Welcome to Day 5!

Today: AI Classification, Project time, Final Presentations,

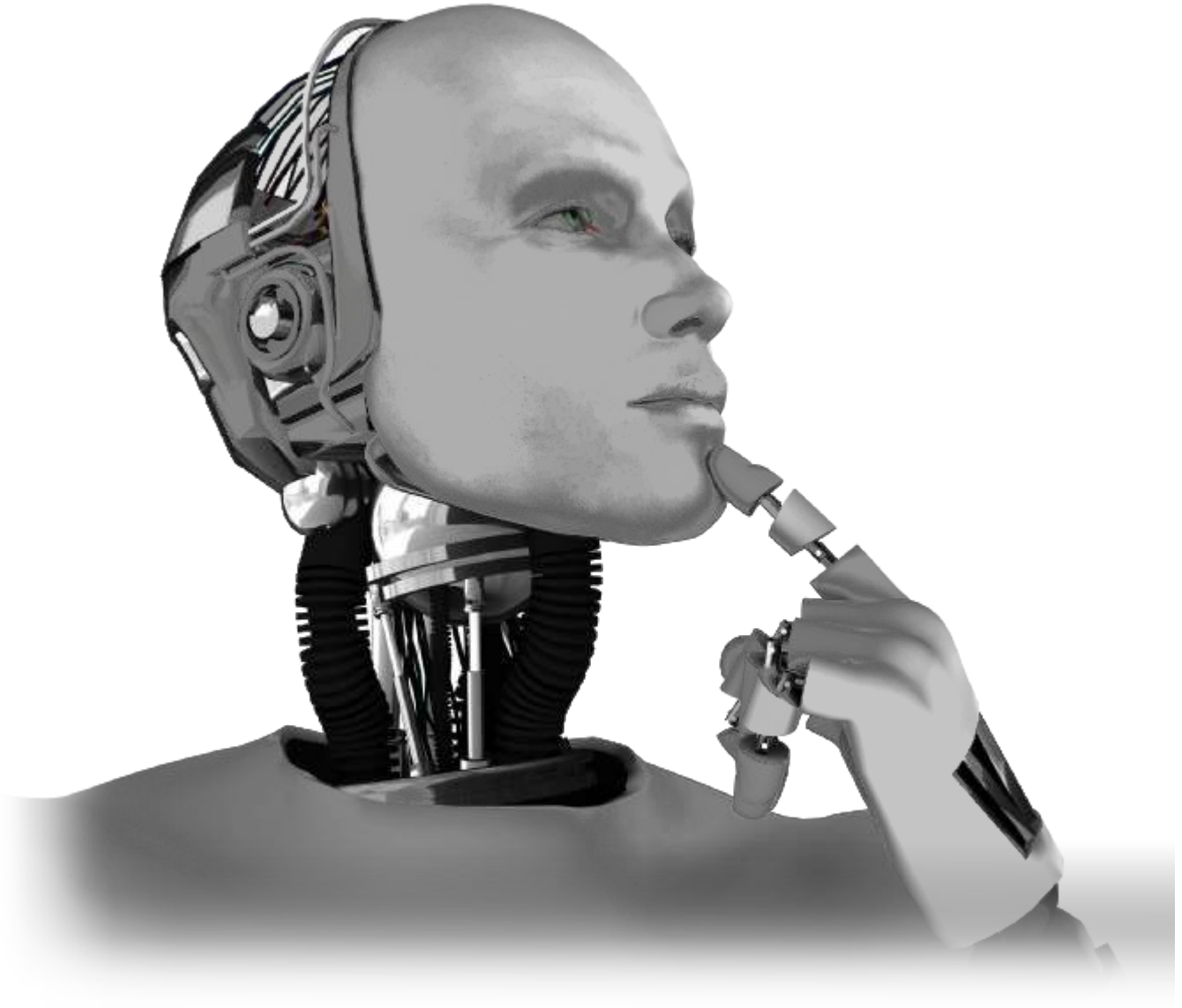
HOW DO THEY KNOW WHAT YOU WANT?

INTRO TO AI/ML

Classification

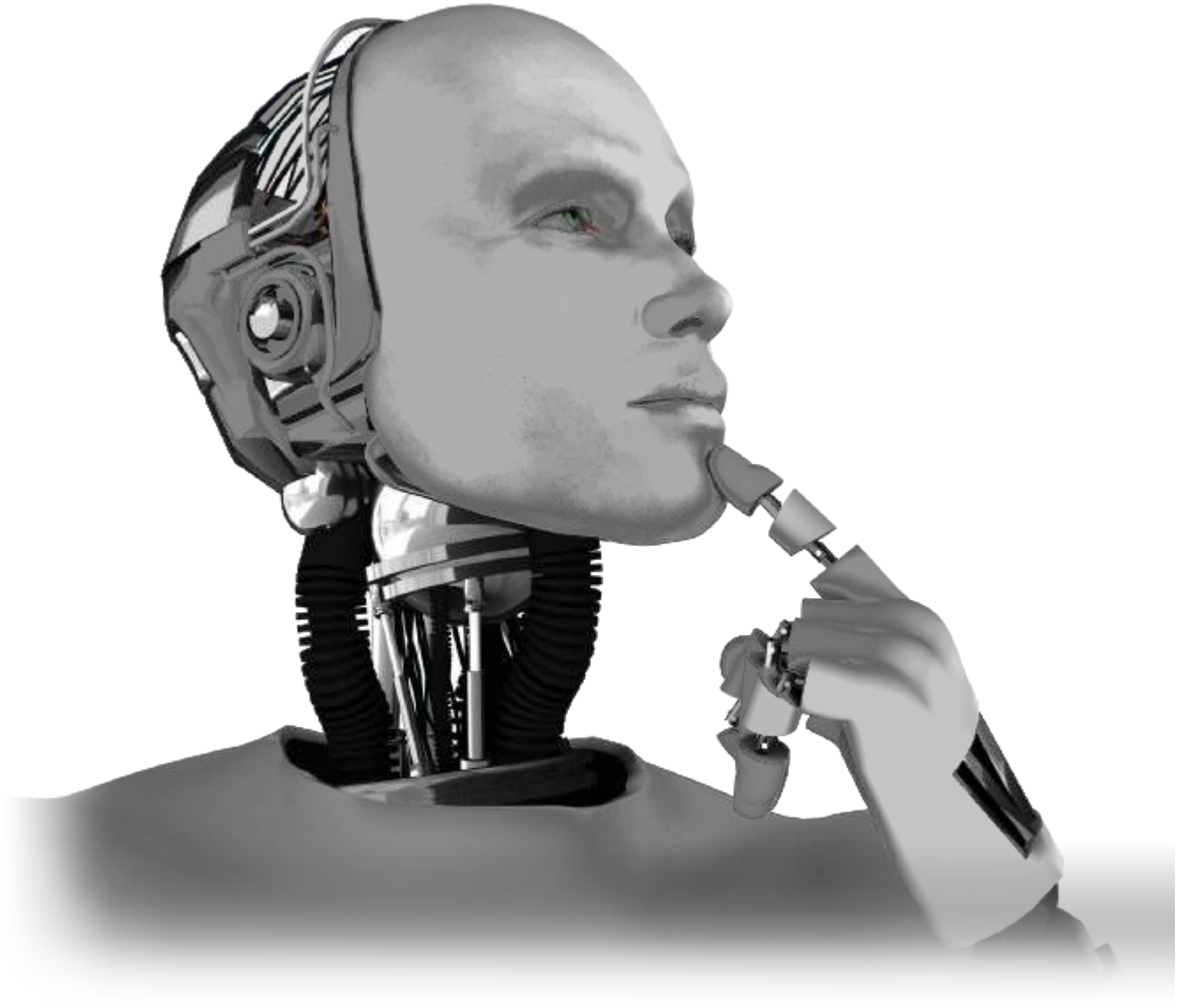
Agenda

- Definition
- Applications
- Classification problem
- Classification systems
- Conclusion
- Lets test ourselves!



Agenda

- Definition
- Applications
- Classification problem
- Classification systems
- Conclusion
- Lets test ourselves!



WHAT IS A.I.?

A branch of computer science that tries to create systems that can model and imitate the intelligent behaviour of the human mind in computers.

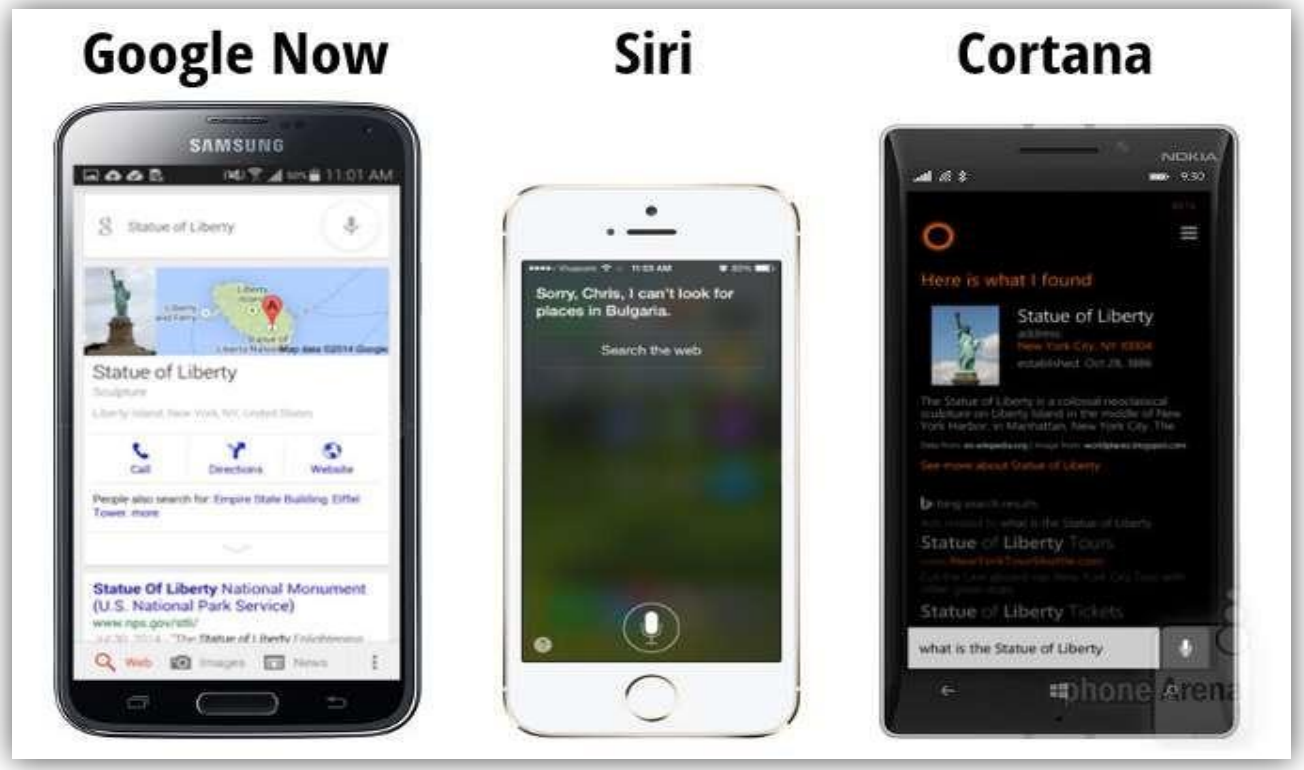
Human beings are the most intelligent creatures

AI enables recognition

data → **Training** → prediction model

 → **Prediction Model** → prediction e.g cat

Applications



Google Now, Siri and **Cortana** are intelligent digital personal assistants on (iOS, Android, and Windows). They all help using **voice**.

We can use them to make calls, send messages, set reminders, take notes, find music, find restaurants, check our calendars, and more.

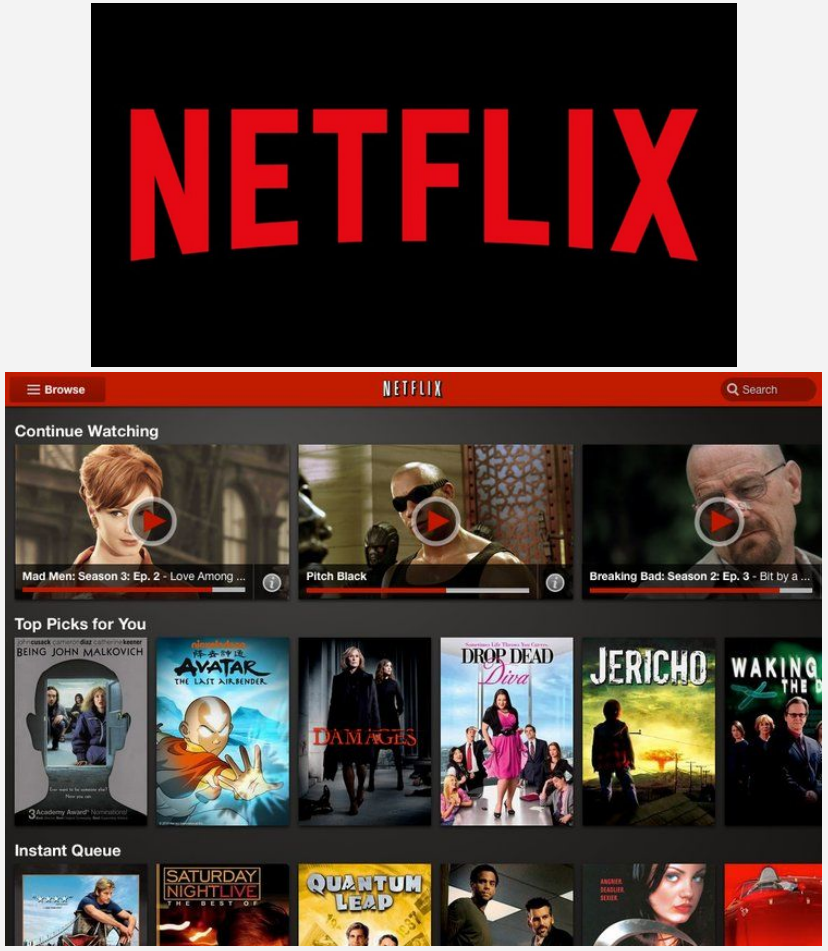
facebook

Gmail

Applications

Spotify

Applications



Netflix is an intelligent system that uses Machine Learning to predict & suggest movie preferences to users.

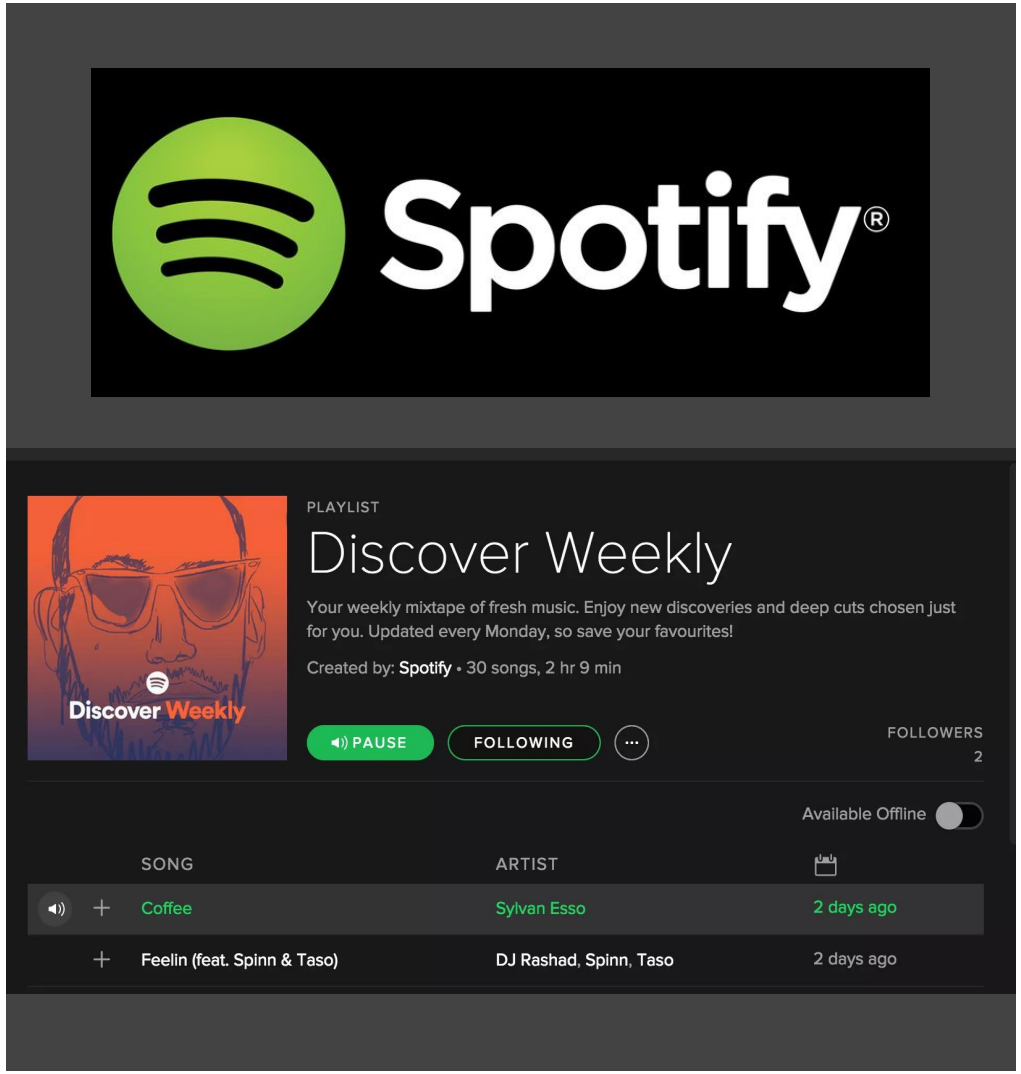
How is Netflix Intelligent?

Applications

Facebook puts content on your Newsfeed and advertises products to you intelligently using Machine Learning.

How is Facebook Intelligent?

