DATA SCIENCE

By: Michel Samir Zaki

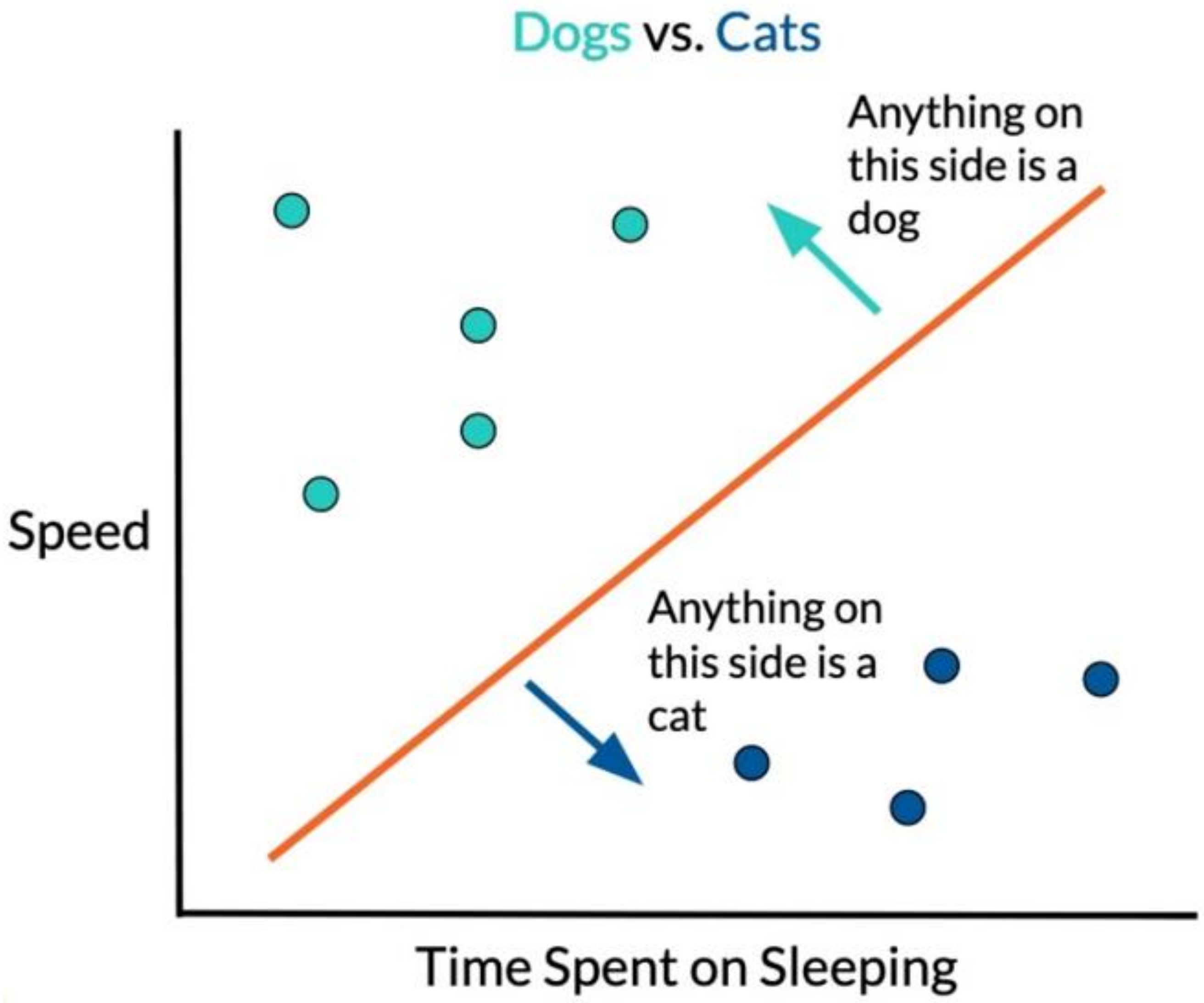
Perceptron

- Designed to mimic the neurons in the human brain

- A single perceptron is a **linear classifier** – it separates two groups using a line.

- The perceptron learns by readjusting itself based on points it **misclassified** (points on the wrong side of the line) at every time step.

