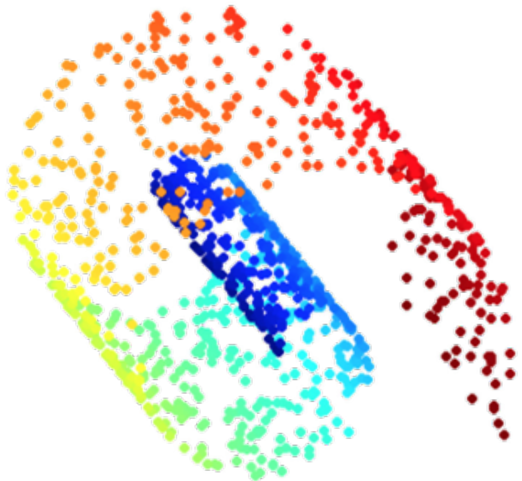
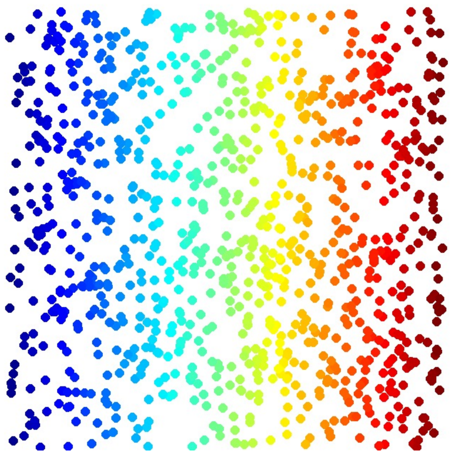
Dimensionality Reduction

Part I

Dimensionality Reduction: Part I



Looks like 3-D

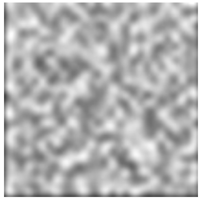


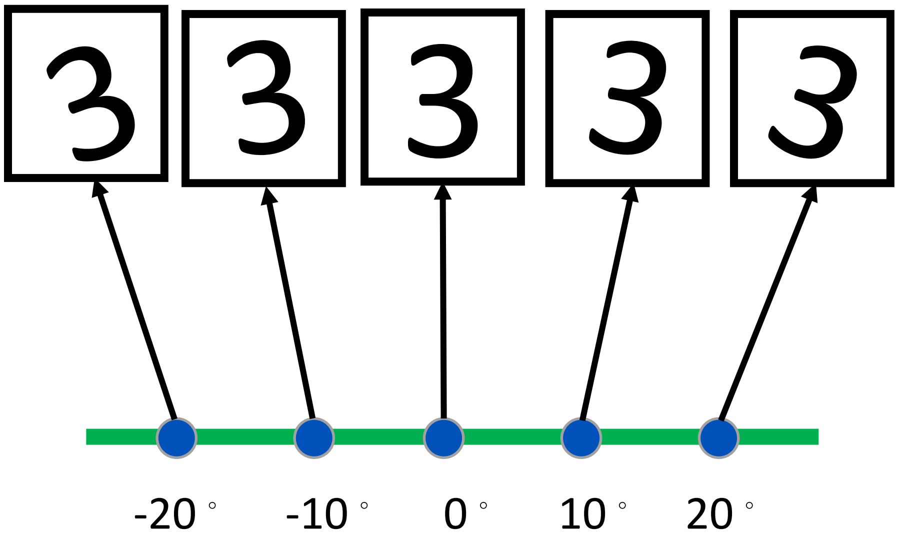
Actually, 2-D

Dimensionality Reduction: Part II

In MNIST, a digit is 28 x 28 dims.

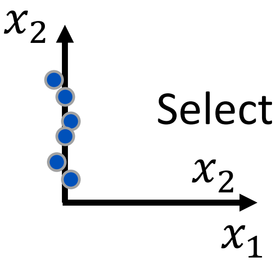
- Most 28 x 28 dim vectors are not digits

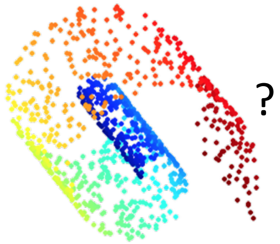




Distributed Representation

- Feature selection

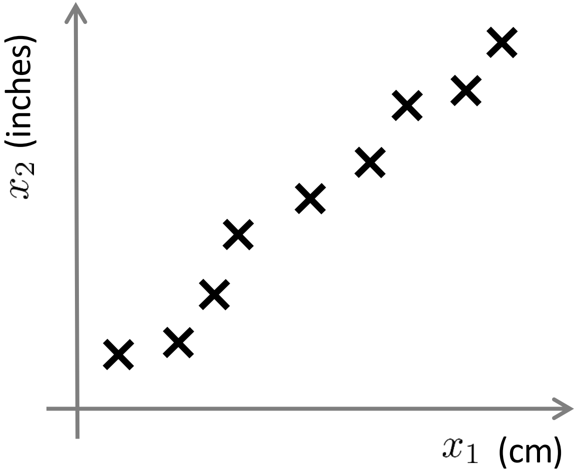




- Principle component analysis (PCA)

