CIS 678 Machine Learning

Feature scaling

Feature Scaling

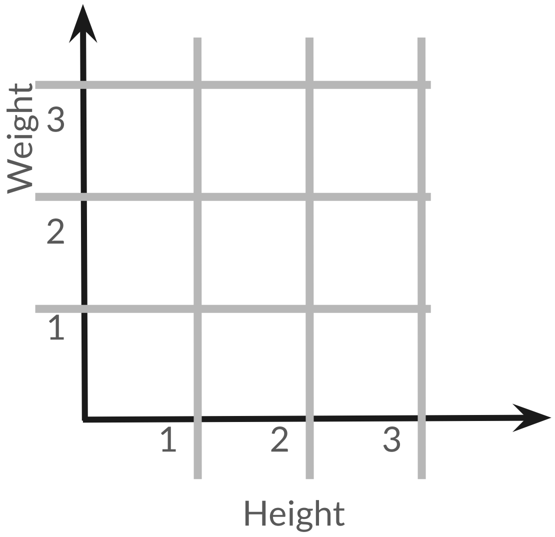
Lets we are given a set of data points

	height(meter)	weight(kg)
0	1.50	70
1	1.70	80
2	1.80	82
3	1.60	75
4	1.75	78

Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

	height(meter)	weight(kg)
0	1.50	70
1	1.70	80
2	1.80	82
3	1.60	75
4	1.75	78

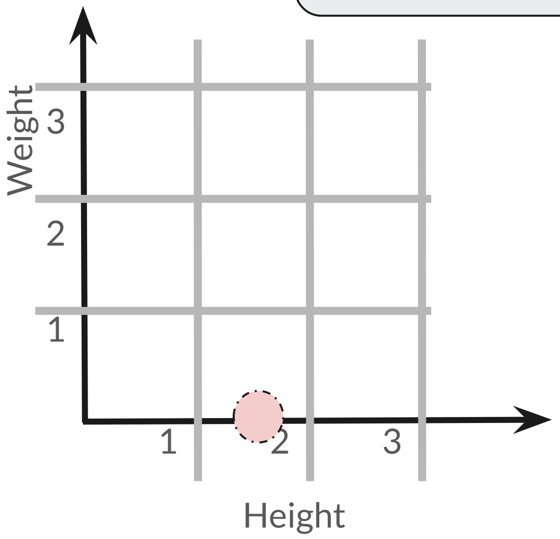


Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