Perceptron



Dogs vs. Cats
Anything on this side is a dog
Anything on this side is a cat
Speed
Time Spent on Sleeping

Perceptron Learning Algorithm (PLA)

- Distance from a point to the line

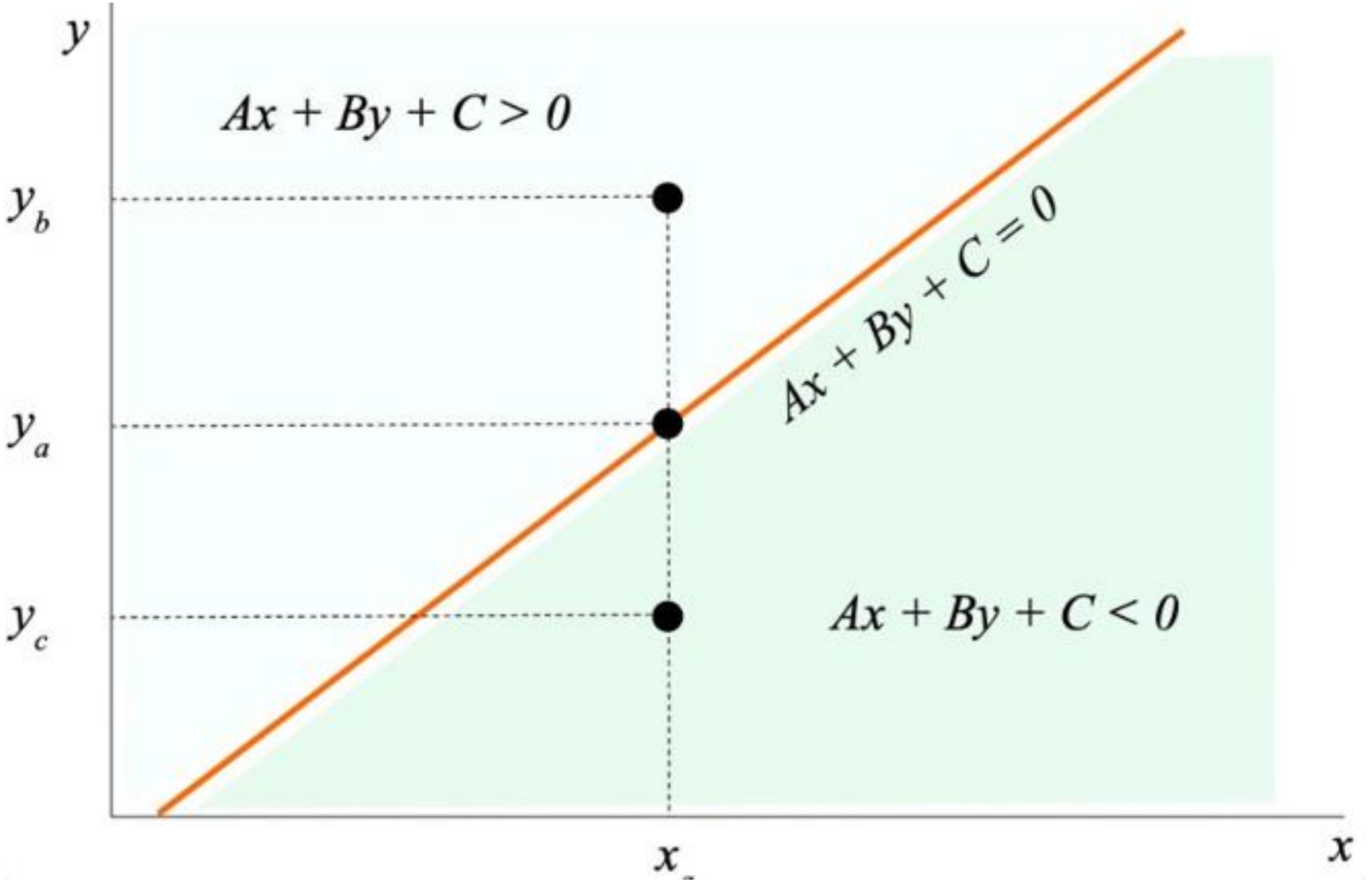
To write an algorithm that finds a line that separates the two groups, we must first find a way to figure out whether a given point is above or below the line.

$Ax + By + C = 0$ is

$$d = \frac{Ax_a + By_a + C}{\sqrt{A^2 + B^2}}.$$

Perceptron Learning Algorithm (PLA)

- Distance from a point to the line



Perceptron Learning Algorithm (PLA)

- Misclassified points

In summary, we can make a prediction for a point (x_a, y_a) by computing

$$\operatorname{sgn}(Ax_a + By_b + C),$$

where the value of 1 denotes the point is part of one group and a value of -1 denotes the point is part of the other.

> The sgn function computes the sign of the input. It is defined as follows:
>
> $$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Given a training dataset – set of points (x, y, a) were x and y are coordinates and a is 1 or -1 depending on which group the point is part of – we can find misclassified points by finding points whose a value is different from the prediction.

Perceptron Learning Algorithm (PLA)

There are two fundamental assumptions about PLA:

1. The data is linearly separable. In other words, a hyperplane (a linear decision boundary) can completely separate the instances of the two classes in the feature space. This hyperplane is the line where the algorithm aims to find an optimal position during training.