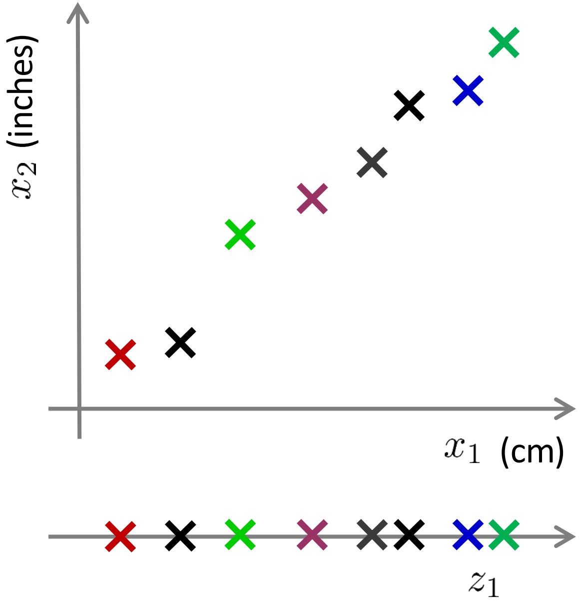
$$z = Wx$$

Data Compression: Part I



Reduce data from 2D to 1D

Data Compression: Part II



Reduce data from 2D to 1D

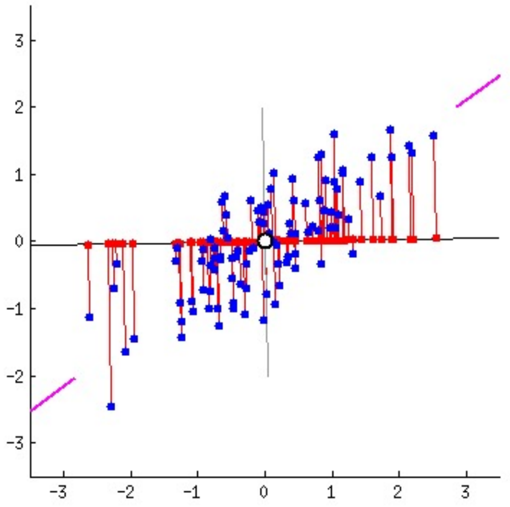
$$x^{(1)} \rightarrow z^{(1)}$$

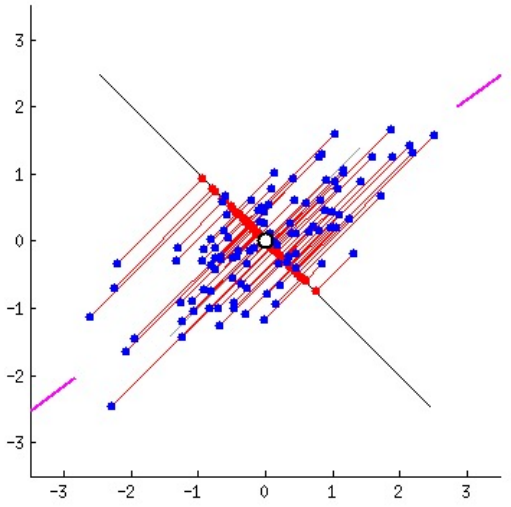
$$x^{(2)} \rightarrow z^{(2)}$$

$$\vdots$$

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

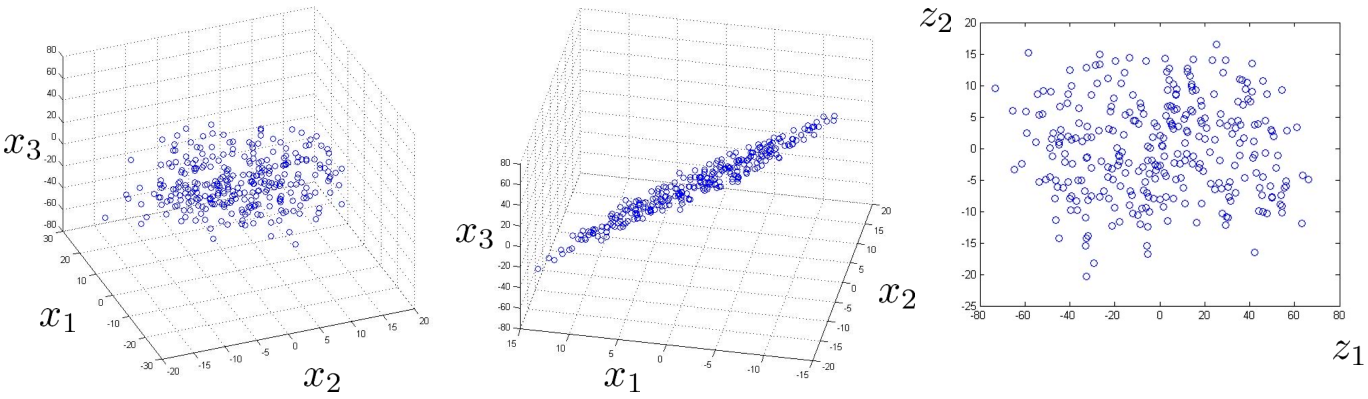
Data Compression: Part III





Data Compression: Part IV

Reduce data from 3D to 2D