Applications

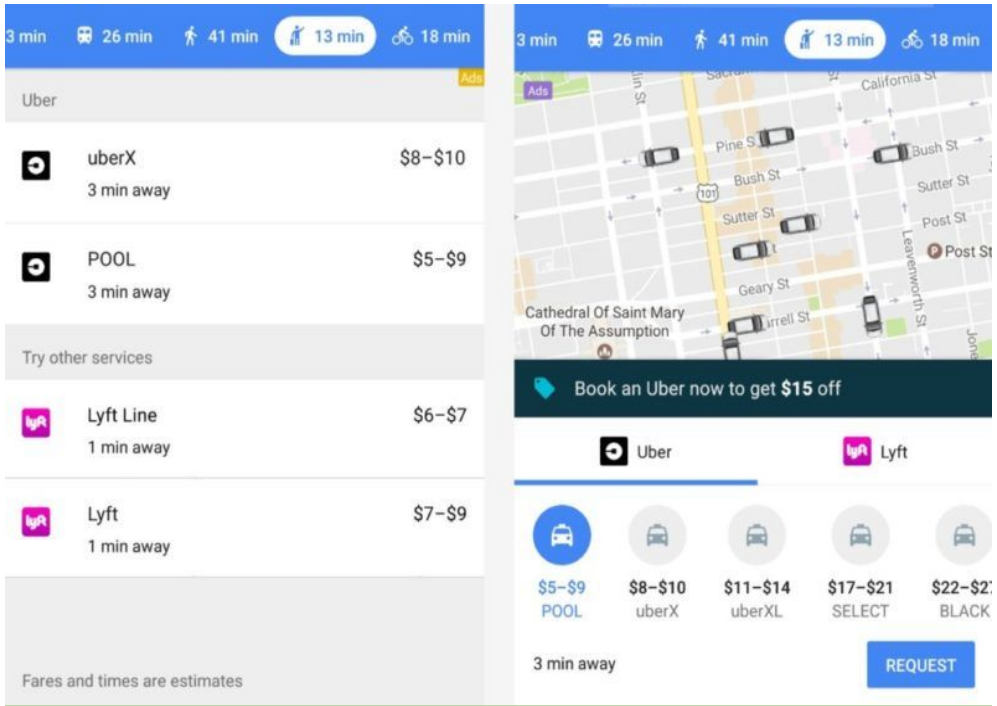


Spotify learns your preferences & filters music lists for you intelligently using Machine Learning.

How is Spotify Intelligent?

OTHER AI APPLICATIONS

Self Driving Cars



Uber Prices

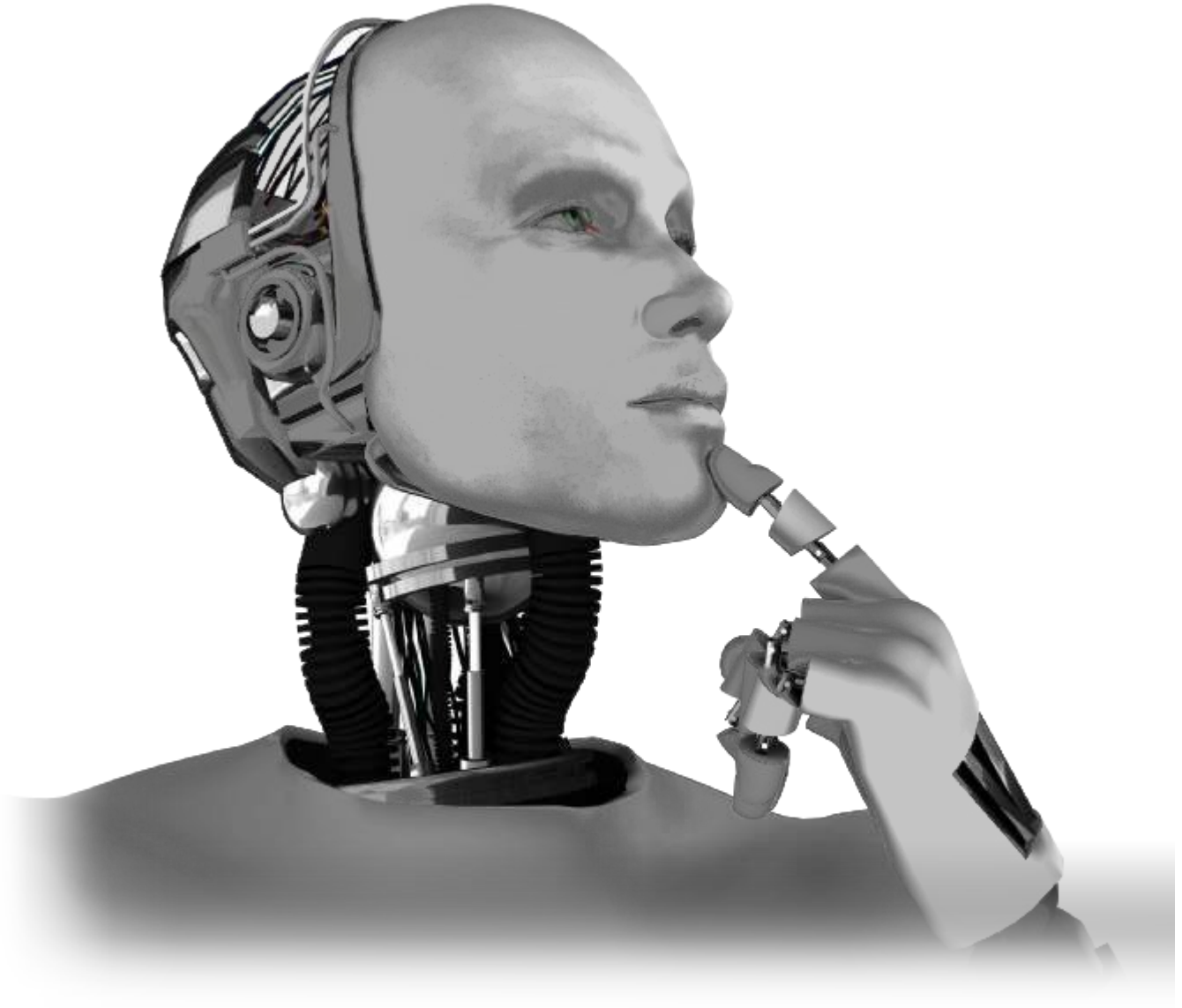
OTHER AI APPLICATIONS

Google Search results

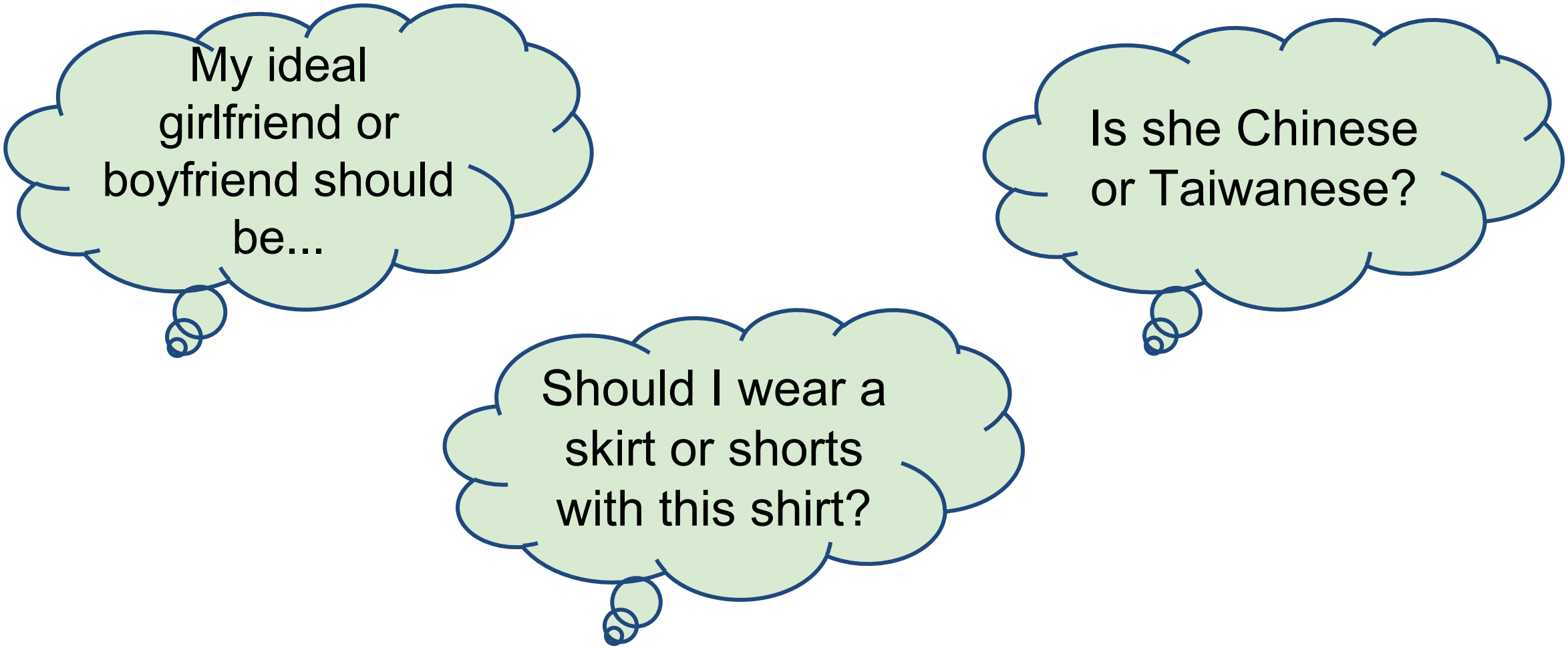
Email Spam Filter

Agenda

- Definition
- Applications
- Classification problem
- Classification systems
- Conclusion
- Lets test ourselves!



CLASSIFICATION



Everyday, all the time we are classifying things.

Classification is a Computable Problem

Example 1: Crossing the street:

- Is there a car coming?
- At what speed?
- How far is it to the other side?

Classification: **Safe to cross or Not safe to cross!!!**

Classification is a Computable Problem

Example 1: Crossing the street:

- Is there a car coming? YES
- At what speed? 80km/hour
- How far is it to the other side? 100 meters

Classification? **Not safe to cross!!!**

Classification is a Computable Problem

Example 2: Nationality:

- Does she only speak English?
- Only knows American music?
- Says “bless you” when someone sneezes?

Classification: **American or Not American!!!**

Classification is a Computable Problem

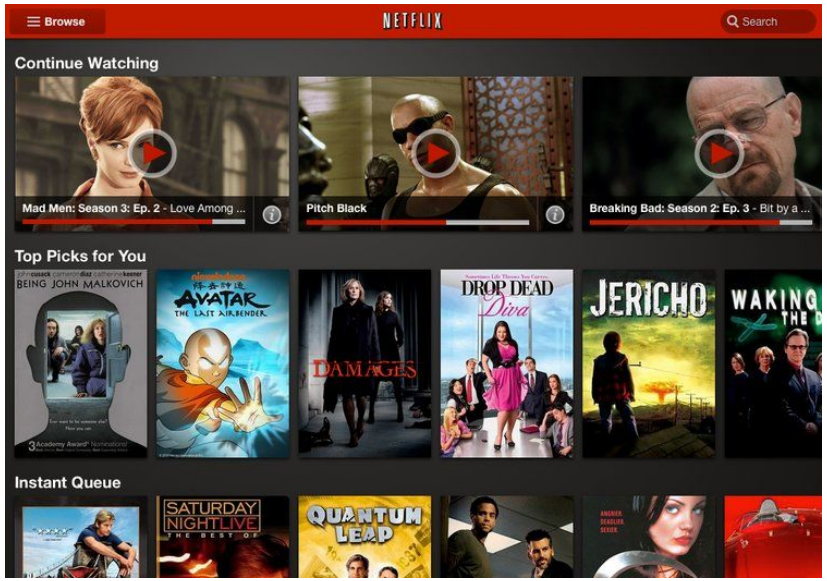
Example 2: Nationality:

- Does she only speak English? NO
- Only knows American music? YES
- Says “bless you” when someone sneezes? YES

Classification? **American**

Classification is a Computable Problem

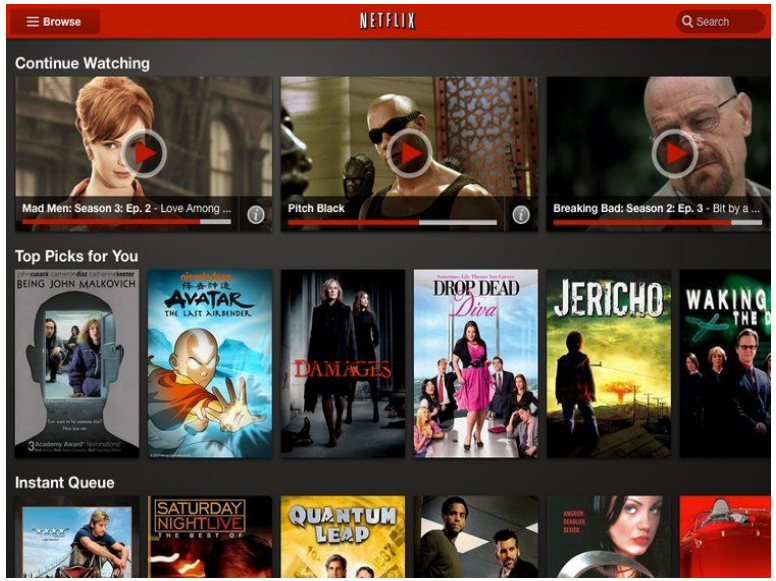
Example 3: Netflix is always computing movies that a user might like:



Classification: **Nitah will enjoy this movie or Not!!!**

Classification is a Computable Problem

Example 3: Assuming Nitah rarely watches horror and there's a movie that:



- Is not in "My Top Picks" section
- Has a rating of 54%
- Is a horror movie

Classification? **Nitah will NOT enjoy this movie**

CLASSIFICATION SYSTEMS

1. Nearest Neighbors (kNNs)
2. Identification trees (ID Trees)
3. Neural Nets
4. Support Vector Machines (SVMs)

To classify, Learning comes first

LEARNING

Have:

Data of all students in this class with their description e.g language spoken, favorite food and nationality

Goal:

Get to know what features distinguish a Korean from a Taiwanese

CLASSIFYING

Have:

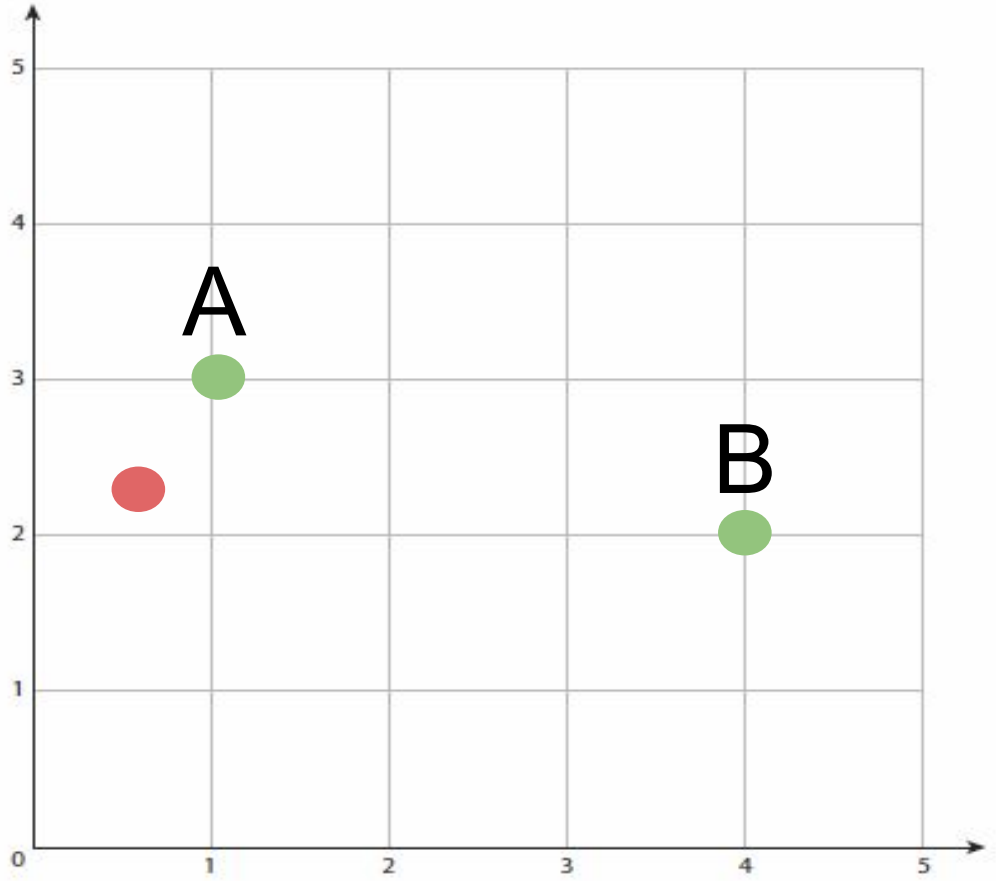
A new student with a list of their description e.g Chris speaks Korean and likes kimchi soup

Goal:

Determine if Chris is Korean or Taiwanese

1. Nearest Neighbors (kNNs)

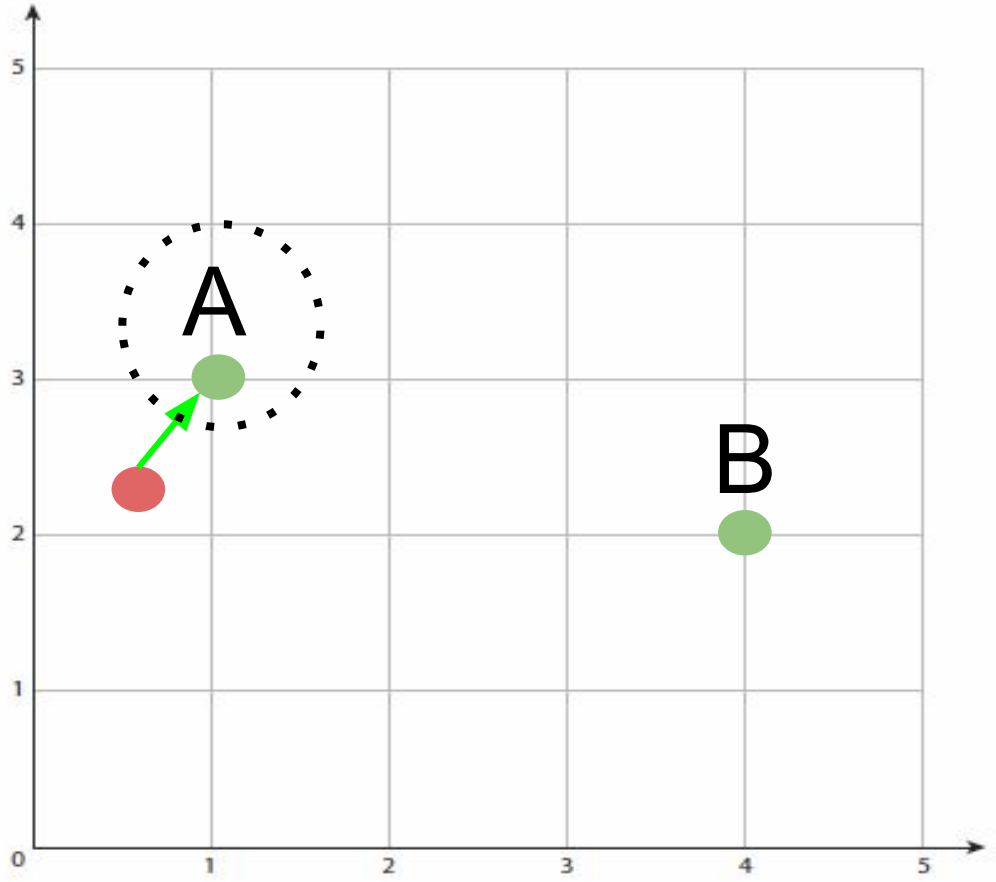
A new item is classified based on the nearest neighbor(s)



GOAL: Classify the red point as either type A or B?