2. The algorithm is designed for binary classification problems, where the goal is to separate data points into two classes. For instance, in a medical scenario, the algorithm could be used to predict whether a person has breast cancer or not, where the two classes are "positive" (having breast cancer) and "negative" (not having breast cancer).

Perceptron Learning Algorithm (PLA)

- The “learning” in perceptron learning

This algorithm will **update** the line Ax+By+C=0 to have new **coefficients (weights)** that ideally separate the two groups in the data.

Algorithm 1 2D Perceptron Learning Algorithm (sketch)

---

procedure LEARN($A, B, C, samples$)
    for each datapoint (x, y, a) in $samples$ **do** // a is 1 or -1 depending on which group datapoint belongs to
        if datapoint is misclassified **then**
            $A \leftarrow A + a \cdot x$
            $B \leftarrow B + a \cdot y$
            $C \leftarrow C + a \cdot 1$
        end if
    end for
end procedure

Perceptron Learning Algorithm (PLA)

- Learning Rate in perceptron learning

The learning rate, “r”, scales how much we correct our perceptron for each misclassified point.

Algorithm 1 2D Perceptron Learning Algorithm (with learning rate)

procedure LEARN($A, B, C, samples$)
    for each datapoint (x, y, a) in $samples$ **do** // a is 1 or -1 depending on which group datapoint belongs to
        if datapoint is misclassified **then**
            $A \leftarrow A + r \cdot a \cdot x$ // r is the learning rate
            $B \leftarrow B + r \cdot a \cdot y$
            $C \leftarrow C + r \cdot a \cdot 1$
        end if
    end for
end procedure

Perceptron Learning Algorithm (PLA)

- Perceptron in Vectors

So far all of the calculations have been done on 2-D space – each point has two coordinates x and y (in 3D space we add a z coordinate to it). But if we switch out the notation from

$$\begin{aligned} A &\leftarrow A + a \cdot x \\ B &\leftarrow B + a \cdot y \\ C &\leftarrow C + a \cdot 1 \end{aligned}$$

to

$$\begin{aligned} w_0 &\leftarrow w_0 + a \cdot 1 \\ w_1 &\leftarrow w_1 + a \cdot x_1 \\ w_2 &\leftarrow w_2 + a \cdot x_2 \end{aligned}$$

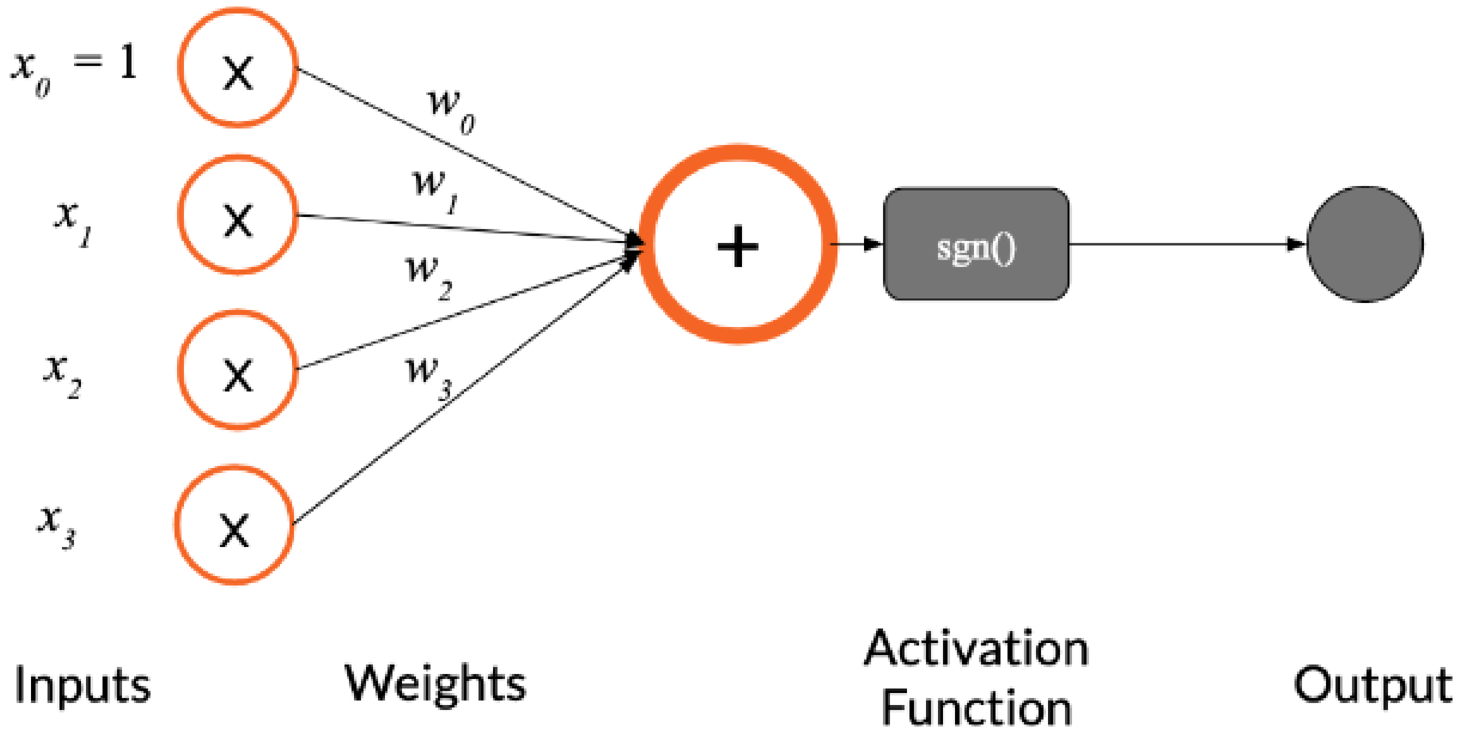
we can then see a clear pattern. Expressed as vectors, we can instead write these updates as

