Normalization

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distoring differences in the ranges of values or losing information.

- MinMax Scaling ✓
- Mean Normalization
- Max absolute
- Robust Scaling

__Weight__

130
67
81
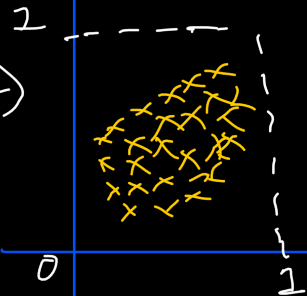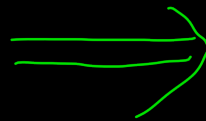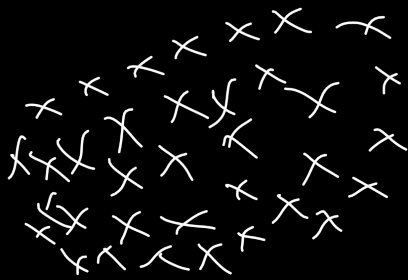61
32
54
⋮
100 wts

Normalize
Min Max Scaling

$Min = 32$

$Max = 130$

$$x'_n = \frac{x_n - x_{min}}{x_{max} - x_{min}}$$

$$= \frac{130 - 32}{130 - 32}$$

$$= 1$$

Min Max Scaling Range = $[0, 1]$

height

Whole data points fit with same distribution in range 0 to 1.

Weight

Mean Normalization

wt
200
100

→ Normalize →

$$\Pi_i^1 = \frac{\Pi_i - \Pi_{mean}}{\Pi_{max} - \Pi_{min}}$$

→ Mean centering

range = [-1 to 1]

Value > Mean +ve
Value < Mean -ve

There is no class in SK learn for this.

This is useful for those algorithms where we need Centered data.
↳ Better we use Standardization

Max Absolute Scaling

Wt

200
100
300

→

$$\Pi_i^1 = \frac{\Pi_i}{|\Pi_{max}|}$$

MaxAbsScaler (SKleam class)

Useful where we have Sparse data (data with more no of zeros)

Robust Scaling

wt
200
300
400

→ $$\Pi_i^1 = \frac{\Pi_i - \Pi_{median}}{IQR}$$

Robustscale class

→ Robust to outliers

(Very useful if data has lots of outliers)

Normalization vs Standardization

few question to ask before Scaling

1) IS feature Scaling required ?

2) (Min Max)

when you have
idea about
min value and
max value

Image Processing → CNN

color channel
0 - 255

Most of the problem are solved using Standardization

→ Better try all if not sure ←
which will work better. It is
all about experiment.