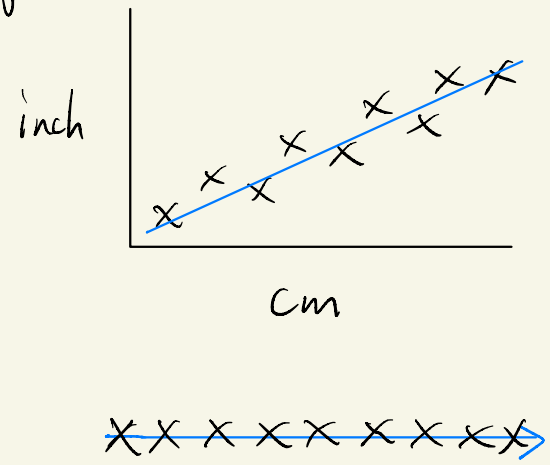
Dimensionality Reduction

May 29, 2021

a.k.a. Data compression

Another unsupervised learning task, like clustering.

e.g. Reduce data from 2D to 1D



✡ Maybe data was rounded off previously.

$x^{(i)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow z^{(i)} = [z]$

$x^{(1)} \in \mathbb{R}^2 \rightarrow z^{(1)} \in \mathbb{R}^1$

$x^{(2)} \in \mathbb{R}^2 \rightarrow z^{(2)} \in \mathbb{R}^2$

$\vdots$

$x^{(m)} \rightarrow z^{(m)}$

✡ Now you can use a 1D feature to represent the data instead of a 2D one.

↳ Results in faster learning algorithm.

✡ You can also do 3D → 2D, 1000D → 100D, etc.

Data Visualization!

$x \in \mathbb{R}^{50}$

eg.

Country	x_1 GDP (trillions of US$)	x_2 Per capita GDP (thousands of intl. $)	x_3 Human Develop-ment Index	x_4 Life expectancy	x_5 Poverty Index (Gini as percentage)	x_6 Mean household income (thousands of US$)	...
Canada	1.577	39.17	0.908	80.7	32.6	67.293	...
China	5.878	7.54	0.687	73	46.9	10.22	...
India	1.632	3.41	0.547	64.7	36.8	0.735	...
Russia	1.48	19.84	0.755	65.5	39.9	0.72	...
Singapore	0.223	56.69	0.866	80	42.5	67.1	...
USA	14.527	46.86	0.91	78.3	40.8	84.3	...
...	...	...	...	...	...	...	

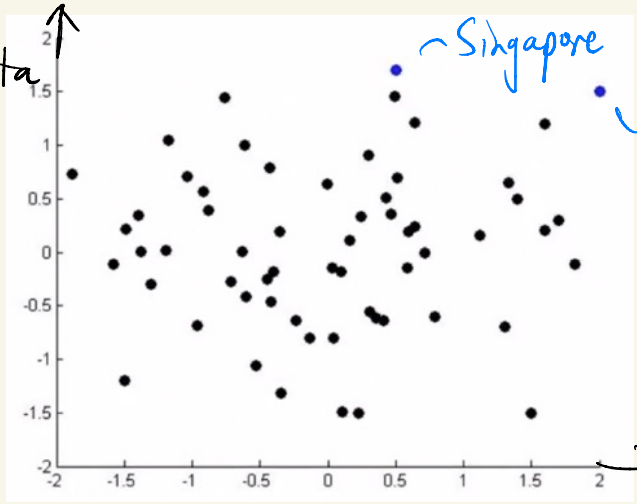
Country	z_1	z_2
Canada	1.6	1.2
China	1.7	0.3
India	1.6	0.2
Russia	1.4	0.5
Singapore	0.5	1.7
USA	2	1.5
...	...	...

Reduce data from 50D to 2D.

$z^{(i)} \in \mathbb{R}^2$

* Sometimes, using dimensionality reduction to allow for plotting data helps w/ understanding the data better.

Plotted!


GDP per capita
Singapore
USA
Z_2
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
Country's GDP
Z_1