Nearest Neighbors: 1-NN

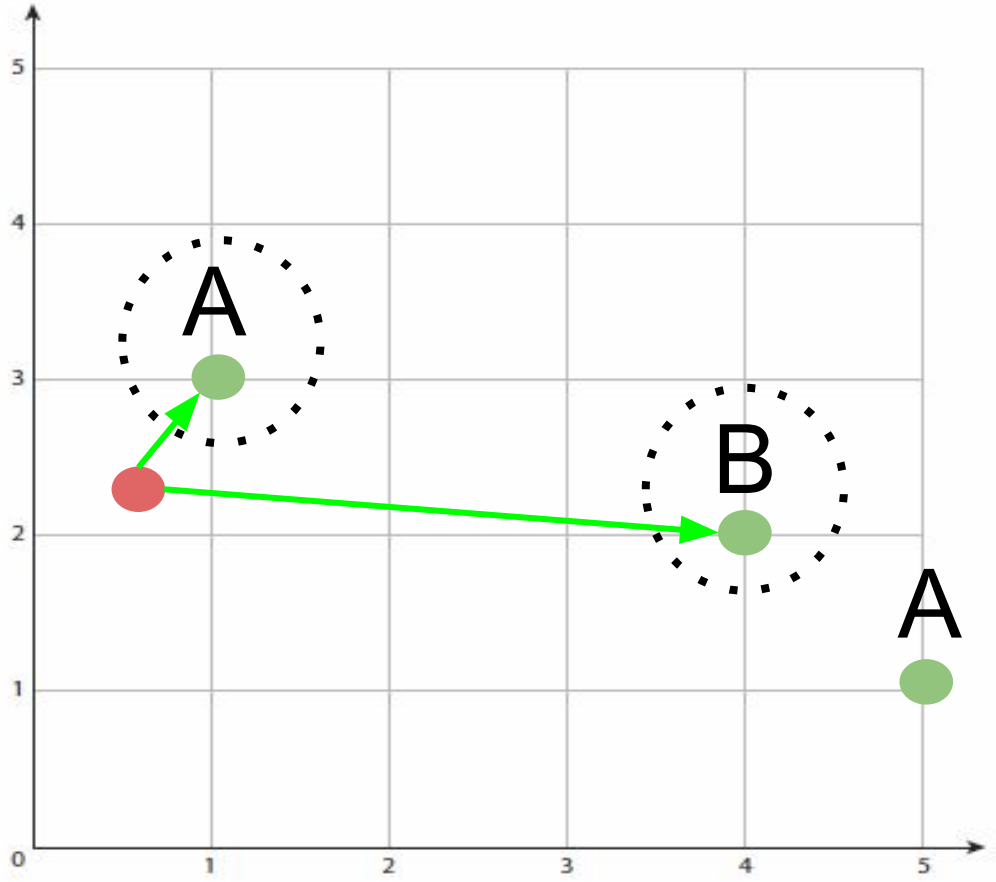
1-NN classifies a new item based on one “the nearest” neighbor



ANSWER: **Type A**

Nearest Neighbors: 2-NN

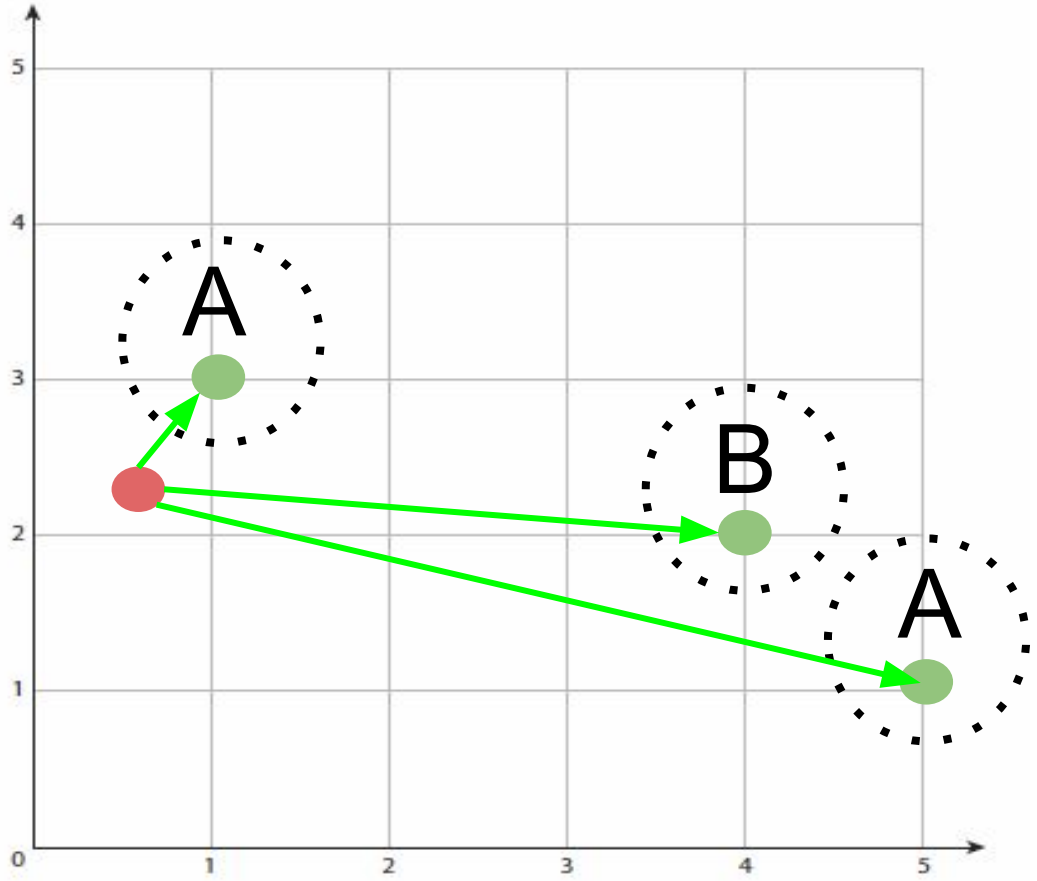
2-NN classifies a new item based on <u>two</u> “nearest” neighbors



ANSWER: **Tie/None/Not sure**

Nearest Neighbors: 3-NN

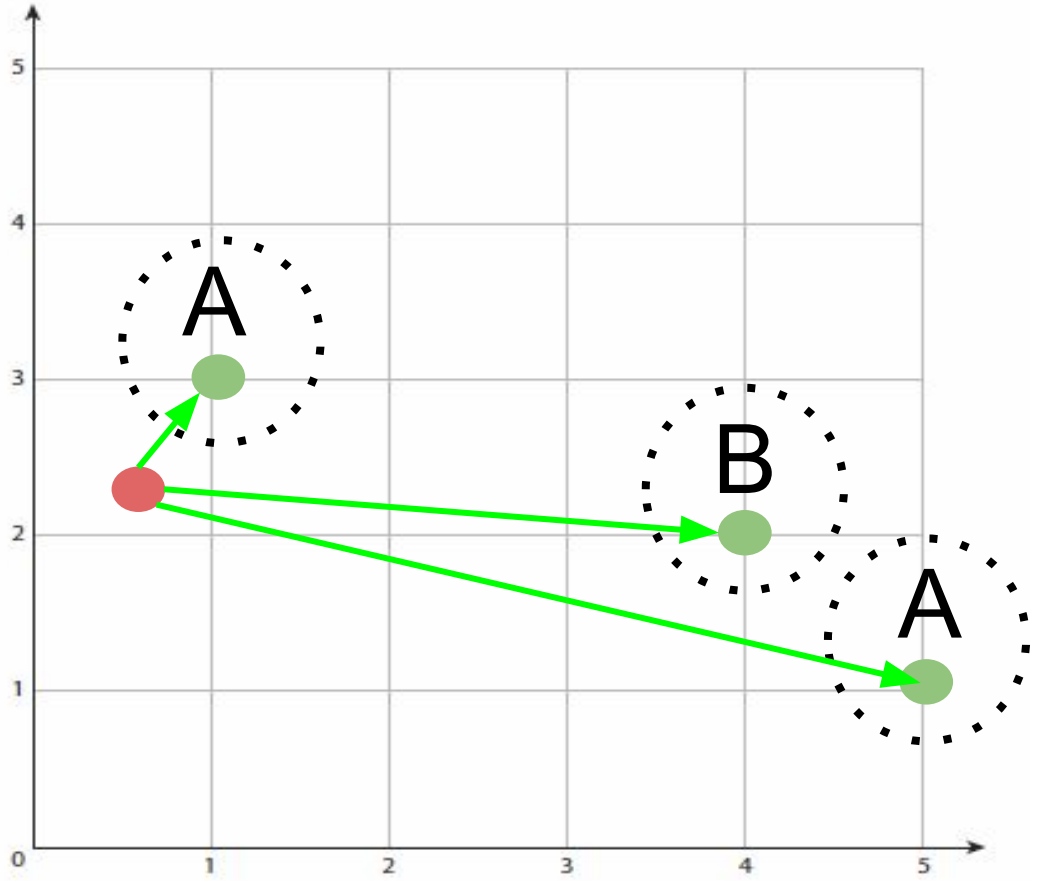
3-NN classifies a new item based on <u>three</u> “nearest” neighbors



ANSWER: **Type A**

Nearest Neighbors: 4-NN

4-NN classifies a new item based on <u>four</u> “nearest” neighbors

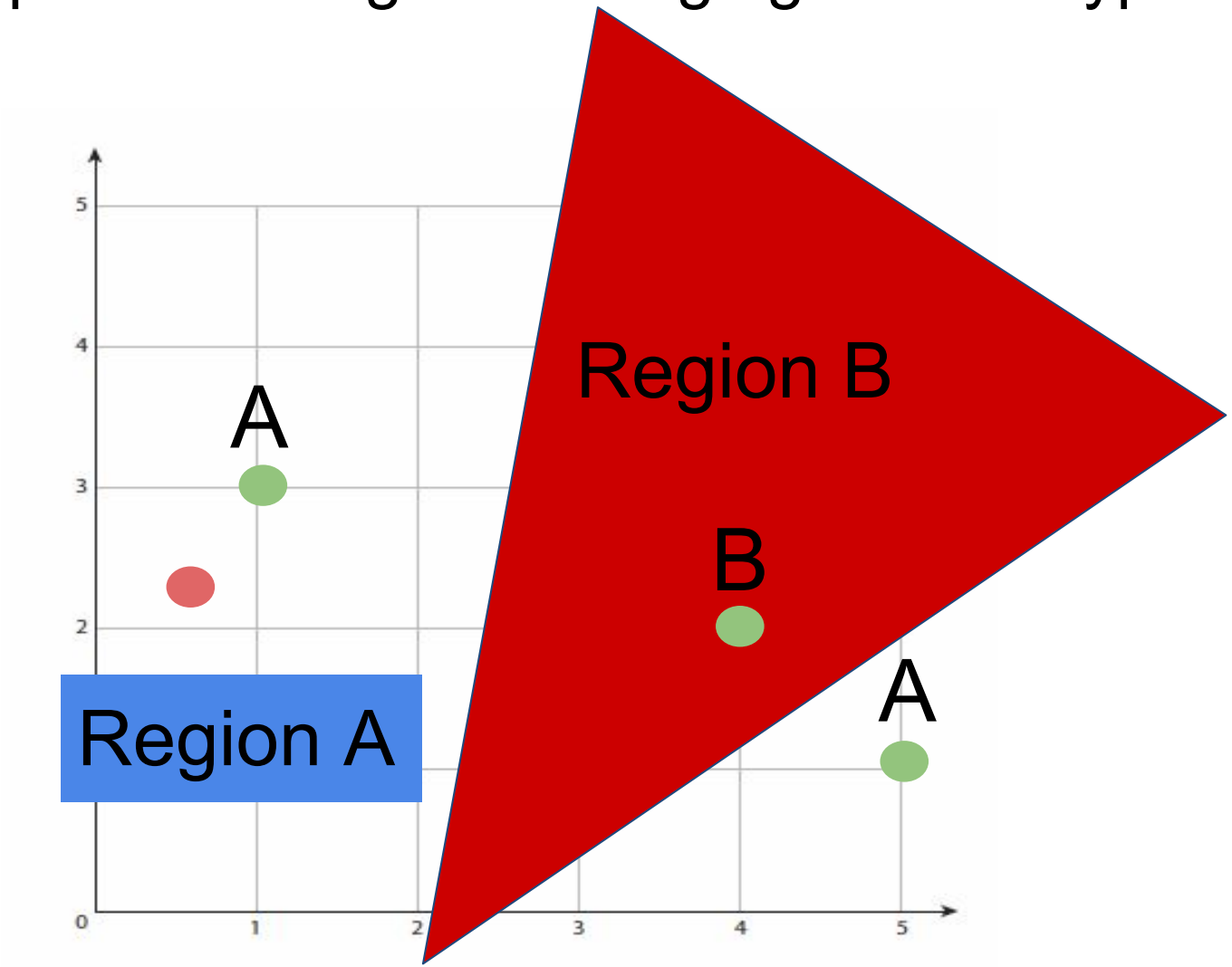


4th neighbor missing

ANSWER: **Unknown**

How a computer classifies using kNNs

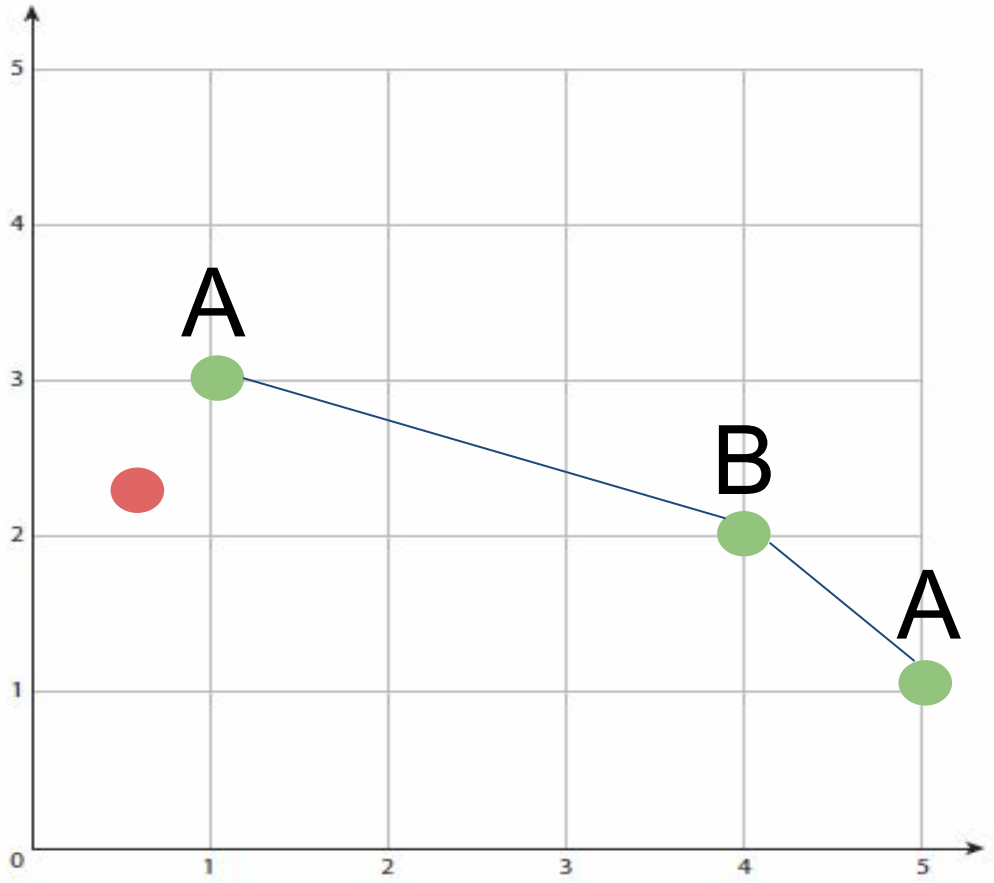
It computes the region belonging to each type



How a computer classifies using kNNs

It computes the region belonging to each type

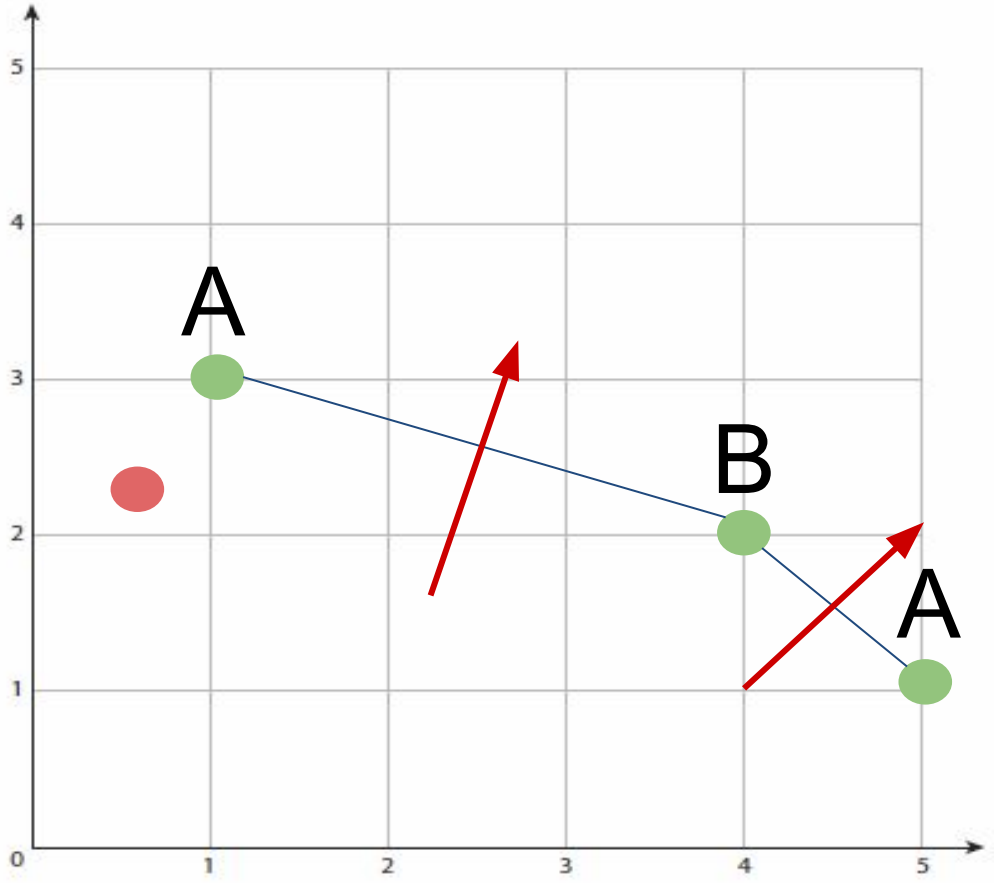
STEP 1: For each pair of different points, draw a straight line joining the two points



