Feature Scaling

↳ Feature Scaling is ALWAYS applied to columns

Applied

X1	X2	X3	X4
64	104	81	62
108	93	55	44
30	35	51	63

↳ Feature Scaling is NOT applied across columns i.e it is not applied for a row

X1	X2	X3	X4
64	104	81	62
108	93	55	44
30	35	51	63

Not applied

Feature Scaling

- Normalization
- Standardization

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

range $[0, 1]$

$$x' = \frac{X - \mu}{\sigma}$$

μ → average, mean

σ → standard deviation

range $[-3, +3]$

Let us consider a dataset:

	Annual income ($)	Age (yrs)
1	70000	45
2	60000	44
3	52000	40

10,000 between 1 and 2; 1 between 2 and 4; 8000 between 2 and 3; 4 between 3 and 4

is person2 similar to person1 or person3

unscaled values (differences)

Comparing Salary to years – not compauable

↳ Important to scale your features

After Normalization

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

	Annual income ($)	Age (yrs)
more similar person 1,2		
1	1	1
2	0.444	0.8
3	0	0

$x' = \frac{45-40}{45-40} = 1$

$x' = \frac{44-40}{45-40} = \frac{4}{5} = 0.8$

$x' = \frac{40-40}{45-40} = 0$

$x' = \frac{52000-52000}{70000-52000} = 0$

$$X^1 = \frac{70,000 - 52000}{70,000 - 52000} = 1$$

$$X^1 = \frac{60,000 - 52000}{70,000 - 52000} = 0.44$$

According to Annual Income column, person 2 is similar to both person 1, person 3

According to Age column person 2 is similar to person 1

Standardization Vs Normalization

Normalization is recommended when we have a normal distribution in most of our features

Standardization works well all the time

Recommended