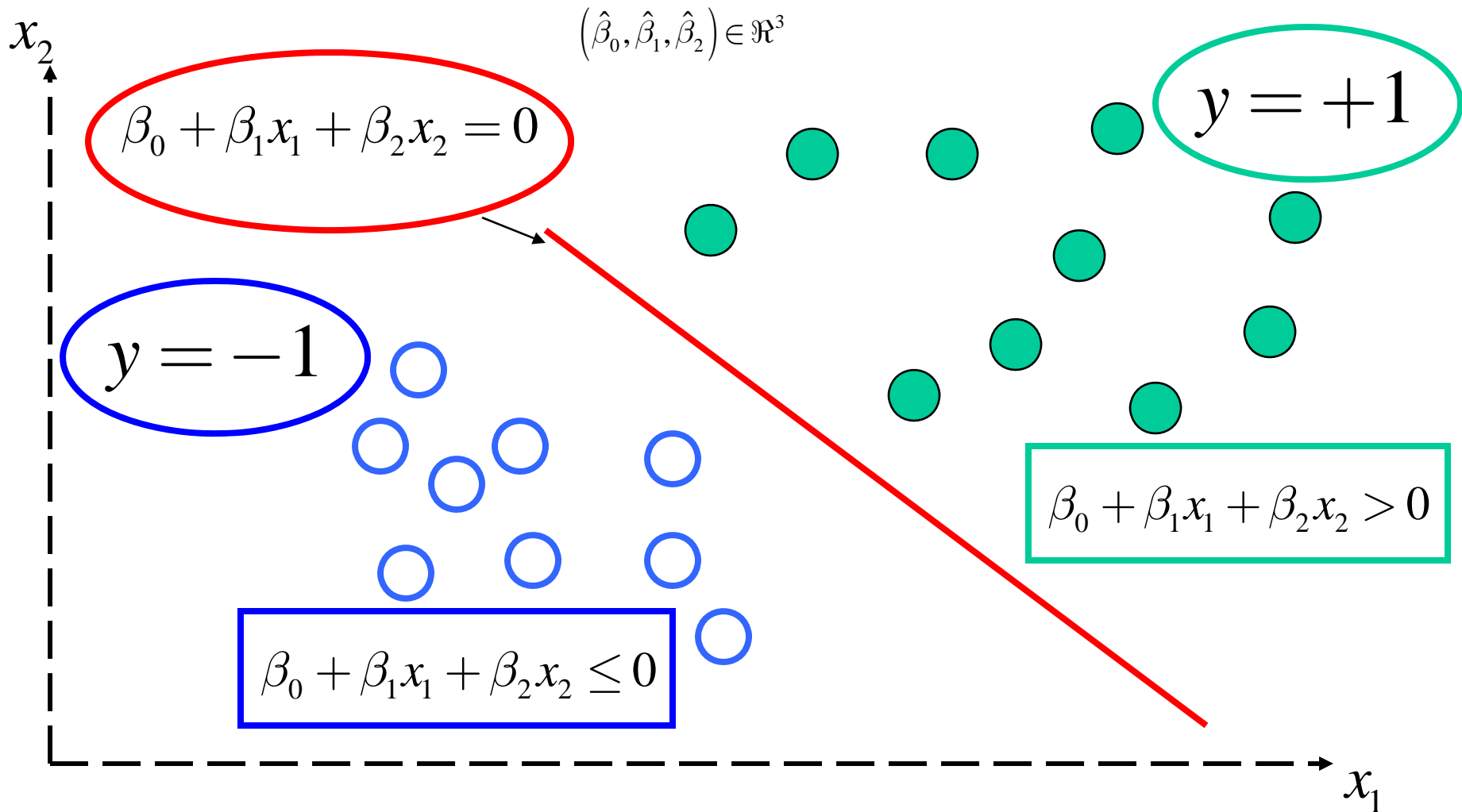


# Linear Separating Hyper-Planes: Discriminative Classifiers

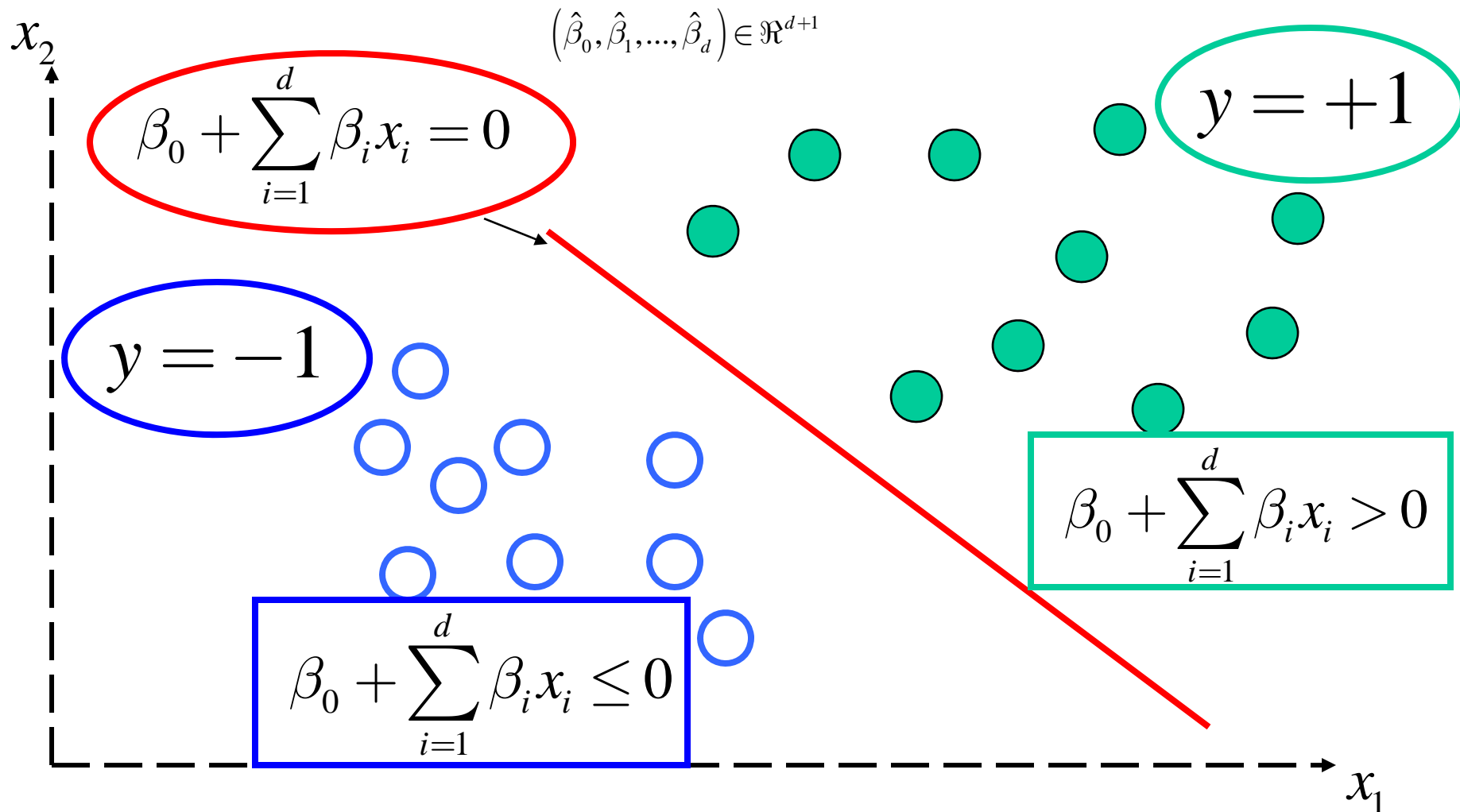
How many lines can separate these points?



# Linear Separating Hyper-Planes (2 dimensions)



# Linear Separating Hyper-Planes (d dimensions)



# Linear Separating Hyper-Planes

- The Model:

$$\hat{y} = M(\mathbf{x}) = \text{sgn}\left[\hat{\beta}_0 + \left(\hat{\beta}_1, \dots, \hat{\beta}_d\right) \cdot \mathbf{x}\right]$$

- Where:

$$\text{sgn}[A] = \begin{cases} 1 & \text{if } A > 0 \\ -1 & \text{otherwise} \end{cases}$$

- The decision boundary:

$$\hat{\beta}_0 + \left(\hat{\beta}_1, \dots, \hat{\beta}_d\right) \cdot \mathbf{x} = \beta_0 + \sum_{i=1}^d \beta_i x_i = 0$$

# Linear Separating Hyper-Planes

- The model parameters are:

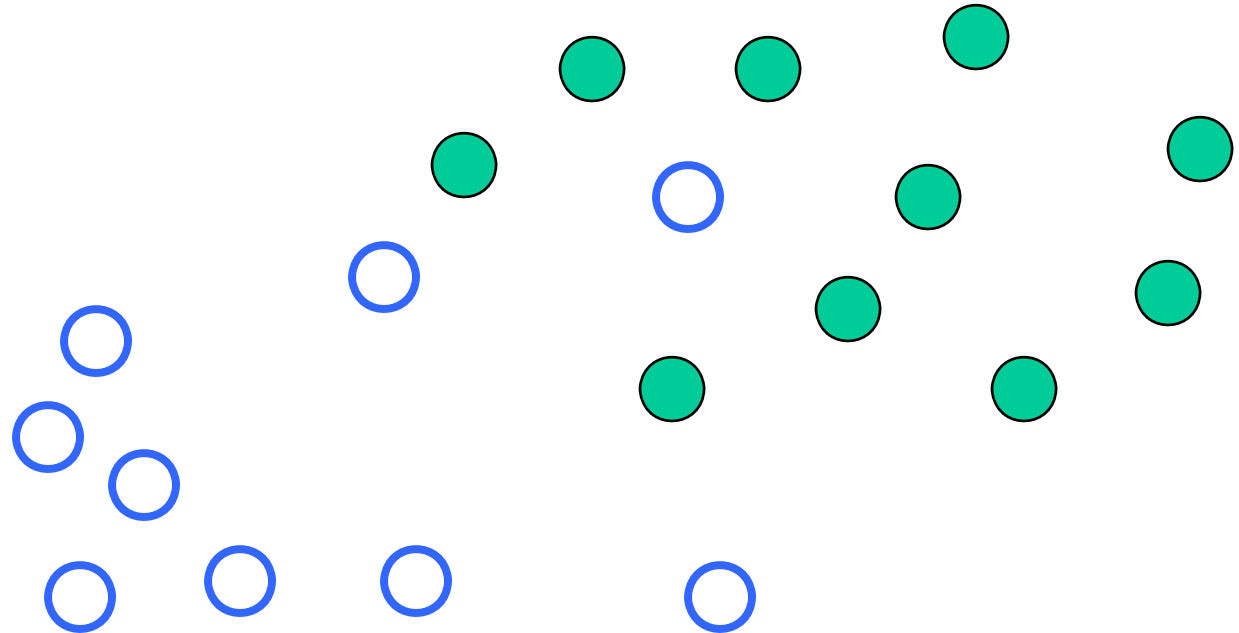
$$\left(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_d\right) \in \mathfrak{R}^{d+1}$$

- The *hat* on the betas means that they are estimated from the data

- Many different learning algorithms have been proposed for determining  $\left(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_d\right)$

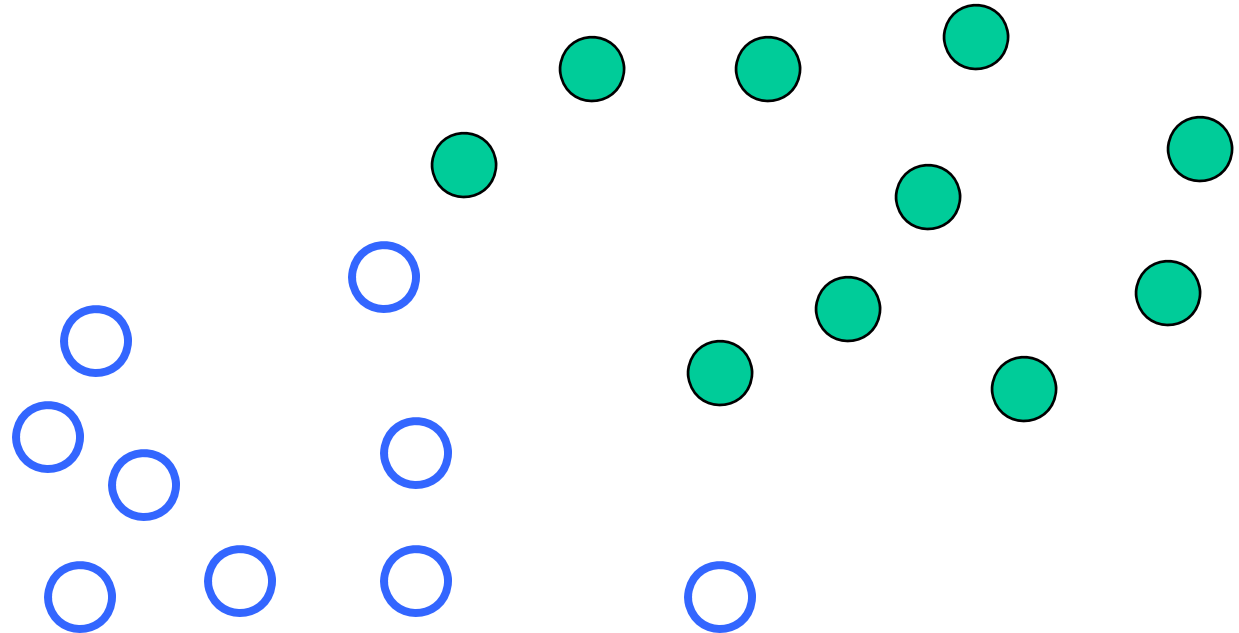
# Is this Data Linearly Separable?

NO!



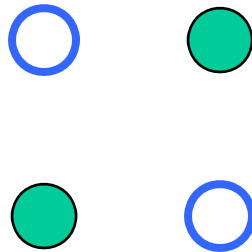
# Is this Data Linearly Separable?

YES!



# Is this Data Linearly Separable?

**NO!**



# Is this Data Linearly Separable?

YES!