	height(meter)	weight(kg)
0	1.50	70
1	1.70	80
2	1.80	82
3	1.60	75
4	1.75	78

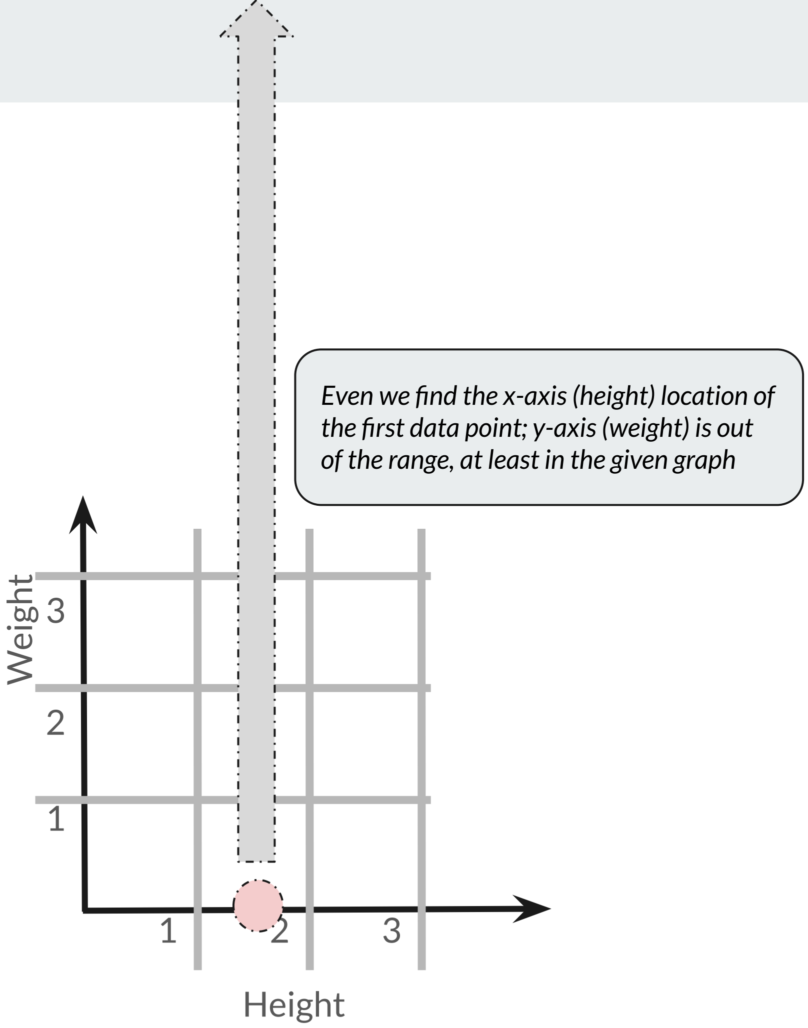
Even we find the x-axis (height) location of the first data point; y-axis (weight) is out of the range, at least in the given graph



Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

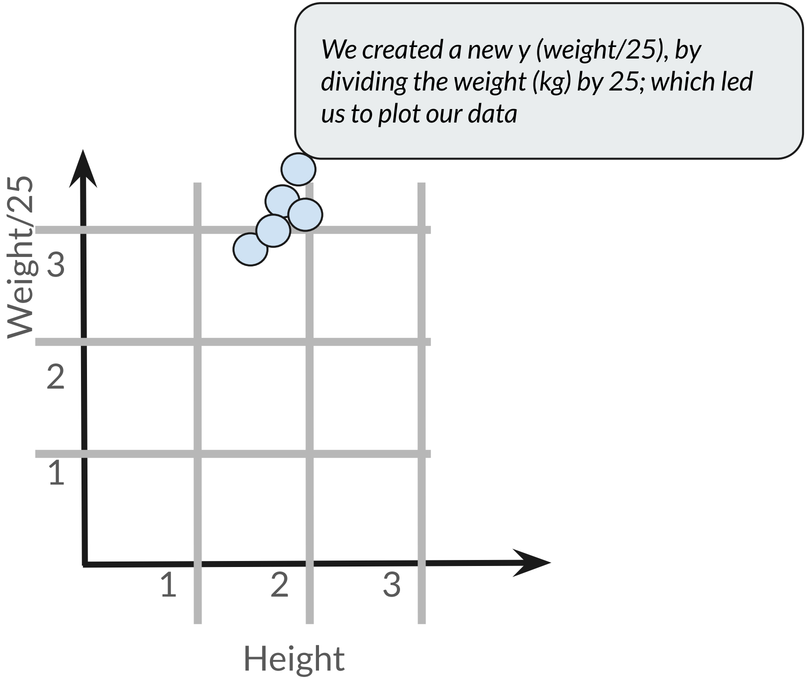
	height(meter)	weight(kg)
0	1.50	70
1	1.70	80
2	1.80	82
3	1.60	75
4	1.75	78



Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

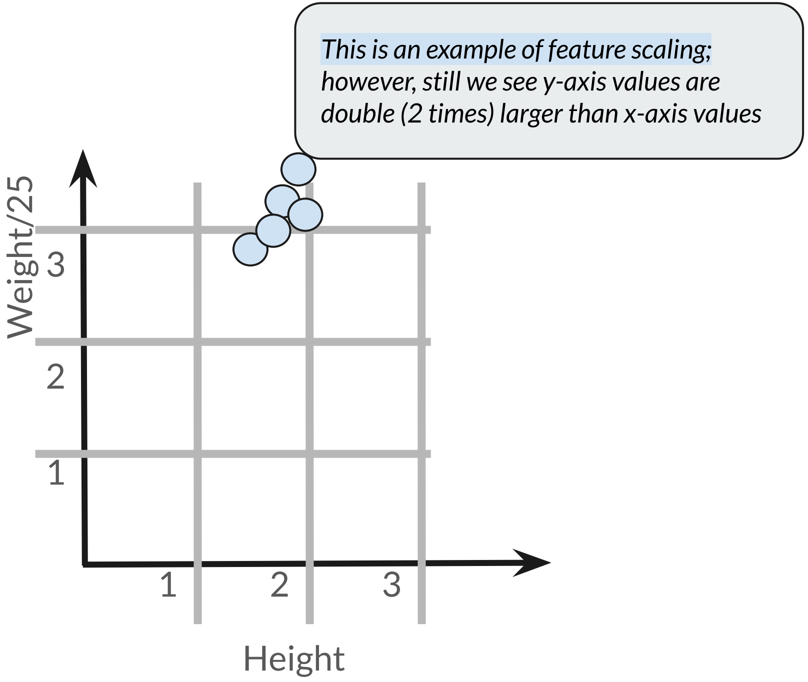
	height(meter)	weight(kg)	weight(kg)/25
0	1.50	70	2.80
1	1.70	80	3.20
2	1.80	82	3.28
3	1.60	75	3.00
4	1.75	78	3.12



Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

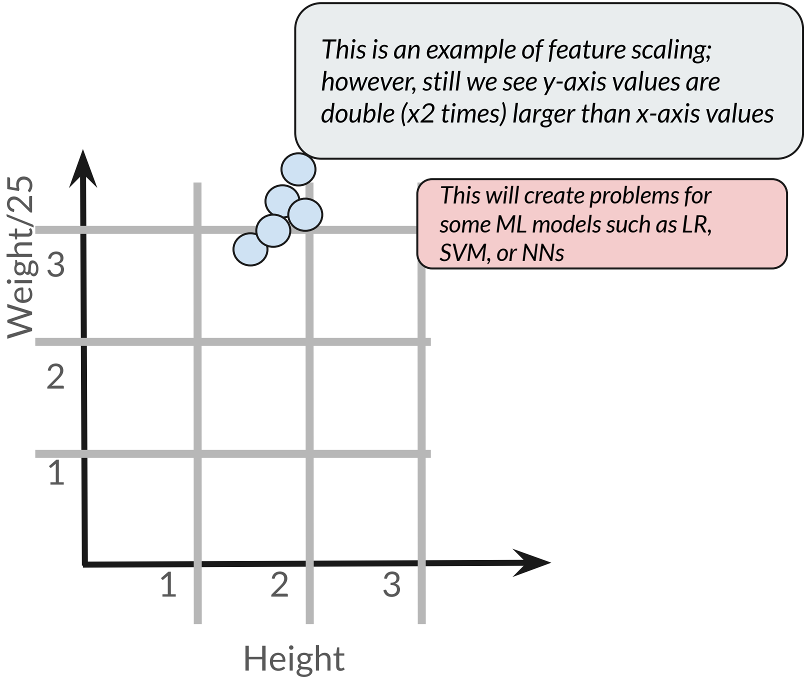
	height(meter)	weight(kg)	weight(kg)/25
0	1.50	70	2.80
1	1.70	80	3.20
2	1.80	82	3.28
3	1.60	75	3.00
4	1.75	78	3.12



Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

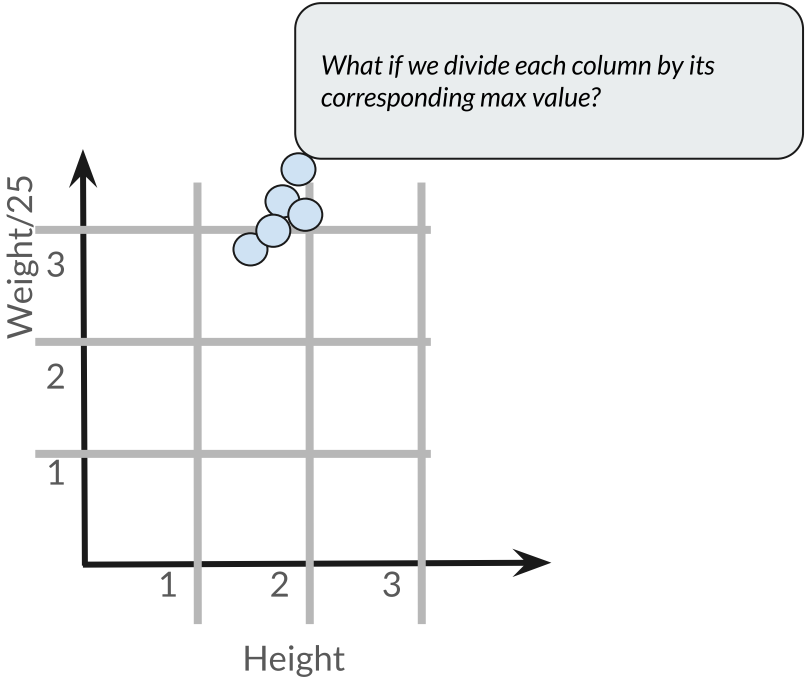
	height(meter)	weight(kg)	weight(kg)/25
0	1.50	70	2.80
1	1.70	80	3.20
2	1.80	82	3.28
3	1.60	75	3.00
4	1.75	78	3.12



Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

	height(meter)	weight(kg)	weight(kg)/25
0	1.50	70	2.80
1	1.70	80	3.20
2	1.80	82	3.28
3	1.60	75	3.00
4	1.75	78	3.12



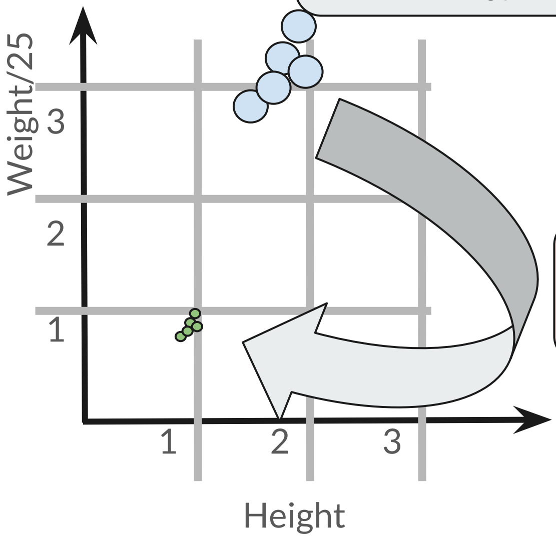
Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

	height(meter)	weight(kg)	weight(kg)/25	height/max(height)	weight/max(weight)
0	1.50	70	2.80	0.83	0.85
1	1.70	80	3.20	0.94	0.98
2	1.80	82	3.28	1.00	1.00
3	1.60	75	3.00	0.89	0.91
4	1.75	78	3.12	0.97	0.95

What if we divide each column by its corresponding max value?

- *We will get all feature values within [0, 1], assuming positive values*



Center of each circle represents corresponding feature values; size of circles only for illustration purpose.