$$\vec{w} \leftarrow \vec{w} + a \cdot \vec{x}.$$

Perceptron Learning Algorithm (PLA)

- Anatomy of a Perceptron

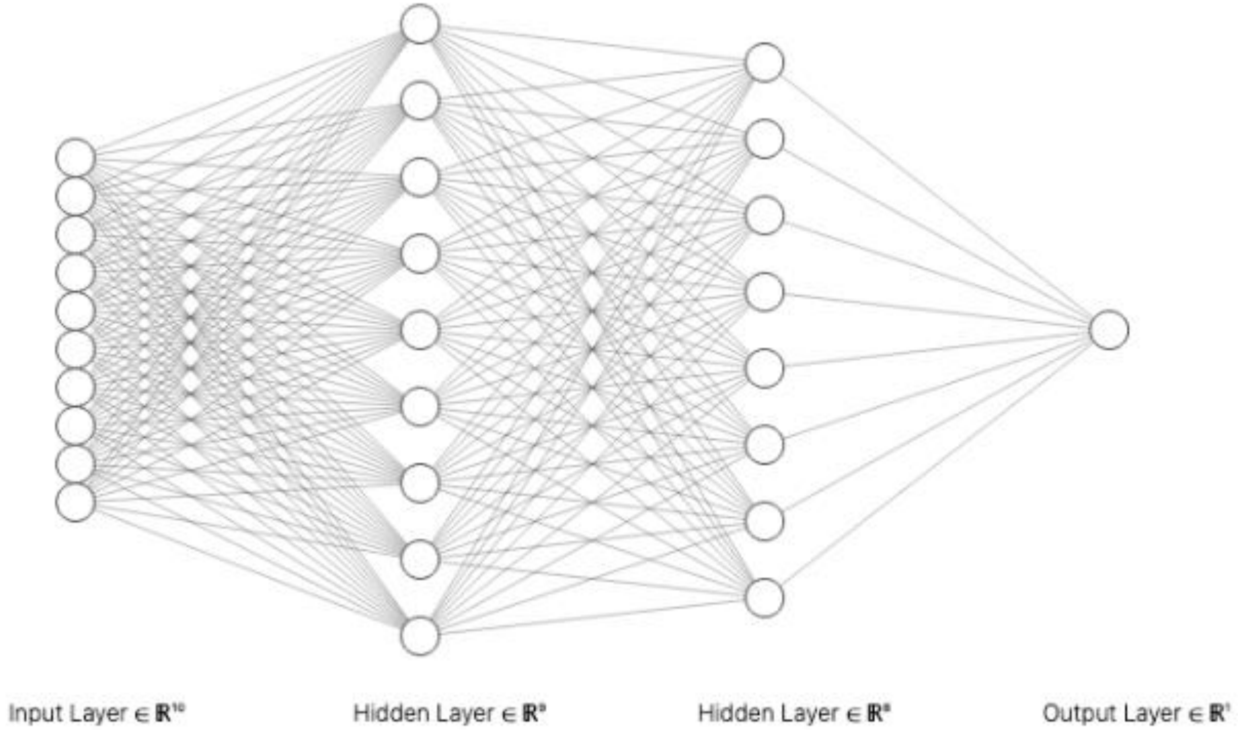
the algorithm requires that the data is **linearly separable**, If the data is not linearly separable, the line will end up bouncing around, never halting



Perceptron Learning Algorithm (PLA)

- **Multi-layered Perceptrons**

are sets of perceptrons linked together to form more complex models. These perceptrons have **non-linear activation functions**, allowing the model to be able to handle data that are **not linearly separable**




Thank You