How a computer classifies using kNNs

It computes the region belonging to each type

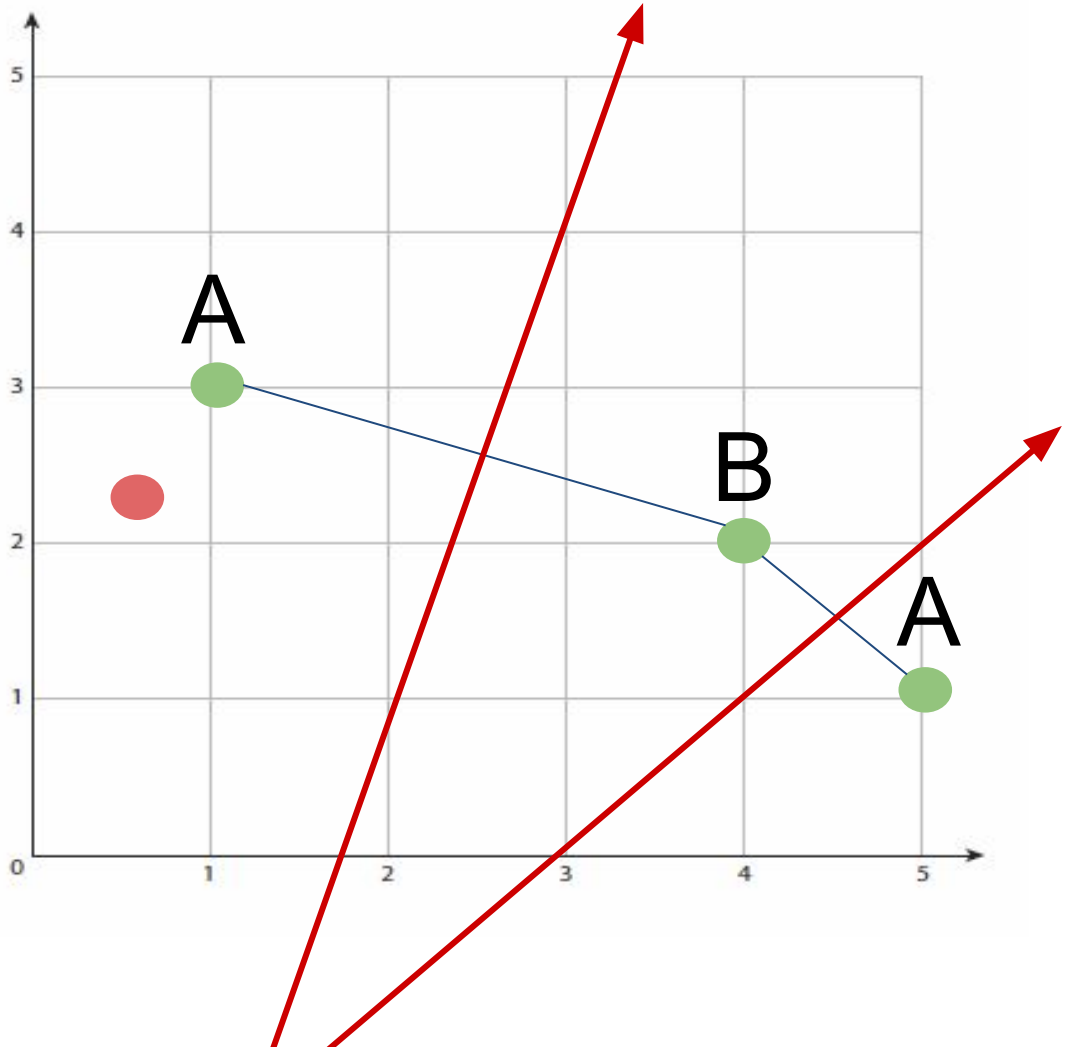
STEP 2: Draw the perpendicular bisectors of the lines



How a computer classifies using kNNs

It computes the region belonging to each type

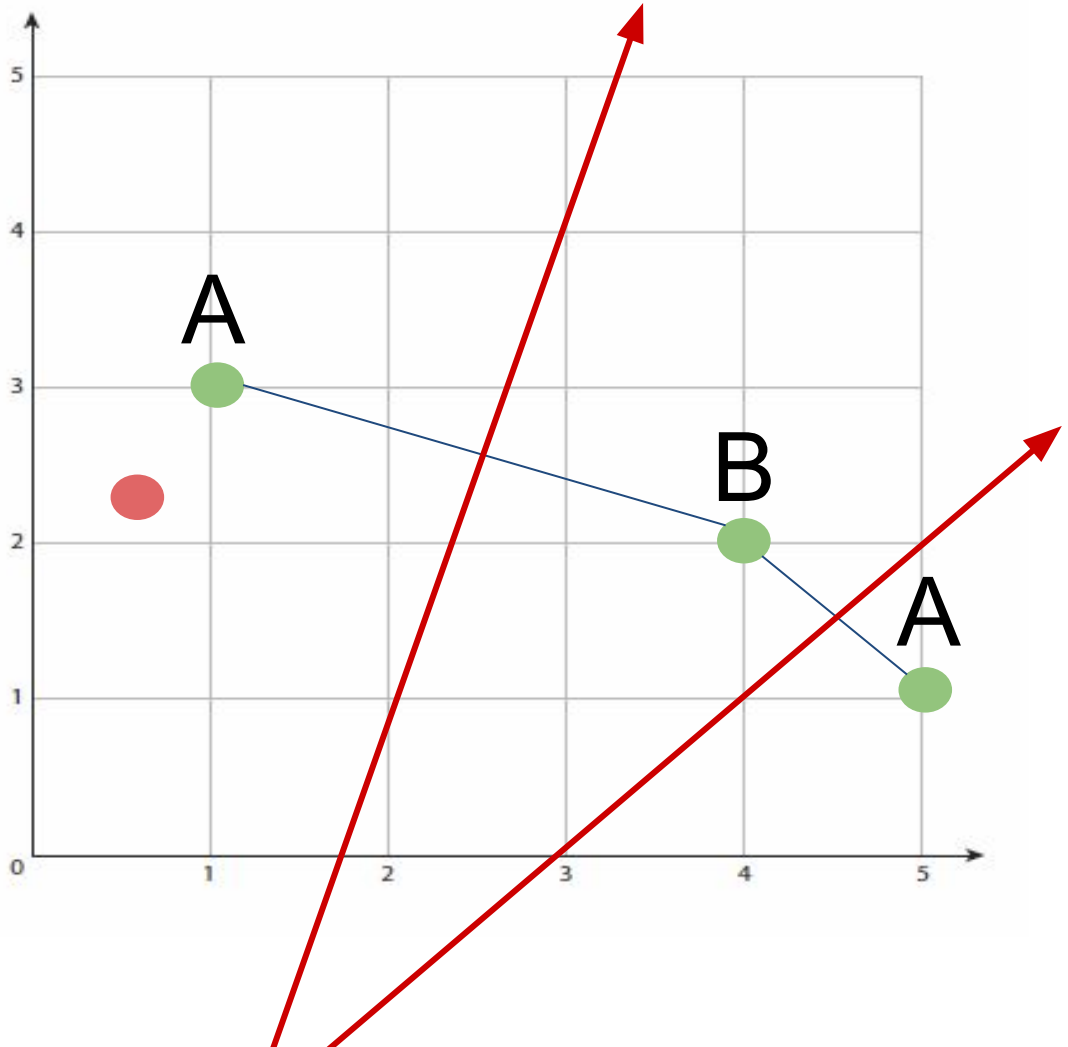
STEP 3: Extend the perpendicular bisectors



How a computer classifies using kNNs

It computes the region belonging to each type

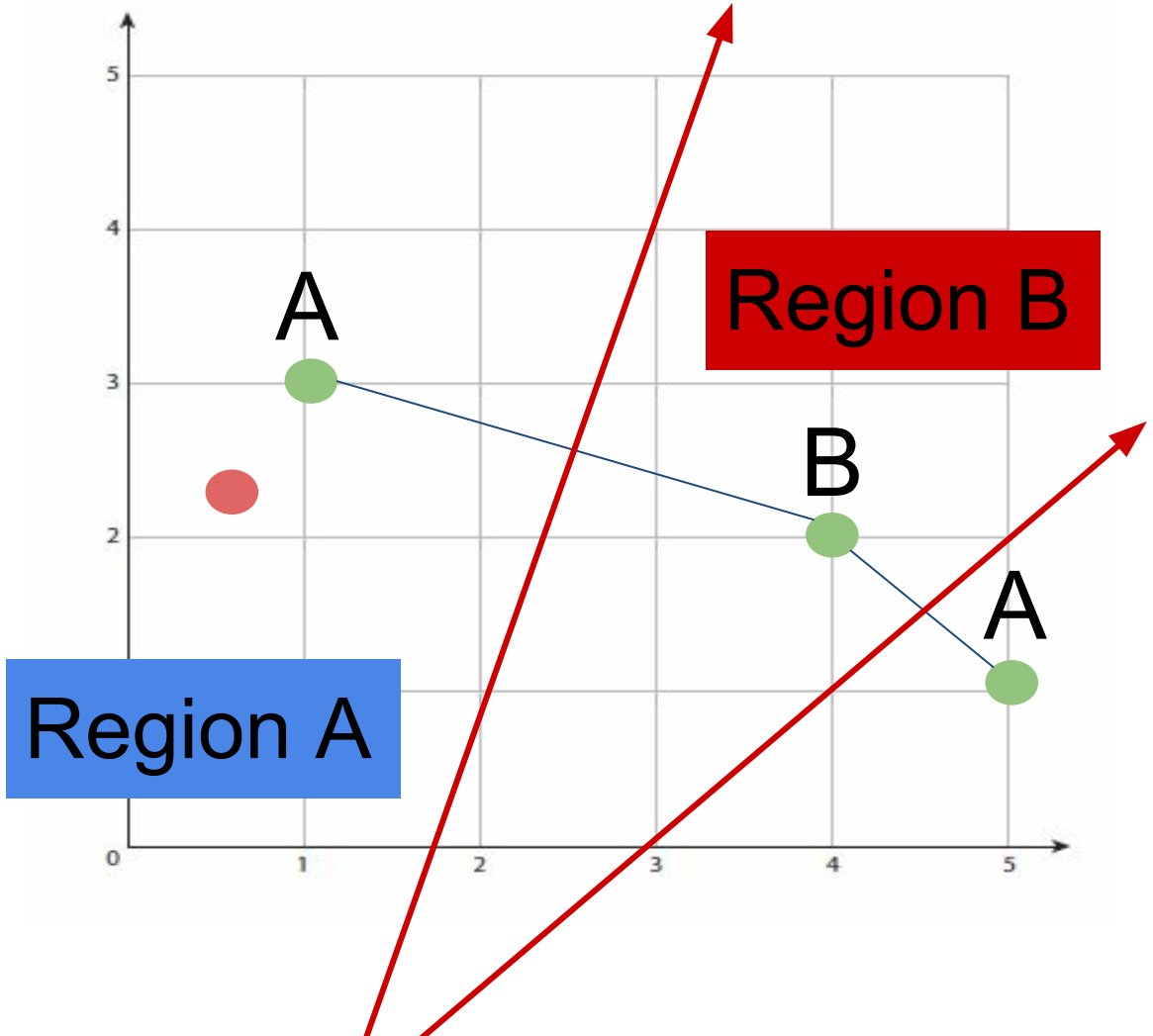
STEP 3: Extend the perpendicular bisectors



How a computer classifies using kNNs

Region A is now clear from Region B

The red point is **type A**



2. Identification Trees

- A simple representation for classifying items into classes
- Also called a decision tree

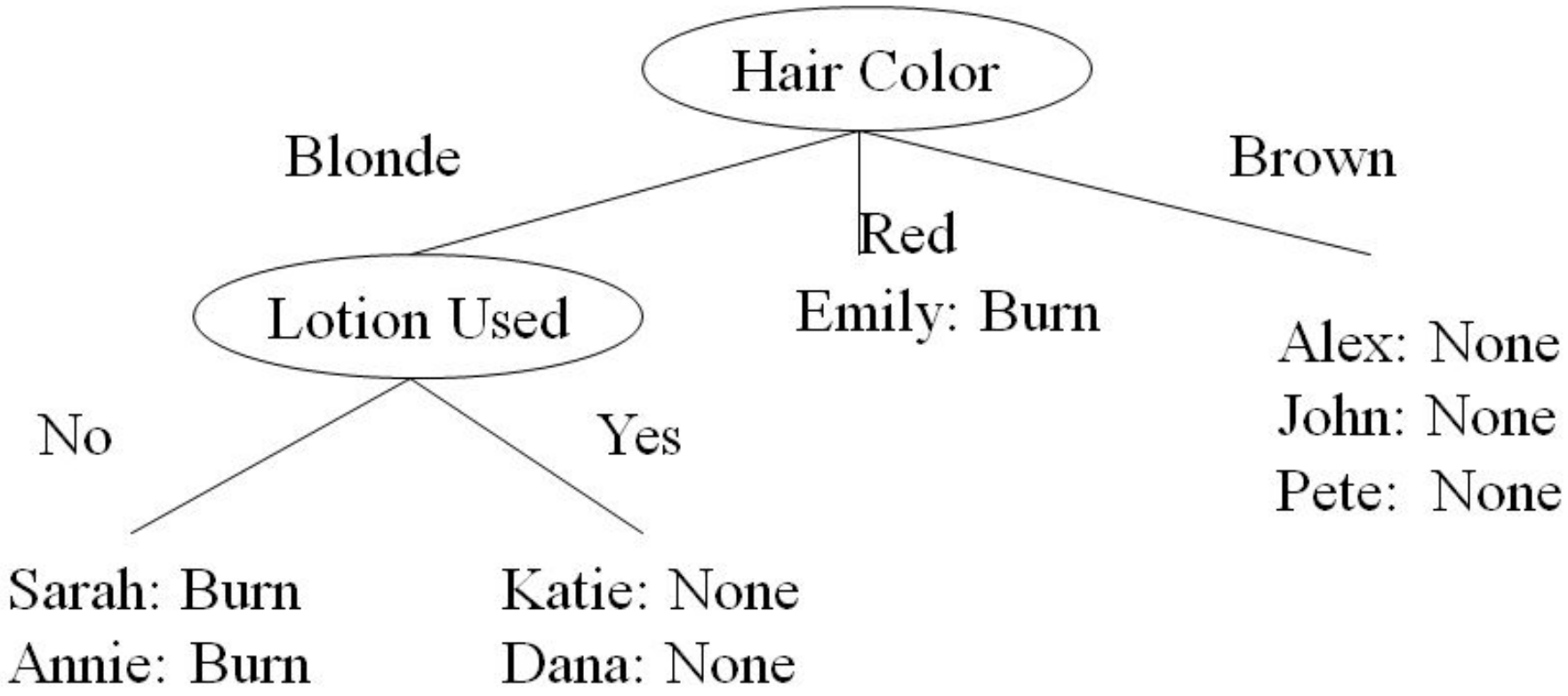
GOAL: Find a set of rules for classifying a new item

Identification Trees: Sunburn Example

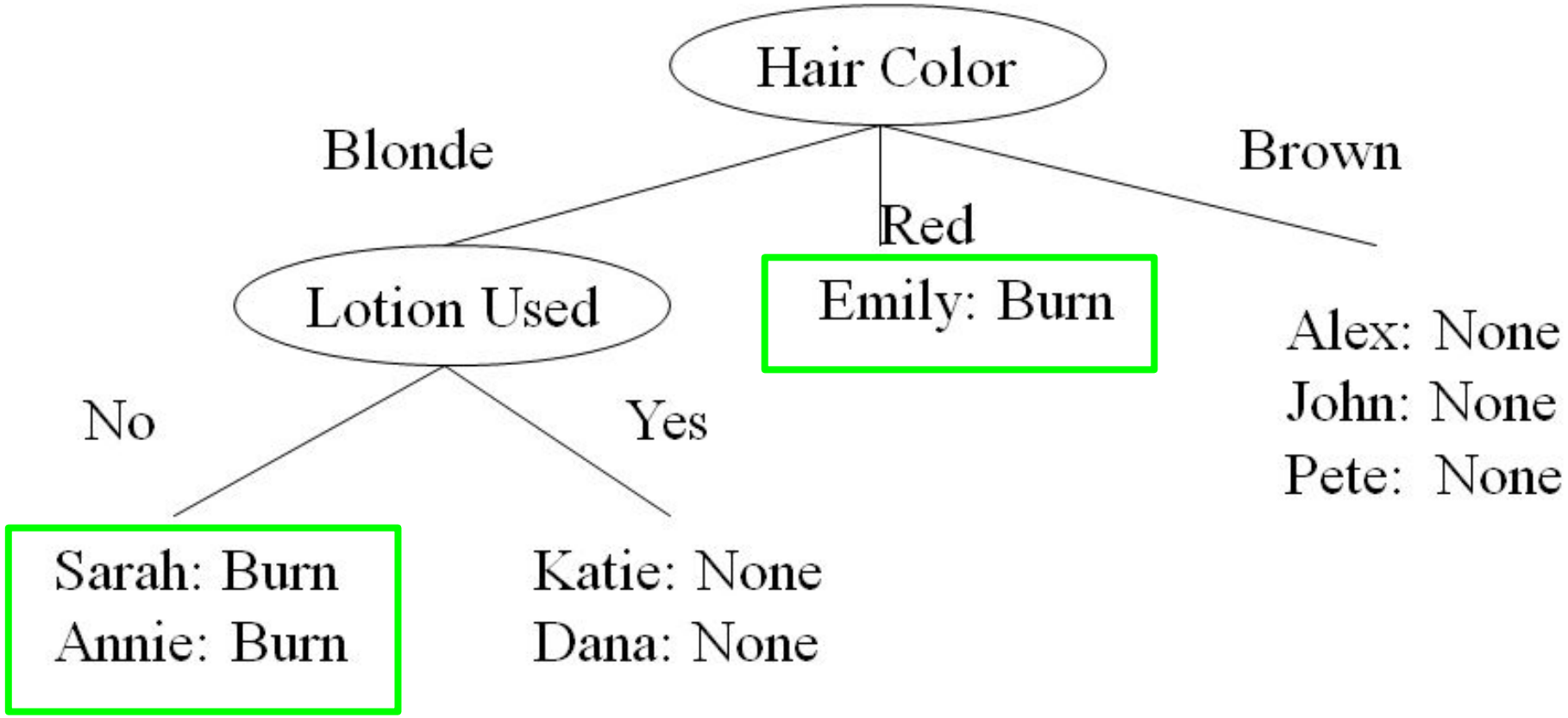
Name	Hair	Height	Weight	Lotion	Result
Sarah	Blonde	Average	Light	No	Burn
Dana	Blonde	Tall	Average	Yes	None
Alex	Brown	Short	Average	Yes	None
Annie	Blonde	Short	Average	No	Burn
Emily	Red	Average	Heavy	No	Burn
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Katie	Blonde	Short	Light	Yes	None

QUESTION: What features make you get a sunburn?

