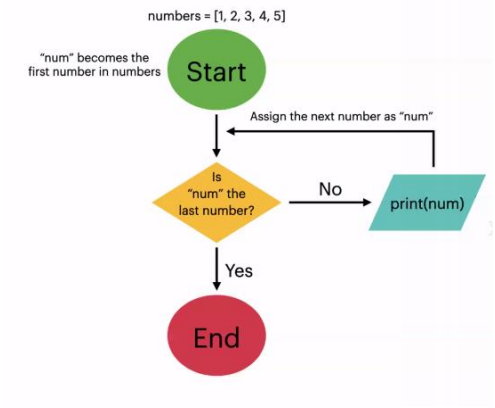
8. Class

Python : For lopp

A **for loop** is a **repetition structure** that iterates over each element in a sequence (like a list, dictionary, or string) one by one.

Key Idea: It allows you to work with each element individually.



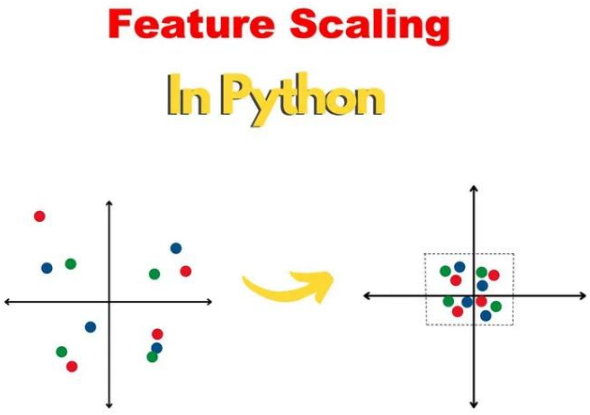
For loop — bu **takrorlash operatori**, u ma'lum bir to'plam yoki ketma-ketlik (masalan, ro'yxat, lug'at, qator) ichidagi har bir elementni navbatma-navbat o'tib chiqish uchun ishlatiladi.

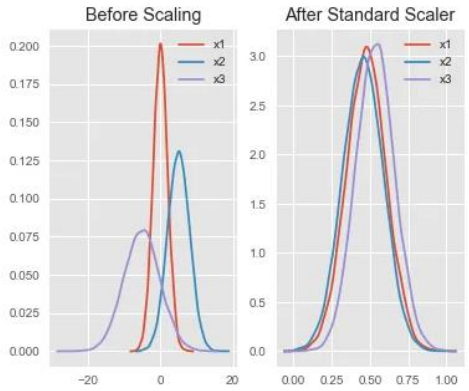
Scaling qilishdan oldin dataga quyidagi ishlovlar berilishi kerak.

SCALING

- Data preprocessing
- missing values
- encoding
- scaling

Scaling bo'lishi uchun raqamli bo'lishi kerak hamma data





Types of Scaling

Here are three common ways to scale data:

1. Min-Max Scaling (Normalization)

- **What it does:** Changes all numbers to fit between 0 and 1 (or another range you want).
- **How it works:** It finds the smallest and largest numbers in your data, then adjusts every number so the smallest becomes 0, and the largest becomes 1.

2. Standardization (Z-Score Scaling)

- **What it does:** Adjusts the numbers so they have a mean of 0 and a standard deviation of 1.
- **How it works:** Subtracts the average (mean) of all numbers, then divides by how spread out the numbers are (standard deviation).

3. Robust Scaling

- **What it does:** Similar to Standardization but focuses on the middle range of numbers (median) instead of the mean. This is good if your data has outliers (extreme values).
- **How it works:** It subtracts the median and divides by the interquartile range (IQR – the range between the 25th and 75th percentiles).
- **Why Use It:** It's not affected by very large or very small numbers like Min-Max and Standardization.

4. Misol bilan tushuntirish

Scalingdan oldin:

- Haqiqiy qiymatlar: $y = [10, 20, 30]$, bashoratlar: $\hat{y} = [12, 19, 28]$.
- $\text{MSE} = \frac{(10-12)^2+(20-19)^2+(30-28)^2}{3} = 3.67$.
- $R^2 = 1 - \frac{\text{SS}_{\text{residual}}}{\text{SS}_{\text{total}}} = 0.9$.