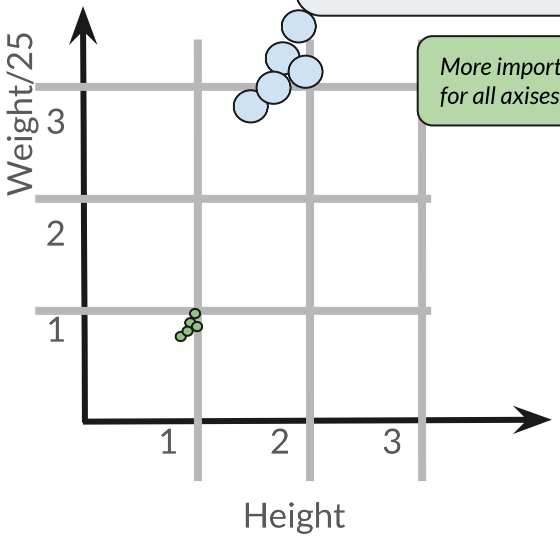
Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

	height(meter)	weight(kg)	weight(kg)/25	height/max(height)	weight/max(weight)
0	1.50	70	2.80	0.83	0.85
1	1.70	80	3.20	0.94	0.98
2	1.80	82	3.28	1.00	1.00
3	1.60	75	3.00	0.89	0.91
4	1.75	78	3.12	0.97	0.95

What if we divide each column by its corresponding max value?

- *We get all feature values within [0, 1], assuming positive values*

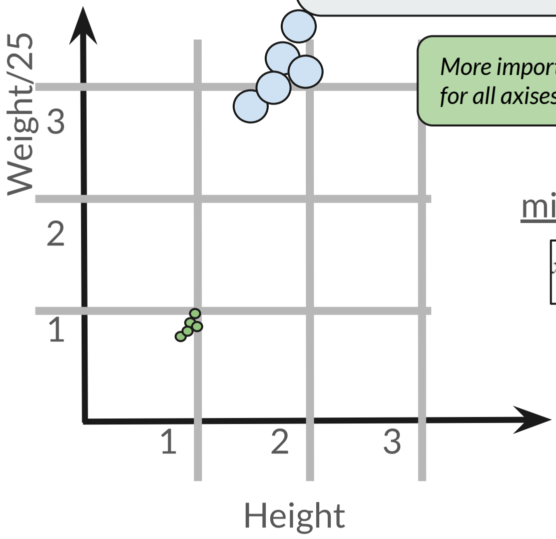


More importantly, now it's fair for all axises; right?

Feature Scaling

Lets we are given a set of data points; we want to plot them on a 2D cartesian plane (vector space)

	height(meter)	weight(kg)	weight(kg)/25	height/max(height)	weight/max(weight)
0	1.50	70	2.80	0.83	0.85
1	1.70	80	3.20	0.94	0.98
2	1.80	82	3.28	1.00	1.00
3	1.60	75	3.00	0.89	0.91
4	1.75	78	3.12	0.97	0.95



What if we divide each column by its corresponding max value?

- *We get all feature values within [0, 1], assuming positive values*

More importantly, now it's fair for all axises; right?

min-max scalar

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

min = 0;
for this particular example

Feature Scaling

We have other scaling as well:

- Standard scalar

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation

Feature Scaling

We have other scaling as well:

- Standard scalar

	height(meter)	weight(kg)	height_ss	weight_ss
0	1.50	70	1.58	1.67
1	1.70	80	-0.28	-0.72
2	1.80	82	-1.21	-1.19
3	1.60	75	0.65	0.48
4	1.75	78	-0.74	-0.24

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation

- *Scaled features are centered across the Zero-axis*

Feature Scaling

We have other scaling as well:

- Standard scalar

	height(meter)	weight(kg)	height_ss	weight_ss
0	1.50	70	1.58	1.67
1	1.70	80	-0.28	-0.72
2	1.80	82	-1.21	-1.19
3	1.60	75	0.65	0.48
4	1.75	78	-0.74	-0.24

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation

- *Scaled features are centered across the Zero-axis*
- *NN community uses this extensively*