* Normalization :

The goal of normalization is to transform features to be in a similar scale.

→ Benifits :

* Helps model converge more quickly during training. when different features GD can "bounce" and slow convergence. "More advanced optimizers like Adagrad and Adam" Protect against this issue.

* Helps model in better Predictions.

* Helps avoid the "NaN" trap when feature values are high. When a value in a model exceeds the floating point Precision limit, It will be set to NaN (Not an Number).

* The model Pays too much attention to features with wide ranges and not for features with narrow ranges.

The 3 Polular normalization methods :

→ Linear scaling
→ Z-Score scaling
→ Log scaling
→ Clipping (Not a true Normalization technique)

⇒ Linear Scaling : LS means converting floating point values from their natural range into a standard range.

$$x' = (x - x_{min}) / (x_{max} - x_{min})$$

original value: x

maximum value in the dataset: x_{max}

scaled value: x'

minimum value in the dataset: x_{min}

ex :

$x = 300$

$x_{min} = 100$

$x_{max} = 900$

$$x' = (300 - 100) / (900 - 100)$$

$$x' = 200 / 800 = 0.25$$

→ LS is a good choice when :

1.) The lower and upper bounds don't change much overtime.

2.) When the outliers are less and aren't extreme.

3.) The feature is uniformly distributed across it's range.

for ex age range lies b/w 0 to 100 where there is a very low chance for outliers so LS is the best suite for above case.

* **Note** : Most real-world features do not meet all the criteria for linear-scaling. Z-score scaling is typically a better normaliza-tion choice than linear scaling.

* Z-Score Scaling :

A Z-Score scaling is the number of standard deviations a value is from the mean. For ex, a value that is 2 standard deviations greater than the mean has a score of +2.0.

$$x' = (x - \mu) / \sigma$$

x' → Z-Score; x → Raw Value; μ → mean; σ → standard deviation

for ex,

mean = 100

Standard deviation = 20

original Value = 130

therefore,

$$x' = (130 - 100) / 20$$

$$x' = 30 / 20 = 1.5$$

* Z-Score is a good choice when the data follows a normal distribution or a distribution similar to that.

* In some situations the Z-Score scaling method will be combined with another form of normalization (usually clipping) to handle those situations.

Que : Suppose our model trains on a feature named "height" because this feature that holds 10 million adult heights of women would Z-Score be a better normalization method?

Ans : Yes, because the height feature conforms the normal distributi

* Log Scaling :

Log scaling computes the logarithm of the raw Value. In theory, the logarithm could be of any base but in Practical it usually calculates the natural logarithm (ln).

$$x' = \ln(x)$$

↳ natural logarithm of x

ex : $X = 54.598$

$x' = \ln(54.598) = 4.0$

→ Log scaling is helpful when the data conforms to a "Powerlaw distribution" :

- Low Values of X have high Values of Y
- As the X Value increaces Y value decreaces quickly and Viceversa.

Ex : Movie ratings :

→ A few movies have lot of user rating (low x Values high Y Values).

→ Most movies have very few user ratings (High x Values low Y Values).

and for other example we can consider book sales :

→ Most Published books sell upto 100 / 200 copies

→ Some books have moderate sell of copies

→ Bestsellers sell more than million copies.

* Clipping :

Clipping is a technique to minimize the influence of extreme outliers. In brief, clipping usually caps the value of outliers to a specific maximum value.

* For example, if a dataset containing a feature named roomPerPerson (total rooms / total occupants) for various houses well 99% of the feature values conform the normal distributio However, the feature contais a few outliers where some of the extreme. How can we reduce the influence of them?

⇒ What if we simply cap the maximum value of roomPerPerson at an arbitrary value, say 4.0?

It means all the values that are greater than 4.0 now becom 4.0. Well this clipped data is more useful than raw data.

We can also clip values after applying other forms of normaliza for ex, we used z-score scaling but some of them have value > 3 then :

→ clip z-scores that are >3 to be equal to 3.

→ clip z-scores less than -3 to become -3.

* Summary :

Normalization technique	Formula	Usecase
Linear scaling	$x' = (x - x_{min}) / (x_{max} - x_{min})$	when feature is uniformly distributed. (Flat-shaped)
Z-score Scaling	$x' = (x - \mu) / \sigma$	when feature is normally distributed (Bell shaped)
Log scaling	$x' = \ln(x)$	when feature is heavily skewed at least of tail (Heavy tail)
clipping	if $x > max$, set $x' = max$ if $x < min$, set $x' = min$	when features have extreme outliers