Scalingdan keyin:

- Haqiqiy qiymatlar: $y' = [0, 0.5, 1]$, bashoratlar: $\hat{y}' = [0.067, 0.467, 0.933]$.
- $\text{MSE} = \frac{(0-0.067)^2+(0.5-0.467)^2+(1-0.933)^2}{3} = 0.0067$.
- Lekin R^2 **o‘zgarmaydi!** Chunki R^2 nisbiy baholash usuli.

---

5. Xulosa

- MSE va R^2 bir-biri bilan bog‘liq, lekin ular turli maqsadlarda ishlatiladi:
  - **MSE**: Absolyut xatolarni o‘lchaydi va hajmga sezgir.
  - R^2: Modelning ishlashini foizda baholaydi va scale-dan ta’sirlanmaydi.

1. Regression va Classification uchun mos metrikalar

Vazifa	Metrika	Ma’nosi	Foydalanish sohasi
Regression	MSE (Mean Squared Error)	Bashorat va haqiqiy qiymatlar o‘rtasidagi kvadrat xatolarning o‘rtacha qiymatini o‘lchaydi.	Regression modellari.
	MAE (Mean Absolute Error)	Bashorat va haqiqiy qiymatlar orasidagi o‘rtacha modul xatosi.	Regression modellari.
	R^2 (Determination Coefficient)	Modelning natijalardagi umumiy o‘zgarishni qanchalik tushuntira olganini foizlarda ko‘rsatadi.	Regression modellari.
Classification	Accuracy	To‘g‘ri tasniflangan sinflar ulushini o‘lchaydi.	Classification modellari.
	Precision va Recall	Muvozanatsiz datasetlarda to‘g‘ri va noto‘g‘ri ajratilgan sinflarni o‘lchaydi.	Imbalanced classification.
	F1-Score	Precision va Recall o‘rtasidagi balansni o‘lchaydi.	Imbalanced classification.
	ROC-AUC	Modelning tasniflash aniqligini baholaydi (to‘g‘ri/noto‘g‘ri sinflar uchun).	Classification modellari.

2. r va R^2 o‘rtasidagi farqlar

Xususiyat	r: Korrelatsiya Koeffitsienti	R^2: Determinatsiya Koeffitsienti
Nima o‘lchanadi?	Ikki o‘zgaruvchi orasidagi bog‘liqlik.	Regression modelning umumiy mosligini (qanchalik yaxshi ishlashini) baholaydi.
Qiymat oraliq	-1 dan $+1$ gacha.	0 dan 1 gacha.
Musbat qiymat	$r = +1$: Mukammal musbat bog‘liqlik.	$R^2 = 1$: Model barcha o‘zgarishni tushuntiradi.
Manfiy qiymat	$r = -1$: Mukammal manfiy bog‘liqlik.	R^2: Manfiy bo‘lmaydi, faqat 0 dan 1 oralig‘ida bo‘ladi.
Bog‘liqlik	$R^2 = r^2$ (faqat oddiy regressiyada).	$R^2 \neq r^2$ (murakkab regressiyada).
Foydalanish sohasi	Korrelatsiya tahlili uchun.	Regression modellarning natijalarini baholash uchun.

Quyidagi jadvalda **scaling kerak bo‘lgan** va **kerak bo‘lmagan** algoritmlar qisqacha, sodda sabablari bilan keltirilgan:

Scaling Kerakmi?	Algoritmlar	Nega?
Kerak	**KNN (K-Nearest Neighbors)**	Masofa asosida ishlaydi, katta diapazon xatolikka olib keladi.
	SVM (Support Vector Machines)	Hyperplane ajratishda katta diapazon muvozanatsizlik keltiradi.
	PCA (Principal Component Analysis)	Dispersiyani hisoblashda katta qiymatlar asosan ta’sir qiladi.
	Gradient Descent (Neural Networks)	Gradient optimallashtirishda katta qiymatlar sekinlashuv va xatoga olib keladi.
Kerak emas	**Decision Tree**	Xususiyatlarning o‘lchami ta’sir qilmaydi, faqat qiymatlarni solishtiradi.
	Random Forest	Har bir daraxt mustaqil ishlaydi va xususiyat diapazonlariga bog‘liq emas.
	Gradient Boosting	Daraxt tuzilmasi asosida ishlaydi, ma’lumot hajmi qaror qabul qilishga ta’sir qilmaydi.