Outline

1. What is Machine Learning??
2. The Wonders of Machine Learning
3. Data, Models and ML Tasks
4. Supervised Learning
    1. Regression
    2. Classification
5. **Unsupervised Learning**
    1. Dimensionality Reduction
    2. Density Estimation

Unsupervised Learning

- Unsupervised learning is ‘understanding data’
- Data: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$
- $\mathbf{x}^i \in \mathbb{R}^d$
- Build models that compress, explain and group data.

Unsupervised Learning Application

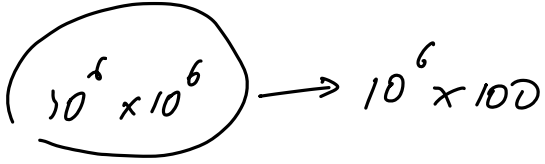
Tweet 1

⋮

Tweet 999999

Group the million tweets into 10 manageable groups

Dimensionality Reduction



E.g.: Represent a million gene expression levels of a million people, using just 100 numbers per person.

Dimensionality reduction: compression and simplification.

Dimensionality Reduction

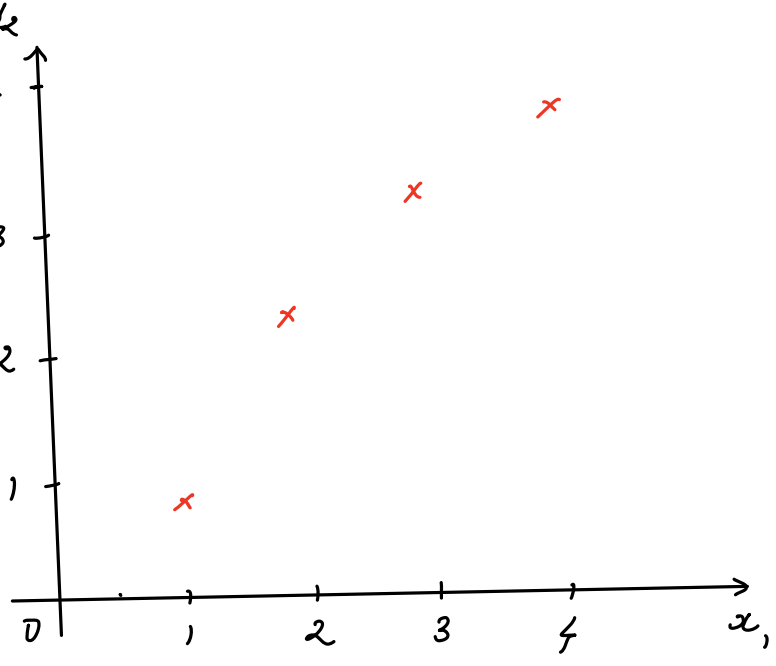
- Data: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$
- $\mathbf{x}^i \in \mathbb{R}^d$
- Encoder $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$
- Decoder $g : \mathbb{R}^{d'} \rightarrow \mathbb{R}^d$
- Goal : $g(f(\mathbf{x}^i)) \approx \mathbf{x}^i$
- Loss $= \frac{1}{n} \sum_{i=1}^{n} \|g(f(\mathbf{x}^i)) - \mathbf{x}^i\|^2$

$d' << d$

Dimensionality Reduction Illustration

$d = 2$, $d' = 1$, $n = 4$

	f	g
$[1, 0.8]$	0.2	$[0.2, 0.2]$
$[2, 2.2]$	-0.2	$[-0.2, -0.2]$
$[3, 3.2]$	-0.2	$[-0.2, -0.2]$
$[4, 3.8]$	0.2	$[0.2, 0.2]$



$f(x) = x_1 - x_2$

$g(u) = [u, u]$

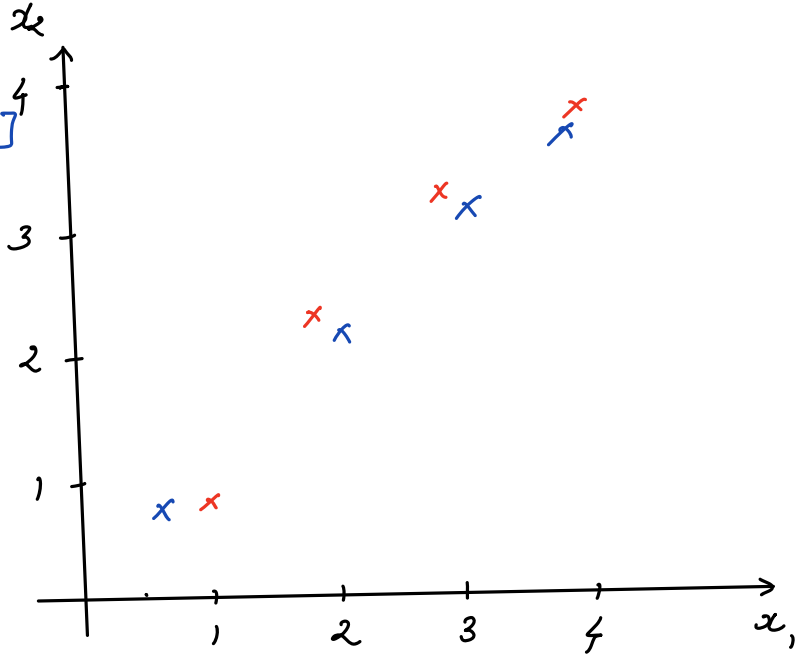
$\breve{f}(x) = \frac{x_1 + x_2}{2}$

$\breve{g}(u) = [u, u]$

Dimensionality Reduction Illustration

$d = 2$, $d' = 1$, $n = 4$

	$\tilde{f}$	$\tilde{g}$
$[1, 0.8]$	0.9	$[0.9, 0.9]$
$[2, 2.2]$	2.1	$[2.1, 2.1]$
$[3, 3.2]$	3.1	$[3.1, 3.1]$
$[4, 3.8]$	3.9	$[3.9, 3.9]$



$f(x) = x_1 - x_2$

$g(u) = [u, u]$

$\tilde{f}(x) = \frac{x_1 + x_2}{2}$

$\tilde{g}(u) = [u, u]$

$10^6 \times 100$

$\hat{f}$, $\tilde{g}$