Data Visualization: Part I

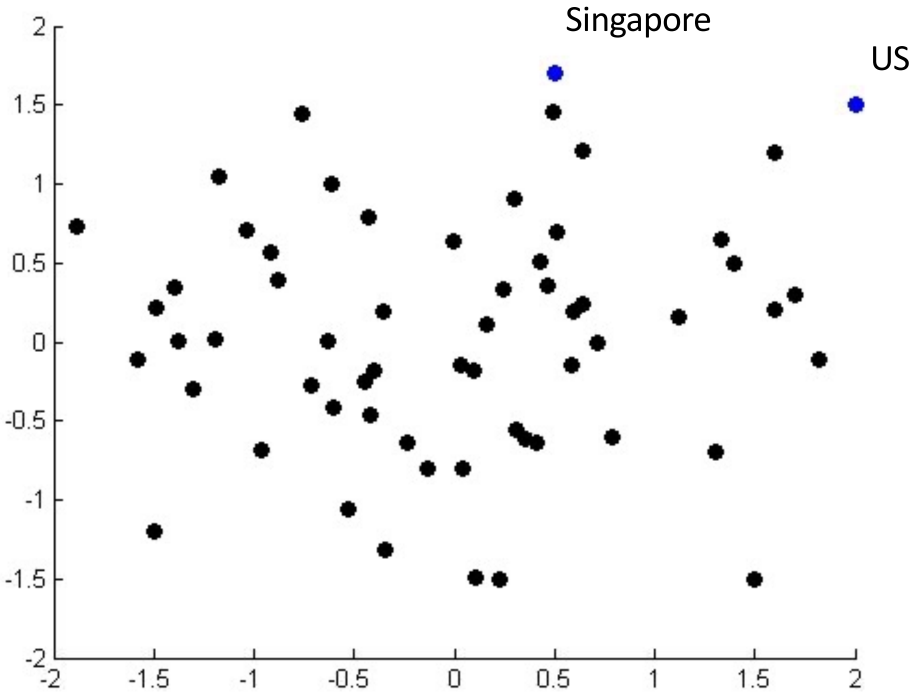
Country	GDP (trillions of US$)	Per capita GDP (thousands of intl. $)	Human Develop-ment Index	Life expectancy	Poverty Index (Gini as percentage)	Mean household income (thousands of US$)	…
Canada	1.577	39.17	0.908	80.7	32.6	67.293	…
China	5.878	7.54	0.687	73	46.9	10.22	…
India	1.632	3.41	0.547	64.7	36.8	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	…
…	…	…	…	…	…	…	

[resources from en.wikipedia.org]

