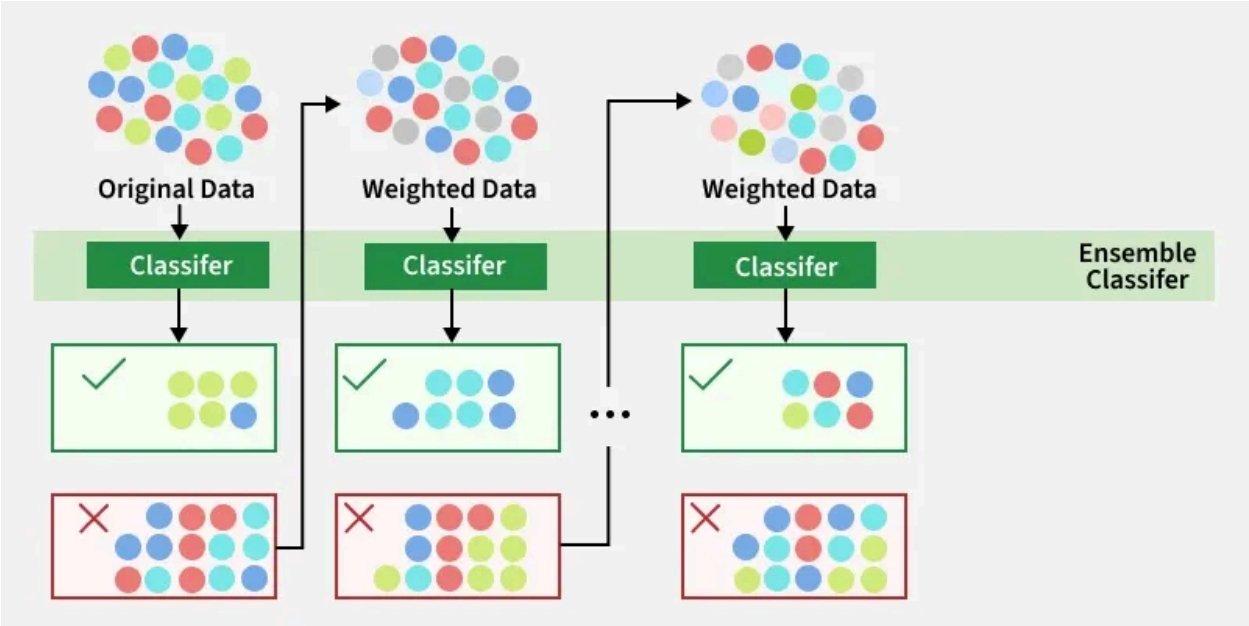
**AdaBoost:-**

AdaBoost is a boosting technique that combines several weak classifiers in sequence to build a strong one. Each new model focuses on correcting the mistakes of the previous one until all data is correctly classified or a set number of iterations is reached.

Think of it like in a class, a teacher focuses more on weak learners to improve its academic performance, similarly boosting work.




## Important warning

If a data point is:

- Noisy

- Wrongly labeled

Its circle keeps growing → AdaBoost overfits

That's why AdaBoost:

- Works great on **clean data**
- Struggles with **noise**

# Real life eg for AdaBoost:-

### Real-world example: Spam Email Detection

#### Goal

Classify emails as:

- **Spam**
- **Not Spam**

---

## What is a "weak classifier" here?

A weak classifier is a **very simple rule**, like:

- **Rule 1:**
  "If email contains the word *FREE*, mark as spam"

- **Rule 2:**
  "If email has more than 3 links, mark as spam"

- **Rule 3:**
  "If sender is unknown, mark as spam"

Each rule:

- Works sometimes
- Fails often
- But better than random guessing

👉 These are **weak classifiers**

---

## Dataset (simplified)

| Email | Contains "FREE" | Many links | Unknown sender | True label |
|---|---|---|---|---|
| E1 | Yes | No | No | Spam |
| E2 | No | Yes | Yes | Spam |
| E3 | No | No | No | Not spam |
| E4 | Yes | No | No | Not spam |
| E5 | No | Yes | No | Spam |

**2.Rule:**

"If email contains 'FREE', mark as spam"

**Results**:

| Email | Correct? |
|---|---|
| E1 | ✅ |
| E2 | ❌ |
| E3 | ✅ |
| E4 | ❌ |
| E5 | ❌ |

Mistakes: **E2, E4, E5**

### Step 3: Update weights (THIS is key)

**AdaBoost says:**

> **"These emails fooled me. Pay MORE attention to them next time."**

**So:**

- **Increase weight of E2, E4, E5**
- **Decrease weight of E1, E3**

**Now:**

- **E2, E4, E5 shout louder**
- **Next classifier *cannot ignore them***

## Train Weak Classifier #2 (with new weights)

**Rule:**

> **"If email has many links → spam"**

**This rule is chosen because it performs well on high-weight emails.**

**Results:**

- **Corrects E2 and E5**
- **Might still fail on E4**

---

## Update weights again

**Now:**

- **Weight of E4 increases more**
- **Others decrease**

**AdaBoost is basically saying:**

> **“Everyone is doing fine except E4. Focus on E4 now.”**

---

## Train Weak Classifier #3

**Rule:**

> **“If sender is unknown → spam”**

**This rule finally gets E4 right.**

---

## Final decision (strong classifier)

**AdaBoost does not take simple majority voting.**

**Instead:**

- **Each weak classifier gets a vote strength**
- **Better classifiers get louder votes**

**Final decision:**

> **“If the weighted vote says spam → spam”**

---

# What “weight” means in ONE LINE

**Weight = how important a data point is when training the next weak classifier**

**OR**

**Weight = penalty for misclassifying that example**