Sunburn Identification Tree



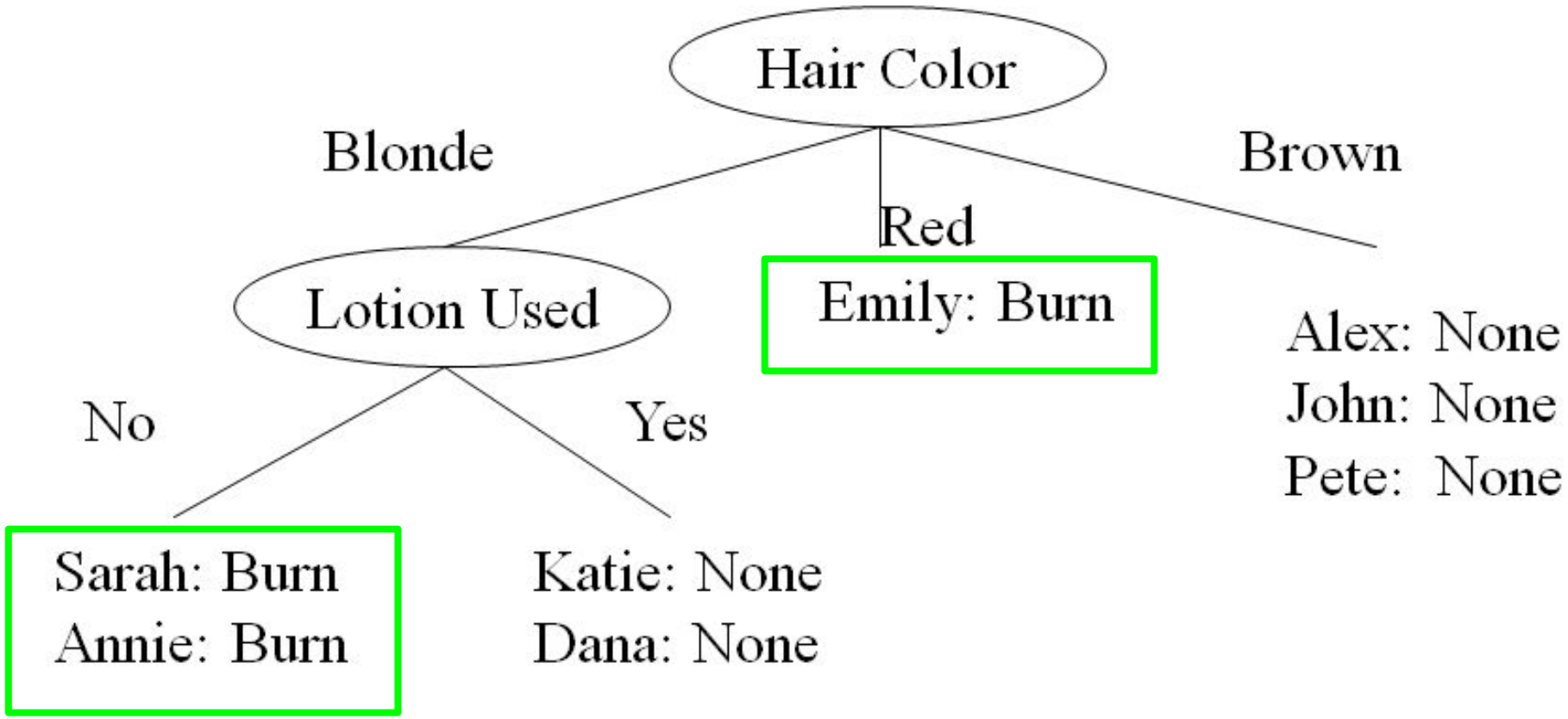
Sunburn Identification Tree



RULES: **1.** If red hair -> sunburn

2. If blonde hair and lotion not used -> sunburn

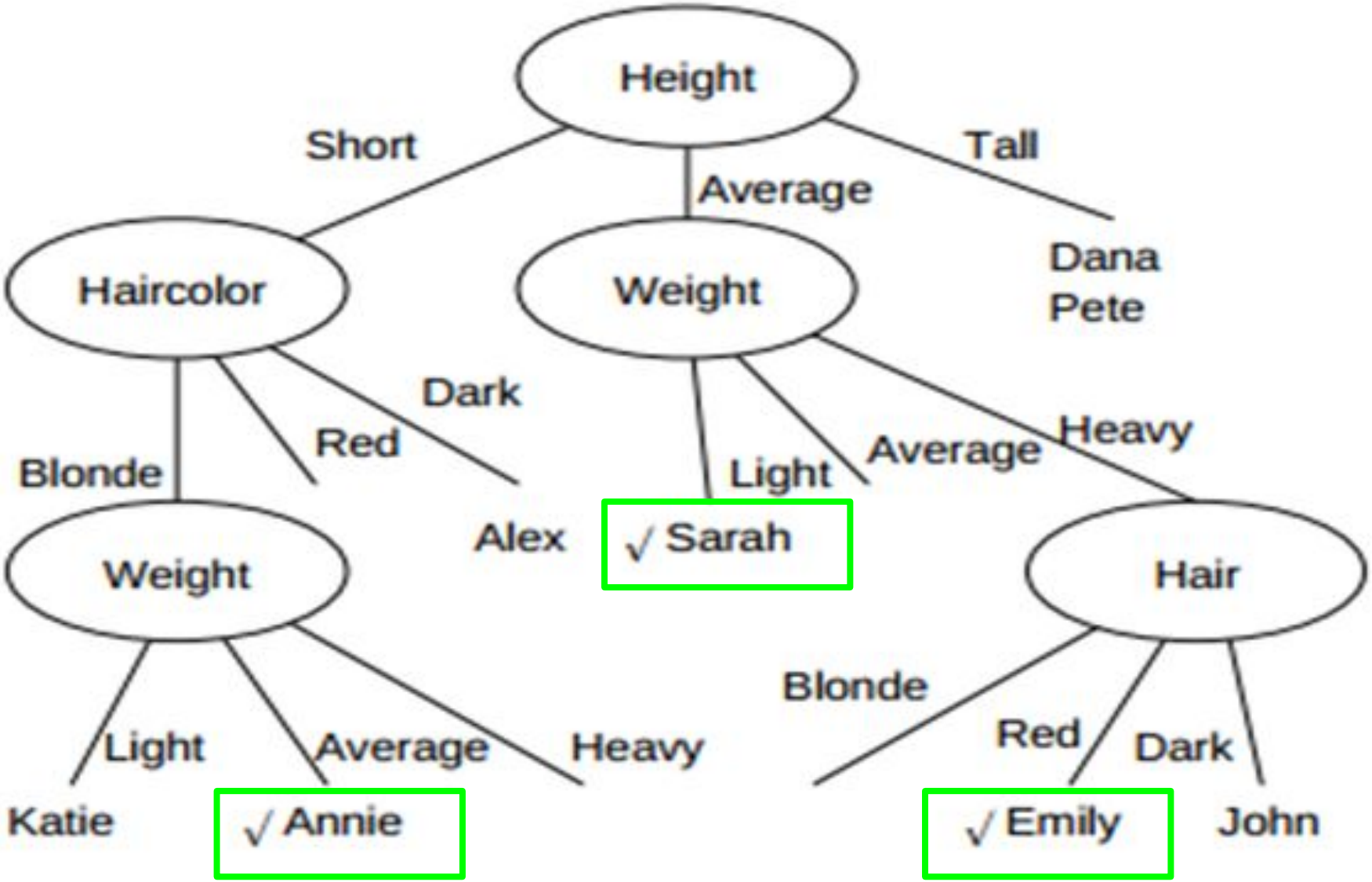
Why did the ID tree use Hair Color as the root feature?



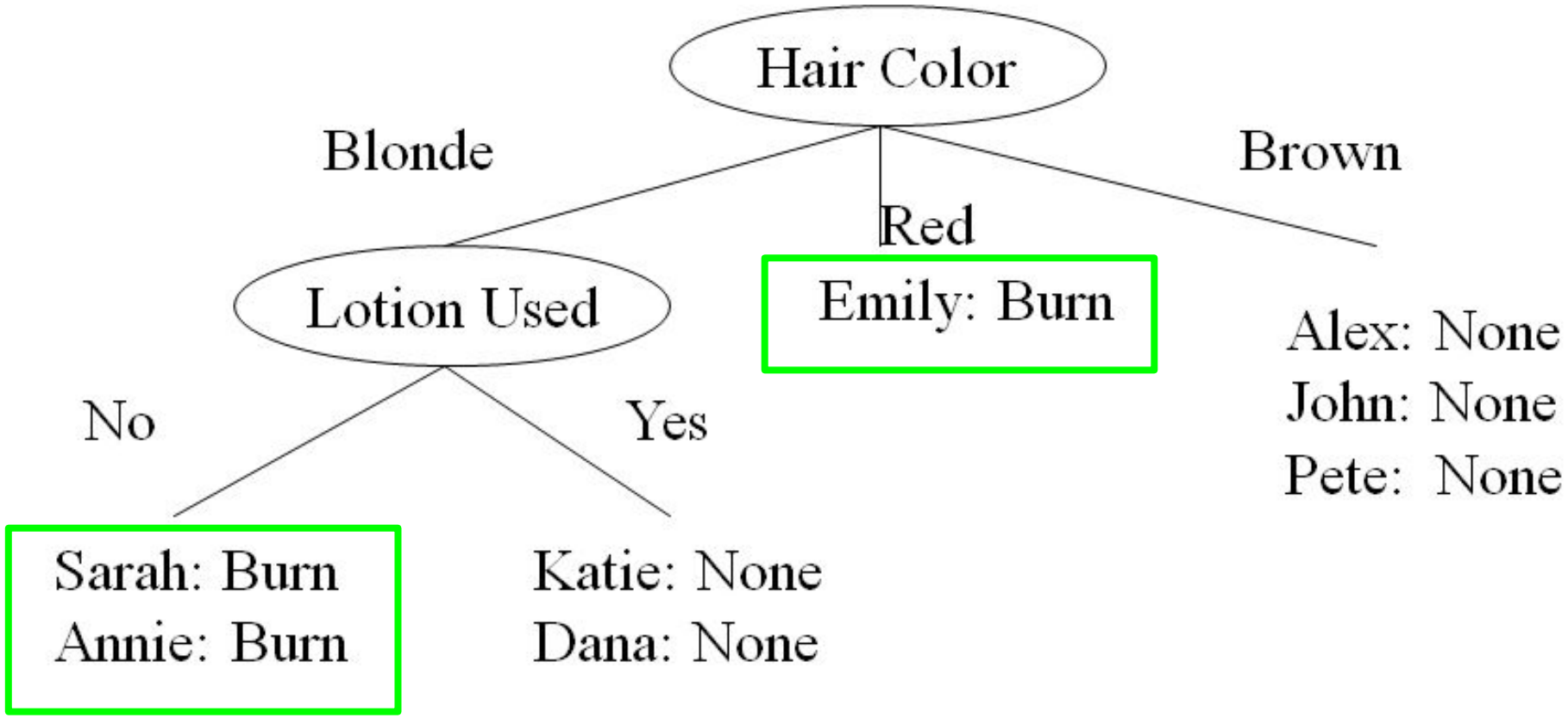
Why did the ID tree use Hair Color as the root feature?

Name	Hair	Height	Weight	Lotion	Result
Sarah	Blonde	Average	Light	No	Burn
Dana	Blonde	Tall	Average	Yes	None
Alex	Brown	Short	Average	Yes	None
Annie	Blonde	Short	Average	No	Burn
Emily	Red	Average	Heavy	No	Burn
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Katie	Blonde	Short	Light	Yes	None

Height as root feature



Hair Color as root feature

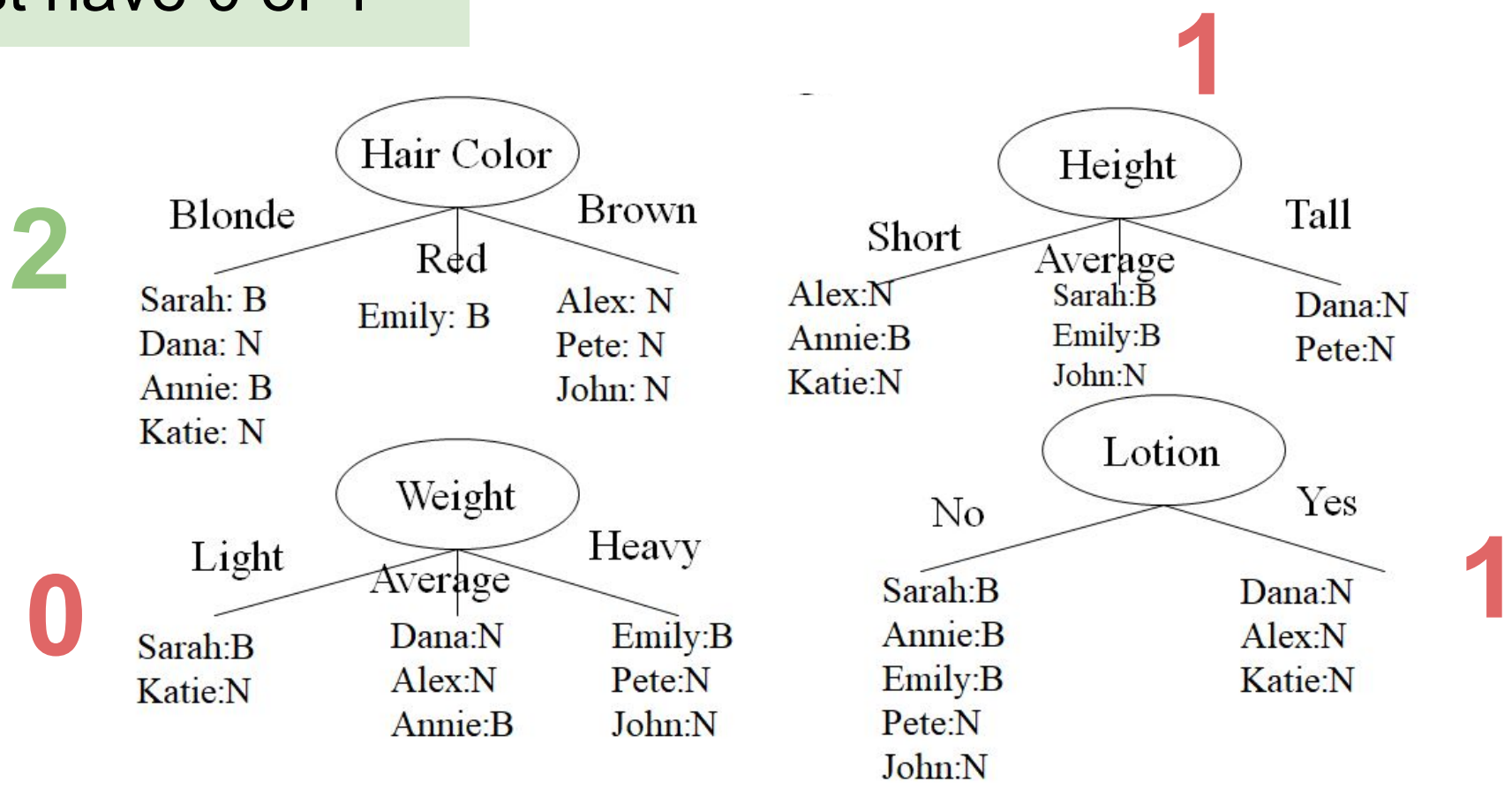


Gives us the smallest tree!

