**🎯 Should we always standardize?**

**No — you should not blindly standardize in every algorithm.**

Why? Because standardization removes the “weight” that comes from the natural scale of the variables.  
If one variable is *intentionally* more important than the others, standardizing would wrongly treat all variables as equally important.

**📌 Rule of thumb:**

| **Scenario** | **Should you standardize?** |
| --- | --- |
| You want all variables to contribute equally | ✅ Yes |
| Some variables are *inherently more important* (domain knowledge, business rules) | 🚫 No |
| You don’t know yet, just exploring | ✅ Yes (good default) |

**🔷 Standardization vs Normalization — Are they the same?**

**No — they are different techniques. But the same principle about *variable importance* applies to both.**

**📌 1️⃣ What is Standardization?**

We transform the data so that:

* Mean = 0
* Standard deviation = 1

Formula:

z=x−mean(x)std(x)z = \frac{x - \text{mean}(x)}{\text{std}(x)}z=std(x)x−mean(x)​

So values are centered around 0 and scaled by their variance.  
Example range: can be negative and positive, e.g., −3-3−3 to +3+3+3.

**📌 2️⃣ What is Normalization?**

We transform the data so that:

* Values are scaled into a fixed range, usually [0, 1].

Formula:

xnorm=x−xminxmax−xminx\_{\text{norm}} = \frac{x - x\_{\text{min}}}{x\_{\text{max}} - x\_{\text{min}}}xnorm​=xmax​−xmin​x−xmin​​

So the smallest value becomes 0, and the largest becomes 1.

**🔷 Why do we use them?**

Both are used when:

* Data has variables with different scales.
* Algorithm is sensitive to scale (e.g., K-means, KNN).

**🔷 What about variable importance?**

👉 Same rule applies:  
If you already know some variables are **more important**, you **shouldn’t blindly normalize or standardize**, because both techniques will equalize their influence.

**📊 Comparison Table**

| **Feature** | **Standardization** | **Normalization** |
| --- | --- | --- |
| Output Range | ≈ -3 to +3 | [0, 1] |
| Uses mean & std? | ✅ Yes | 🚫 No |
| Uses min & max? | 🚫 No | ✅ Yes |
| Handles outliers better? | ✅ Yes | 🚫 No |
| Equalizes scales? | ✅ Yes | ✅ Yes |
| Removes variable importance? | ✅ If misused | ✅ If misused |