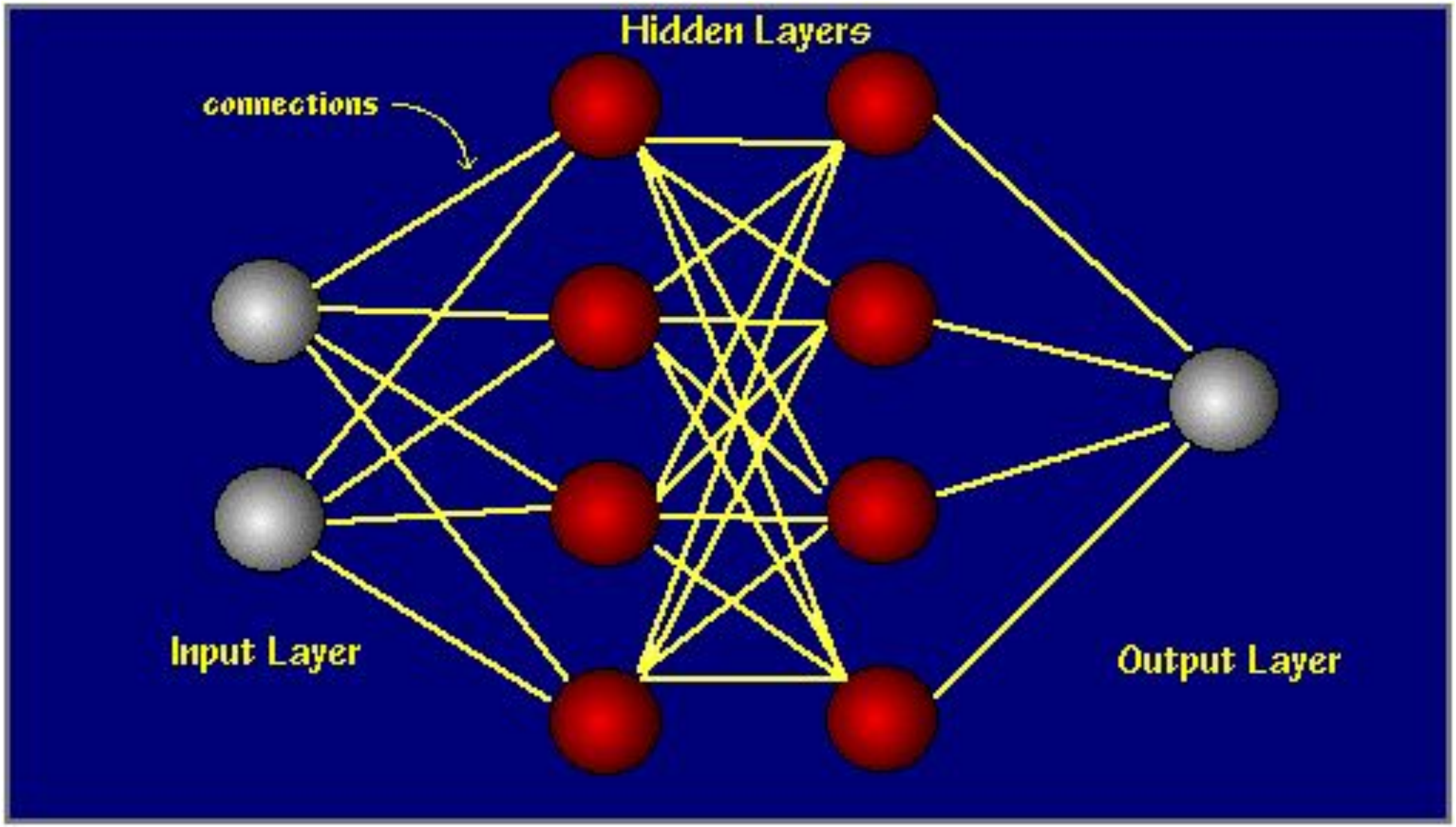
But why does starting with Hair Color give us the smallest tree?

Because Hair Color has least disorder

Has 2 homogenous branches while the rest have 0 or 1



3. Neural Networks



Neural Networks enable classification

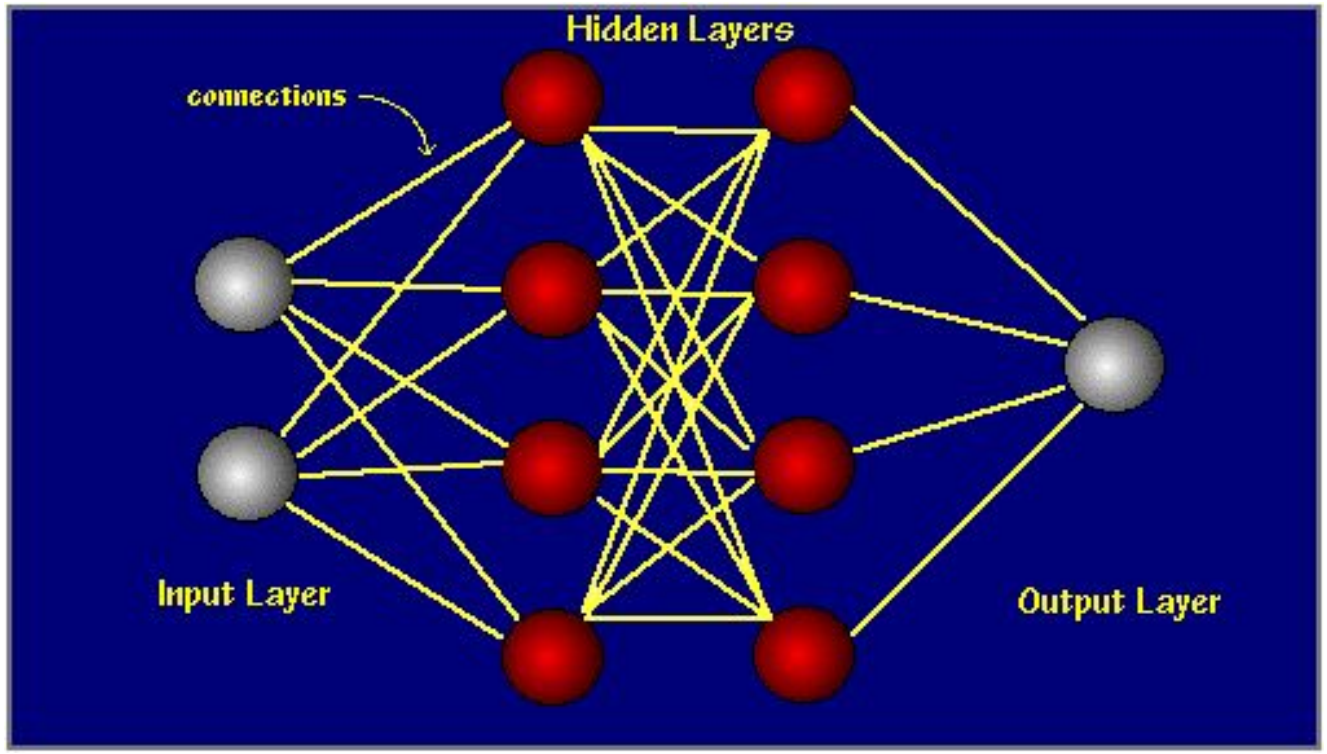
Input

Neural network hidden layers

CAT

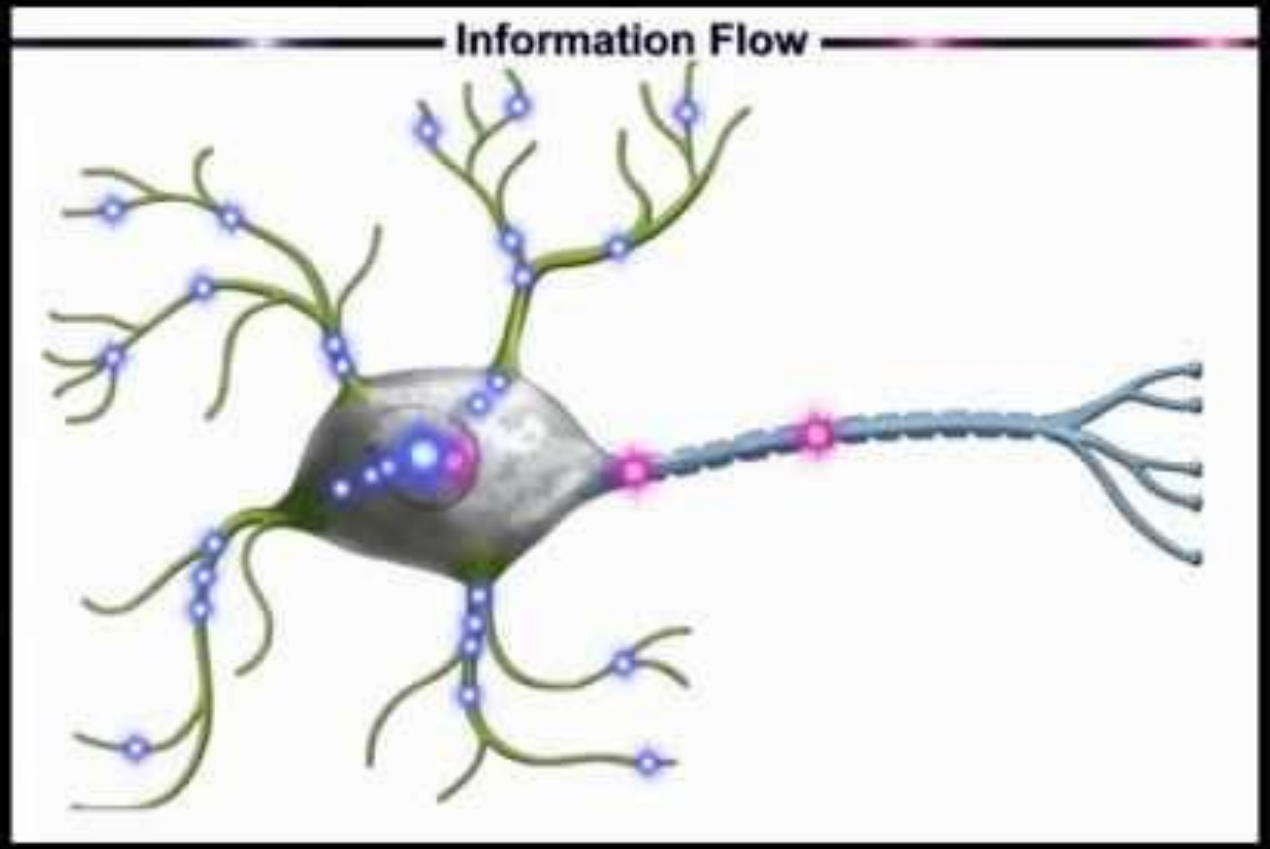
Output

What is a Neural Network?

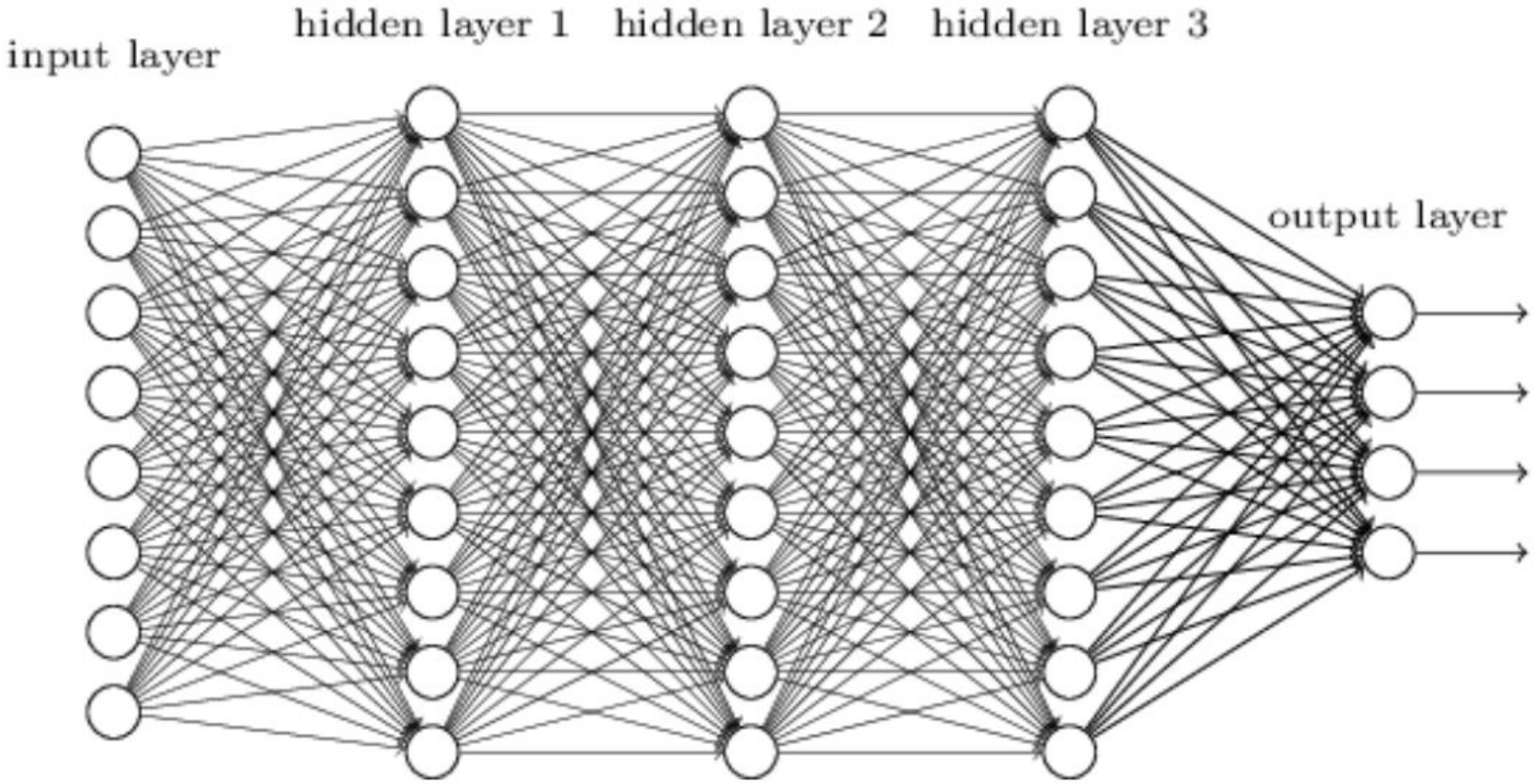


A Neural Net is a series of algorithms that attempts to identify underlying relationships in a data set by using a process that mimics the way the human brain operates.

What is a Neural Network?



A Neural Net is basically interconnected layers of nodes



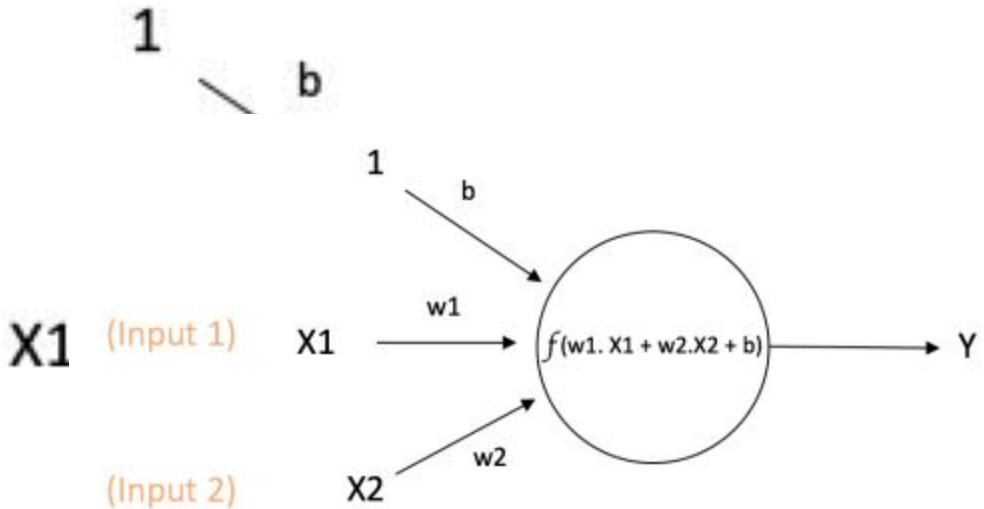
A Neural Net Example

(Input 1)

Height

(Input 2)

Weight



(Output)

Boy

or

Girl

Output of neuron = Y= $f(w1.\ X1 + w2.X2 + b)$

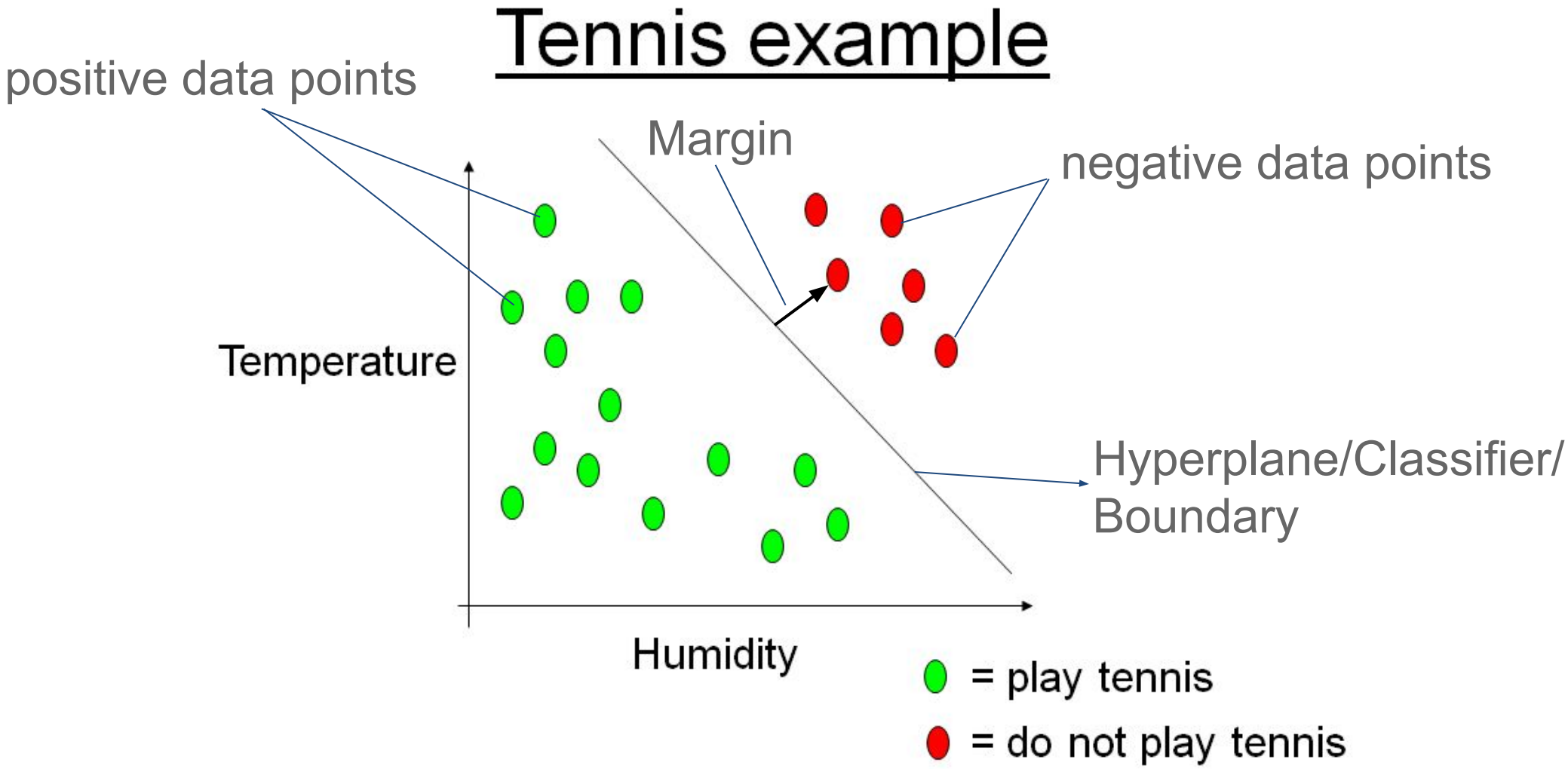
Output of neuron = Y= $f(w1.\ X1 + w2.X2 + b)$

4. Support Vector Machines (SVMs)

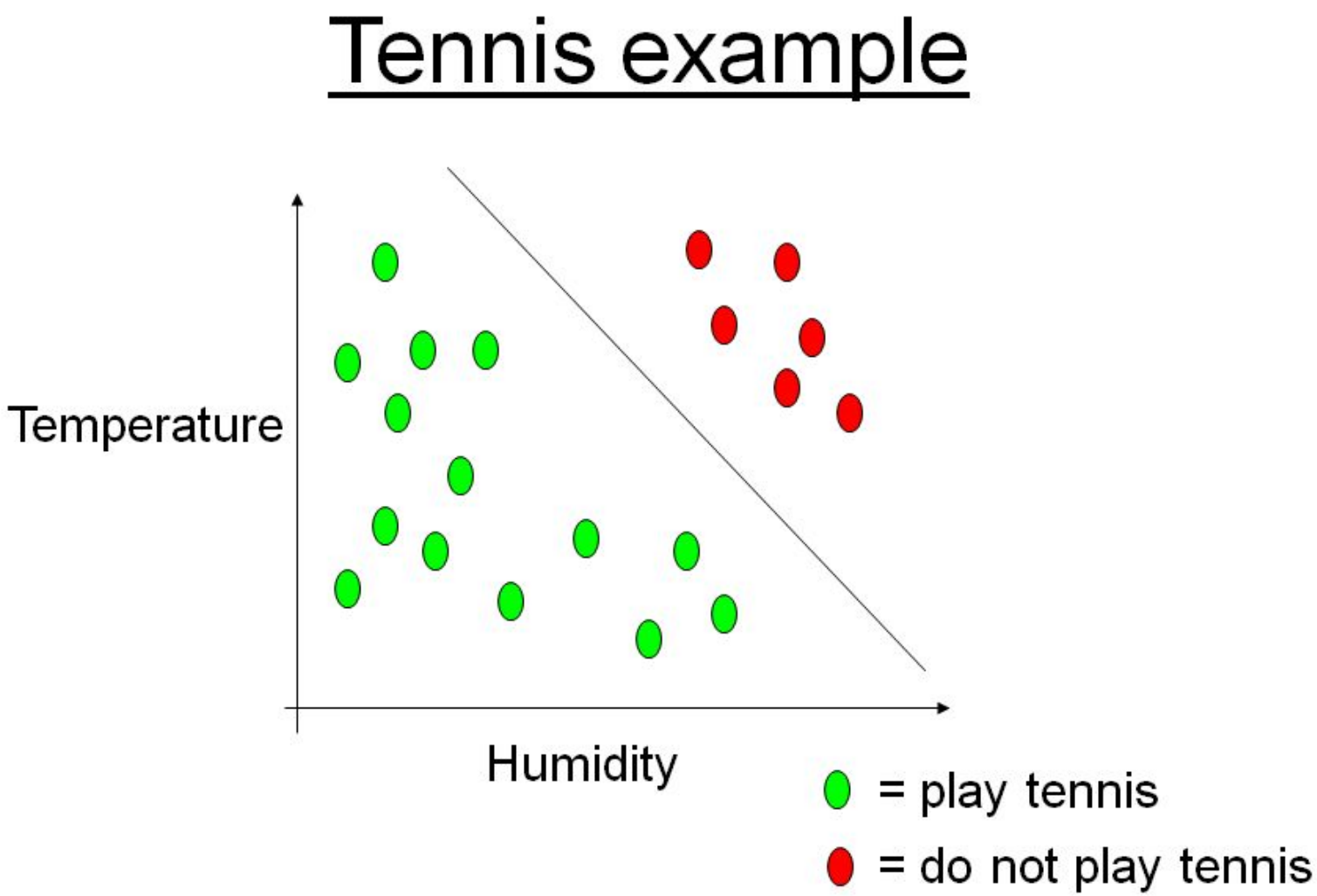
SVM is a supervised machine learning algorithm used to classify items using a computed hyperplane

GOAL: Find the most optimal boundary/classifier

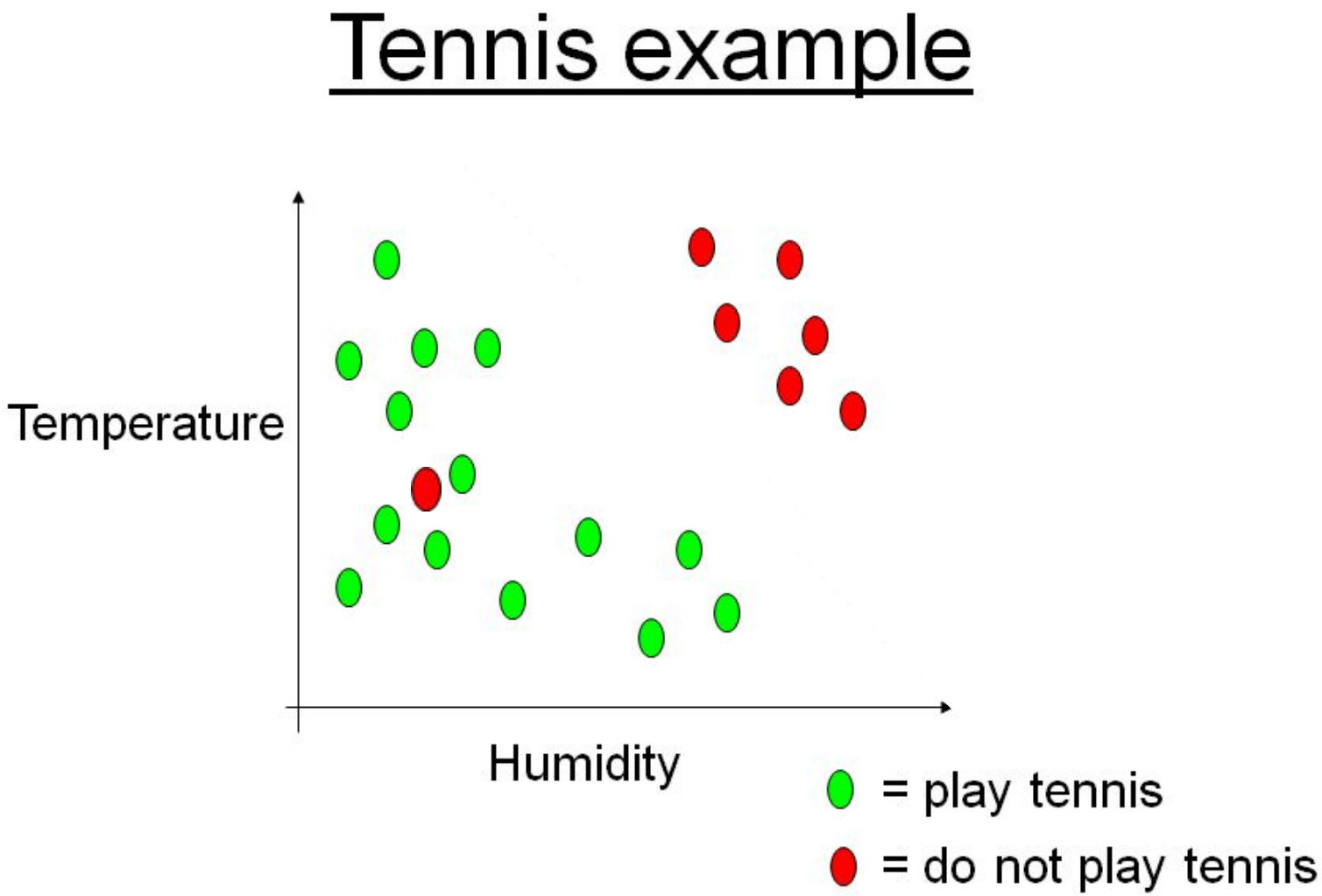
SVMs: Example



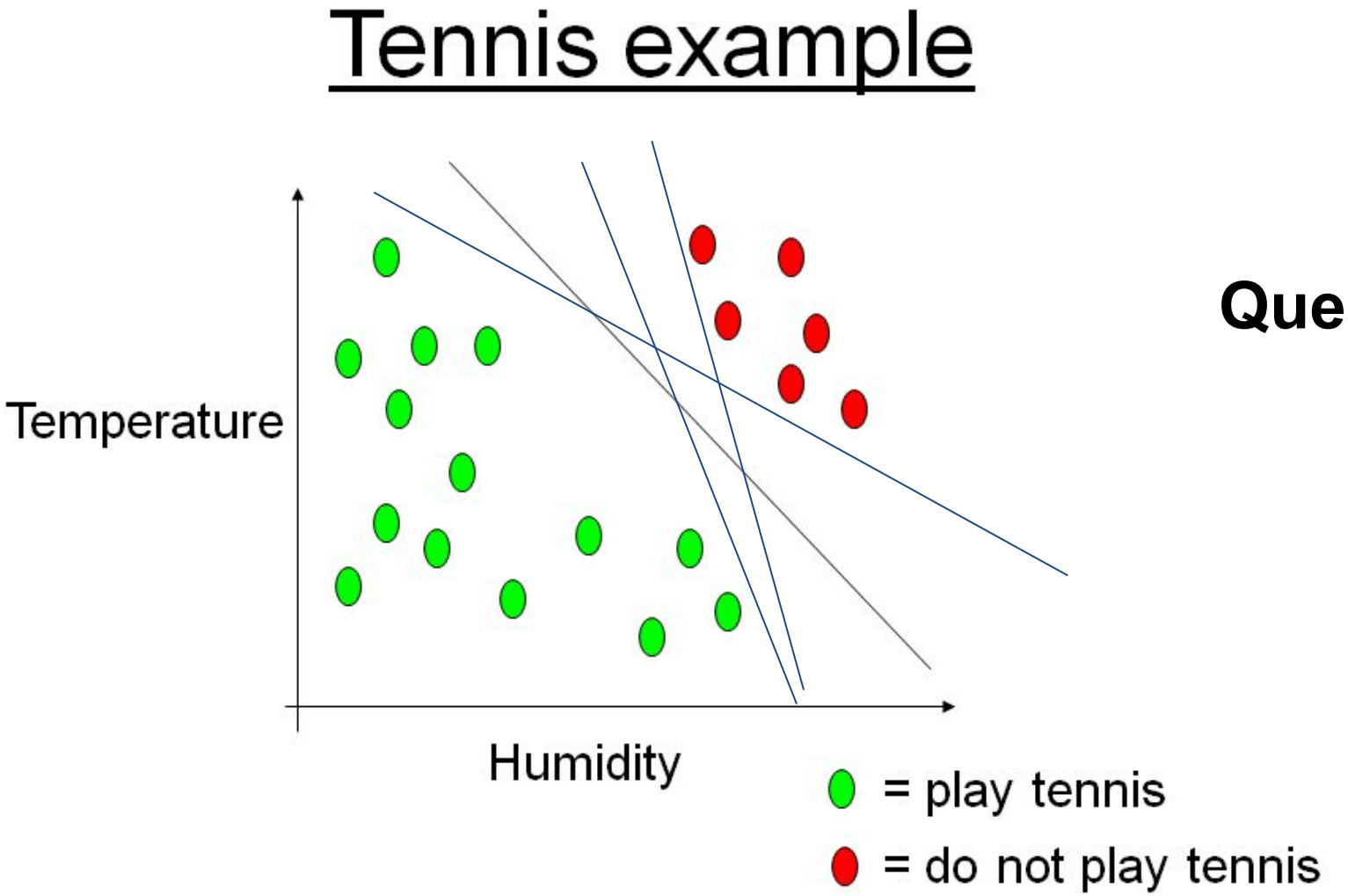
Linearly Separable Dataset



Non-Linearly Separable Dataset

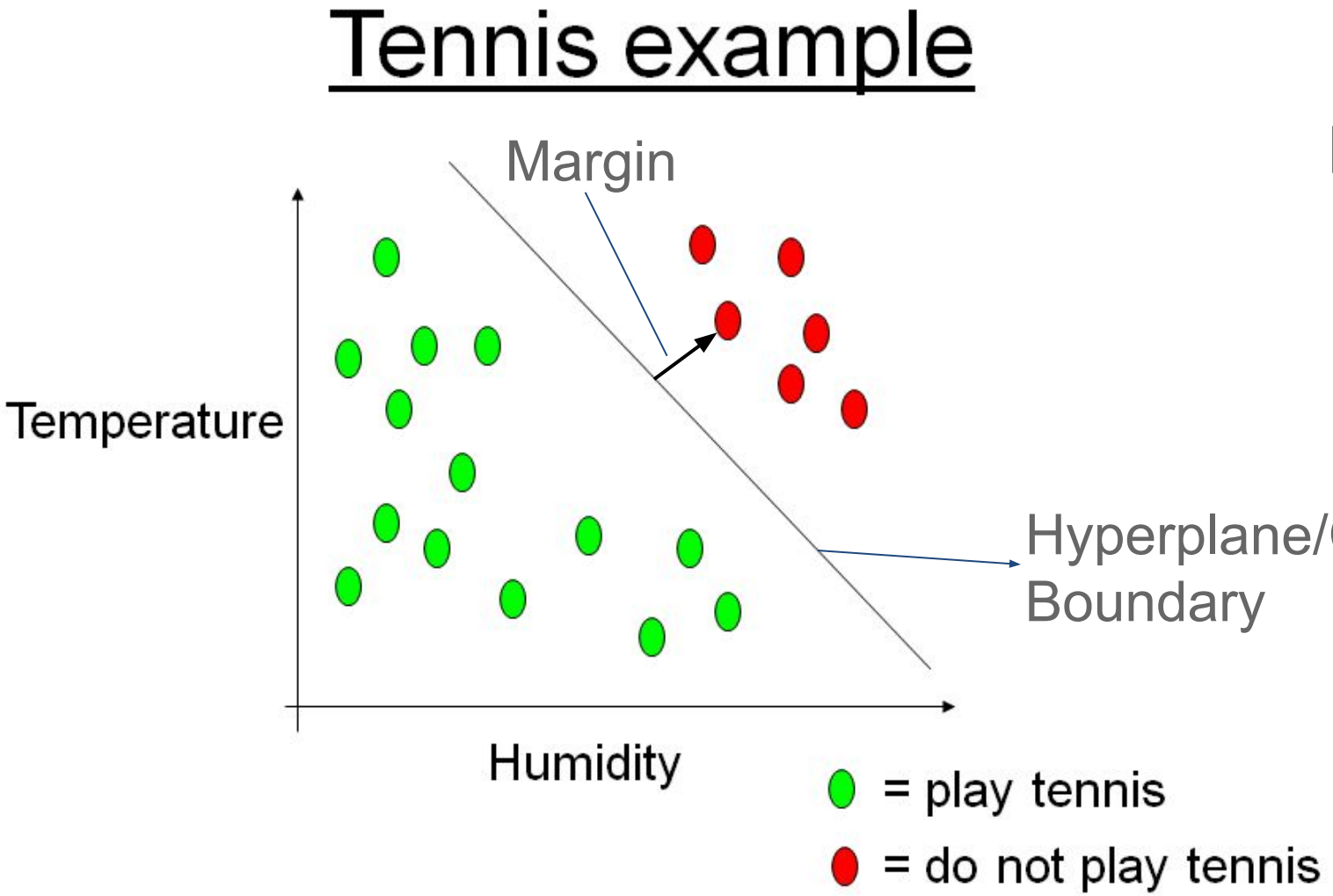


Many possible classifiers



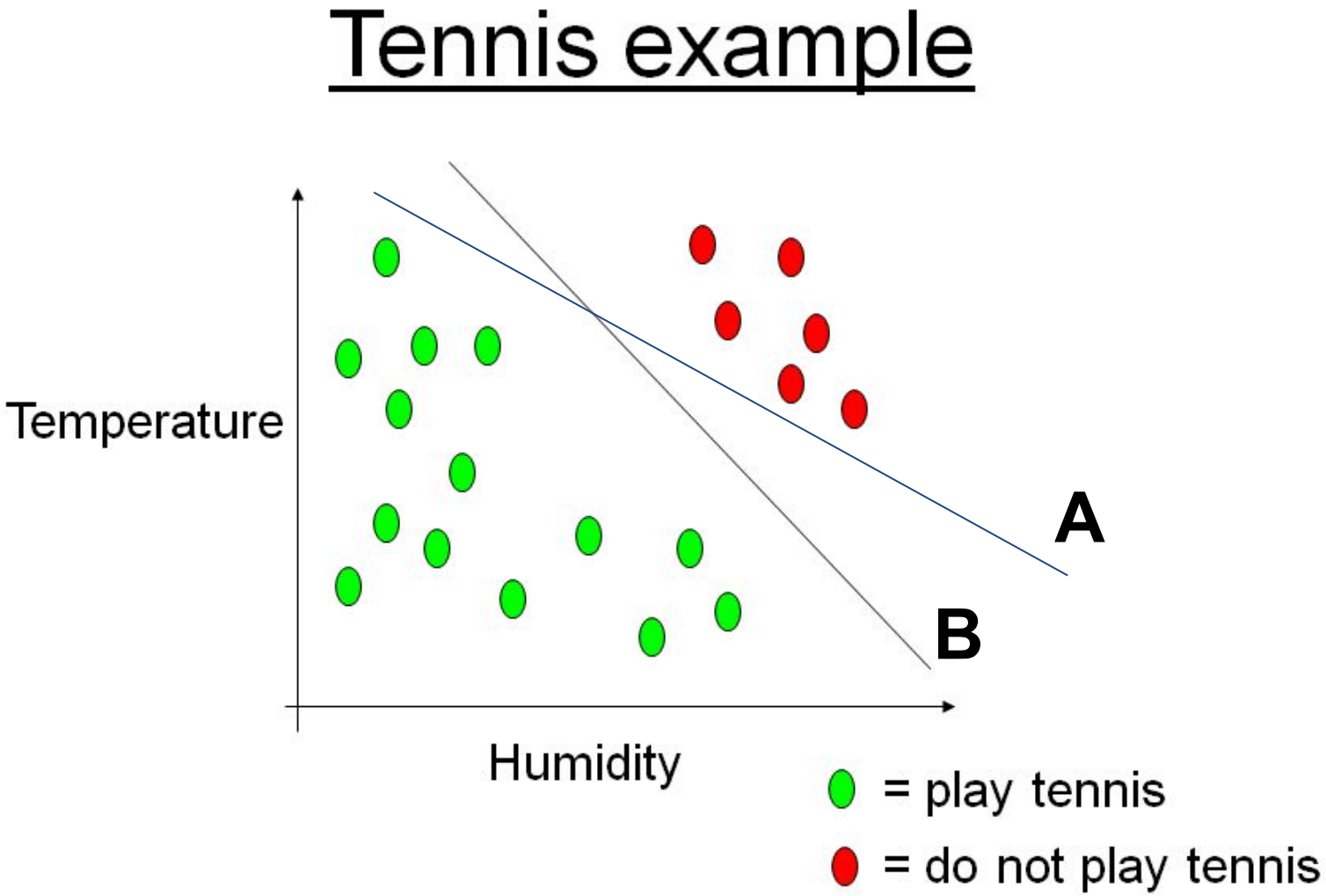
Question: Which is the best classifier? Why?

SVM Goal: Maximize Margin



Margin: Smallest distance from the classifier to the closest data point

Which is a better classifier?

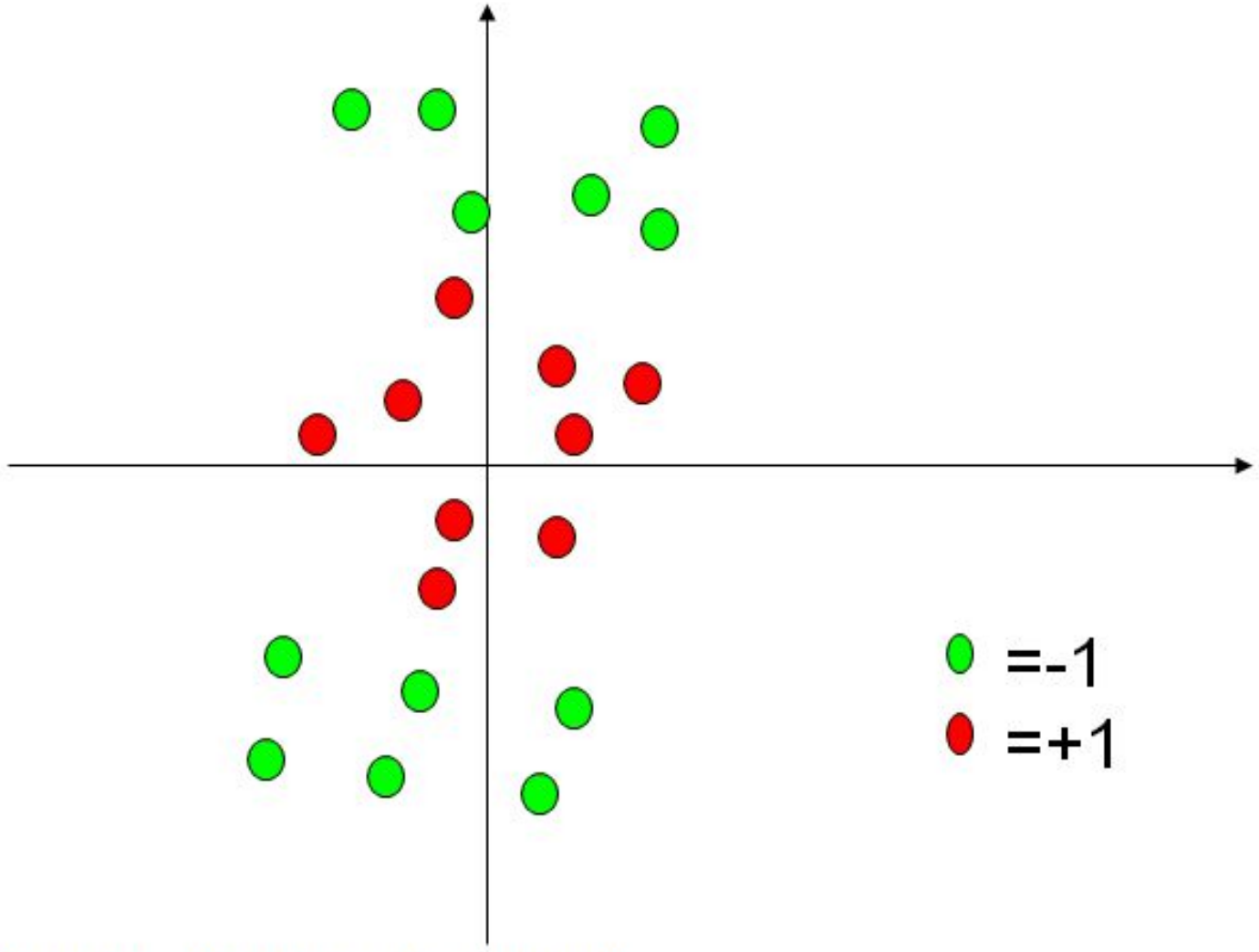


A OR B?

B because has a larger margin

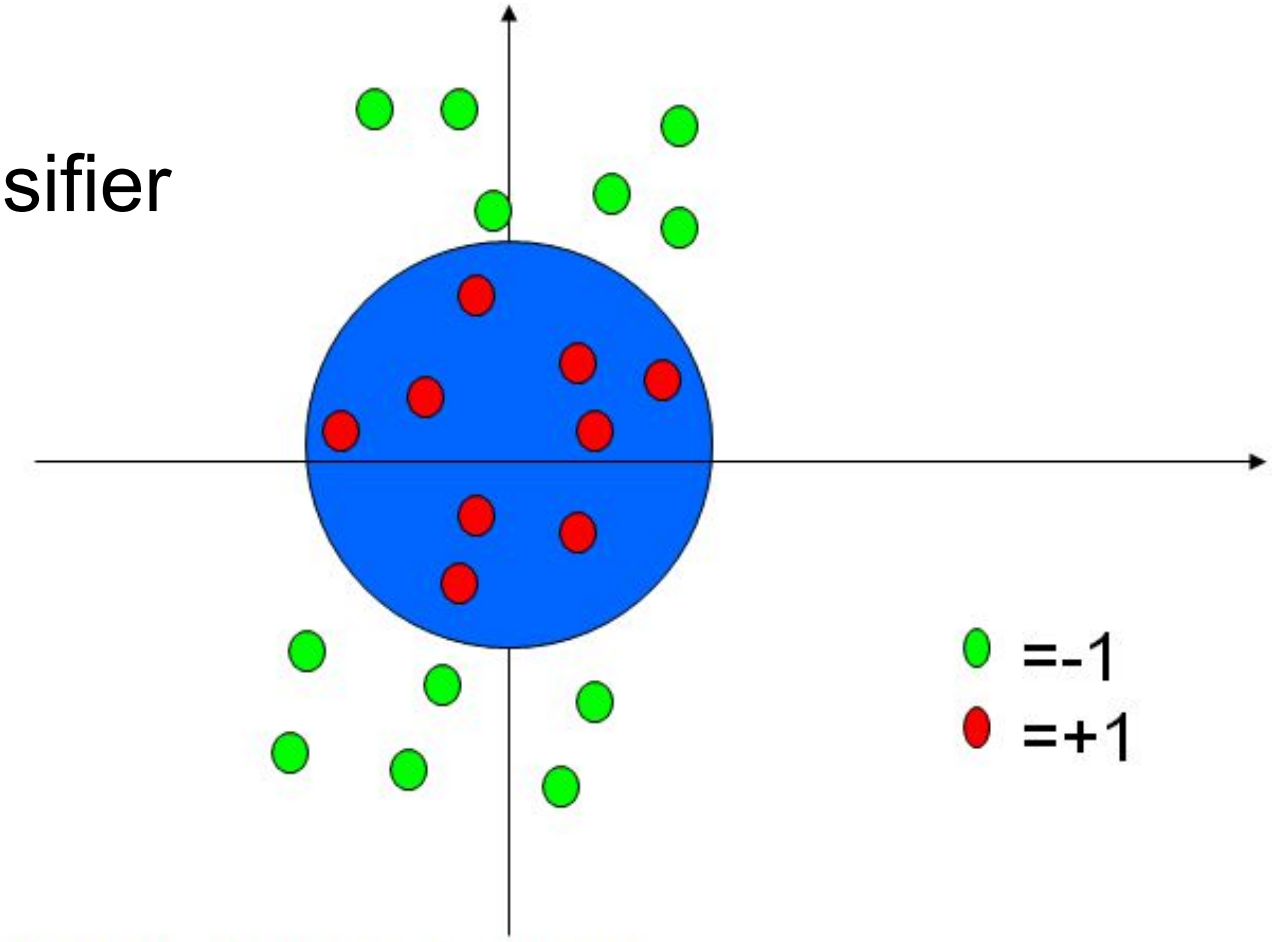
Because has a lot of room for new data points without needing many changes

Problem with Linear SVMs

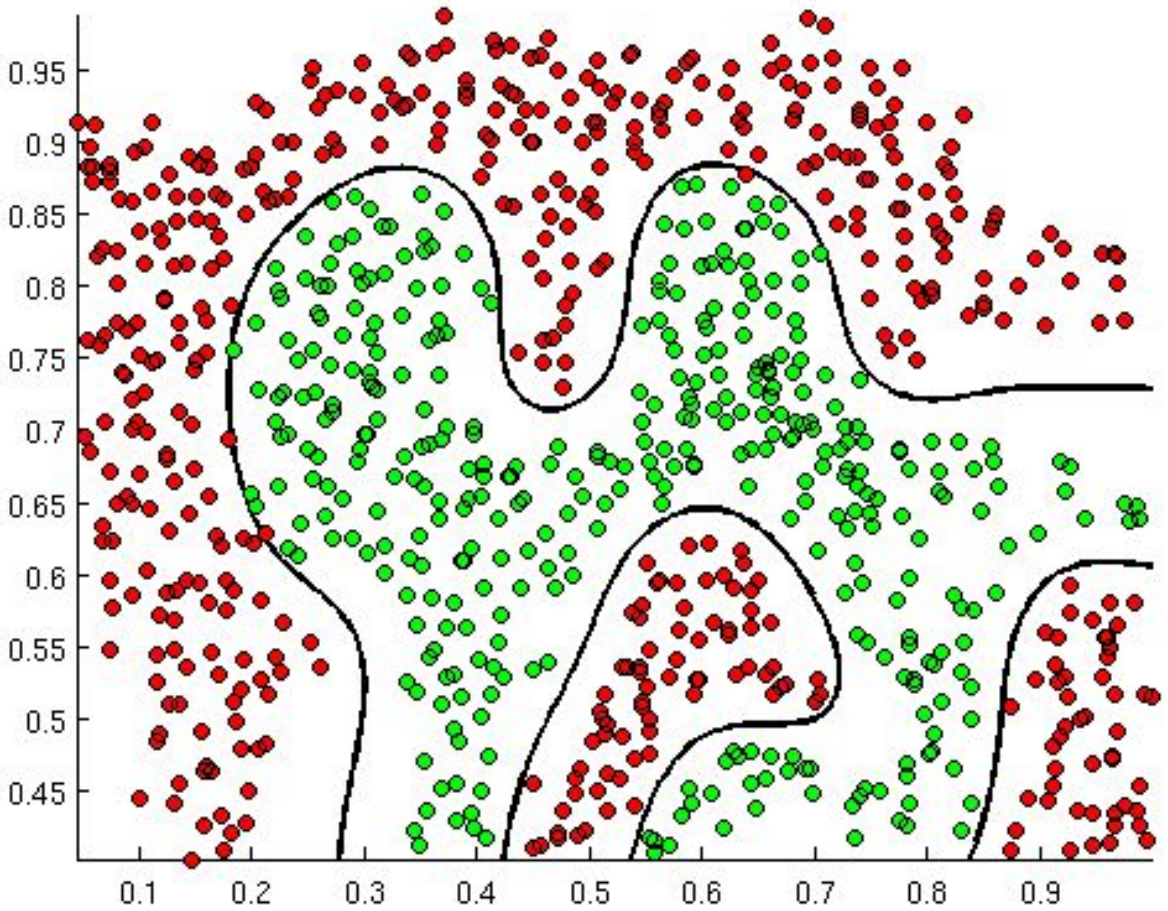


Non-linear SVMs make them powerful

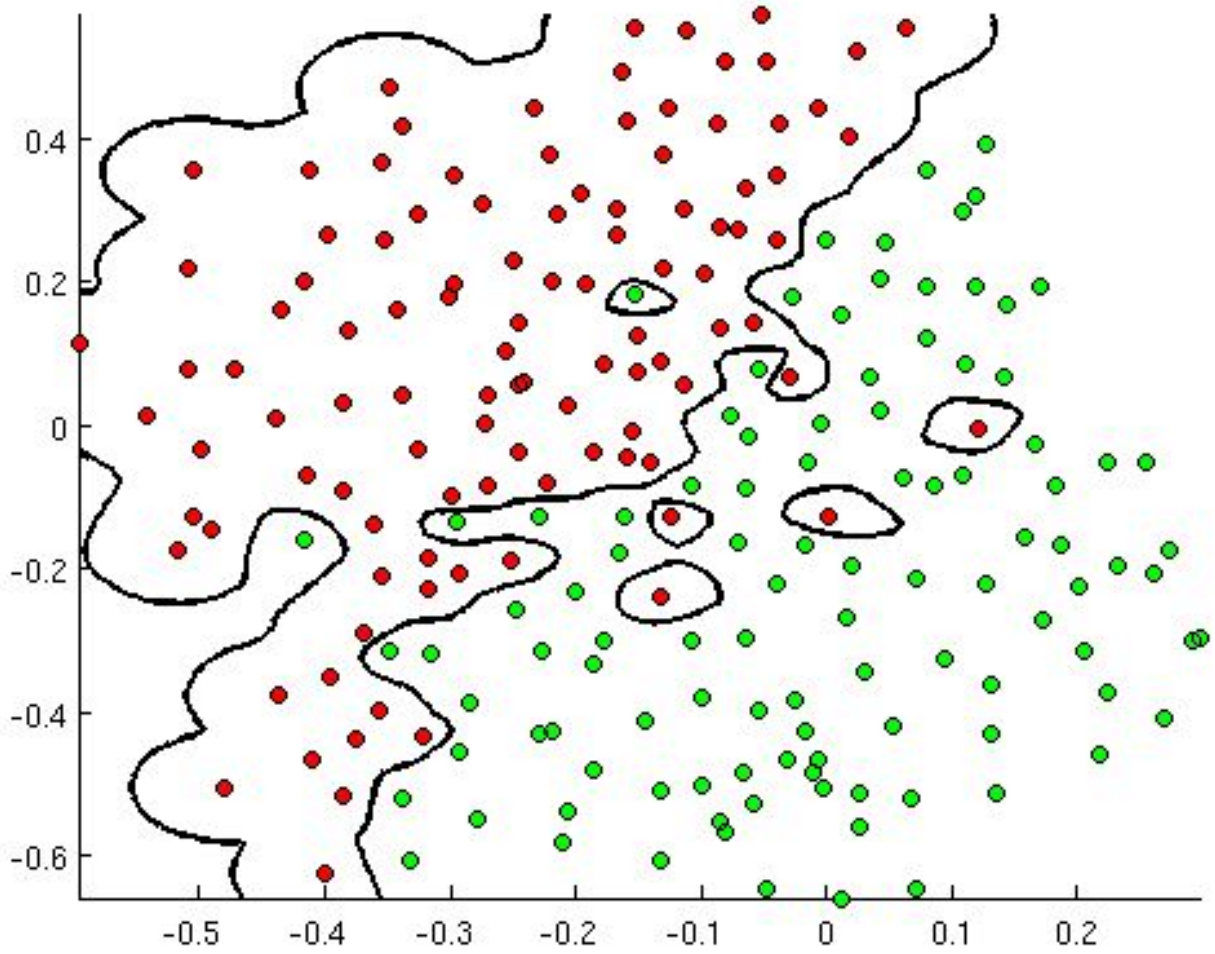
Circular classifier



SVMs can be irregular in shape



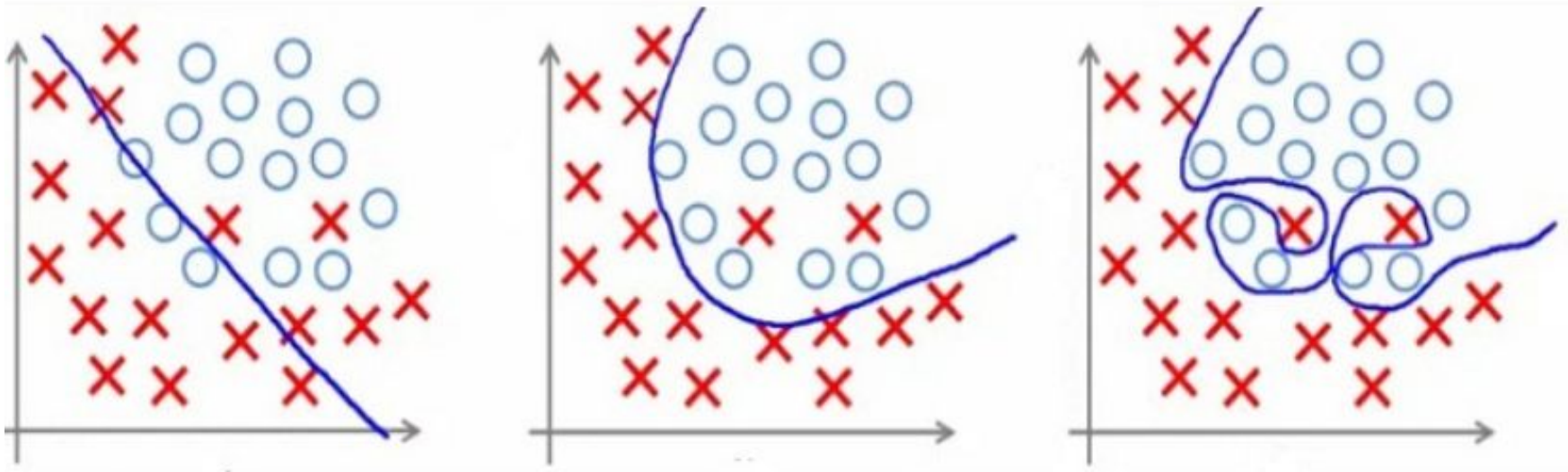
SVMs can have multiple disconnected boundaries



GOAL: Get best possible classifier

A lot of math involved!

Over & Under fitting are huge challenges for SVMs



Under-fitting

(too simple to explain the variance)

Appropriate-fitting

Over-fitting

(forcefitting -- too good to be true)

GOAL:

Be as general as possible

Team Project Time!