Data Visualization: Part II

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
…	…	…

Data Visualization: Part III

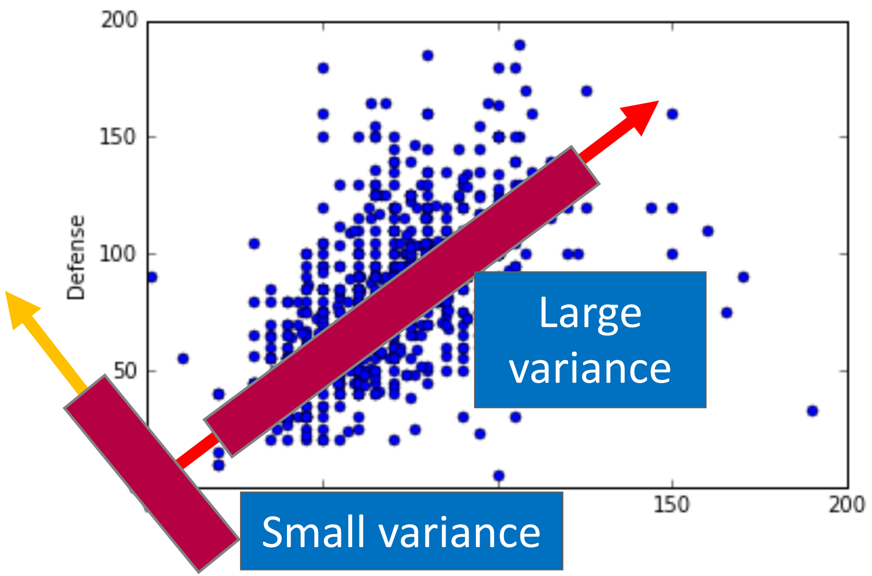


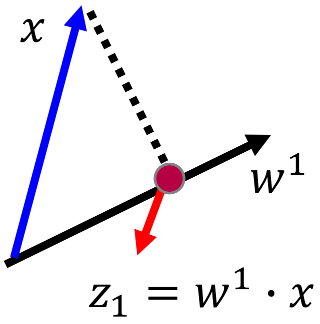
PCA: Part I

$$z = Wx$$

Reduce to 1-D:

$$z_1 = w^1 \cdot x$$





Project all the data points x onto w^1, and obtain a set of z_1

We want the variance of z_1 as large as possible

$$Var(z_1) = \frac{1}{N}\sum_{z_1}(z_1 - \bar{z_1})^2 \quad \|w^1\|_2 = 1$$

PCA: Part II

$$z = Wx$$

Reduce to 1-D:

$$z_1 = w^1 \cdot x$$

$$z_2 = w^2 \cdot x$$

$$W = \begin{bmatrix} (w^1)^T \\ (w^2)^T \\ \vdots \end{bmatrix}$$

Orthogonal matrix

Project all the data points x onto w^1, and obtain a set of z_1

We want the variance of z_1 as large as possible

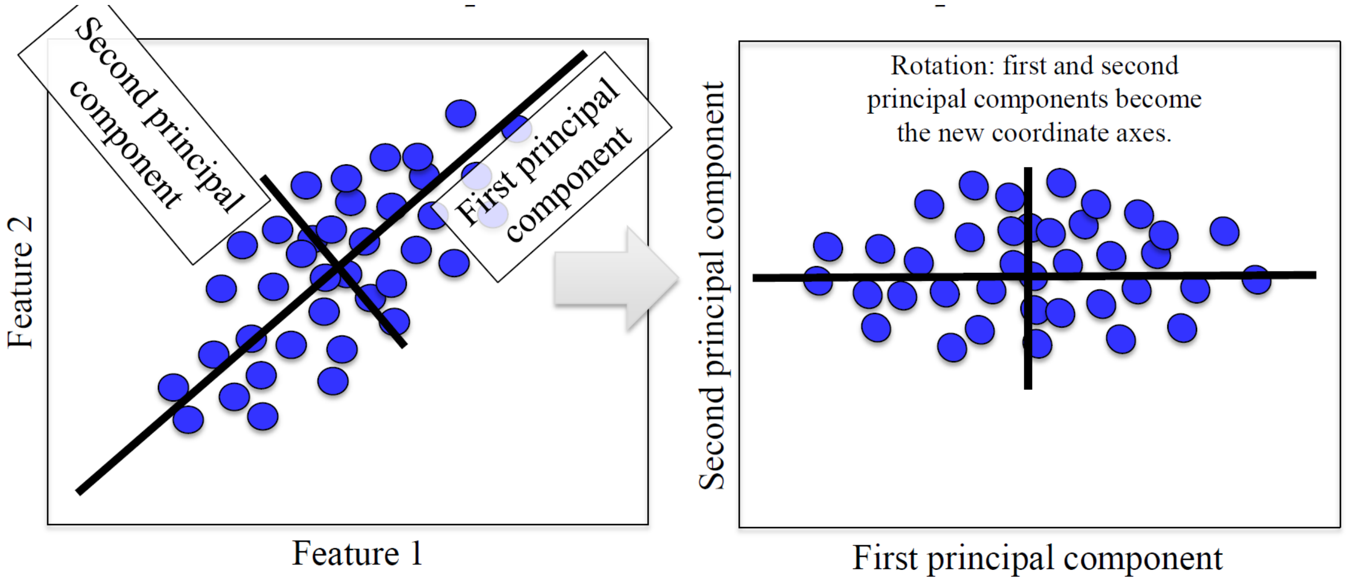
$$Var(z_1) = \frac{1}{N} \sum_{z_1} (z_1 - \bar{z_1})^2 \quad \|w^1\|_2 = 1$$

We want the variance of z_2 as large as possible

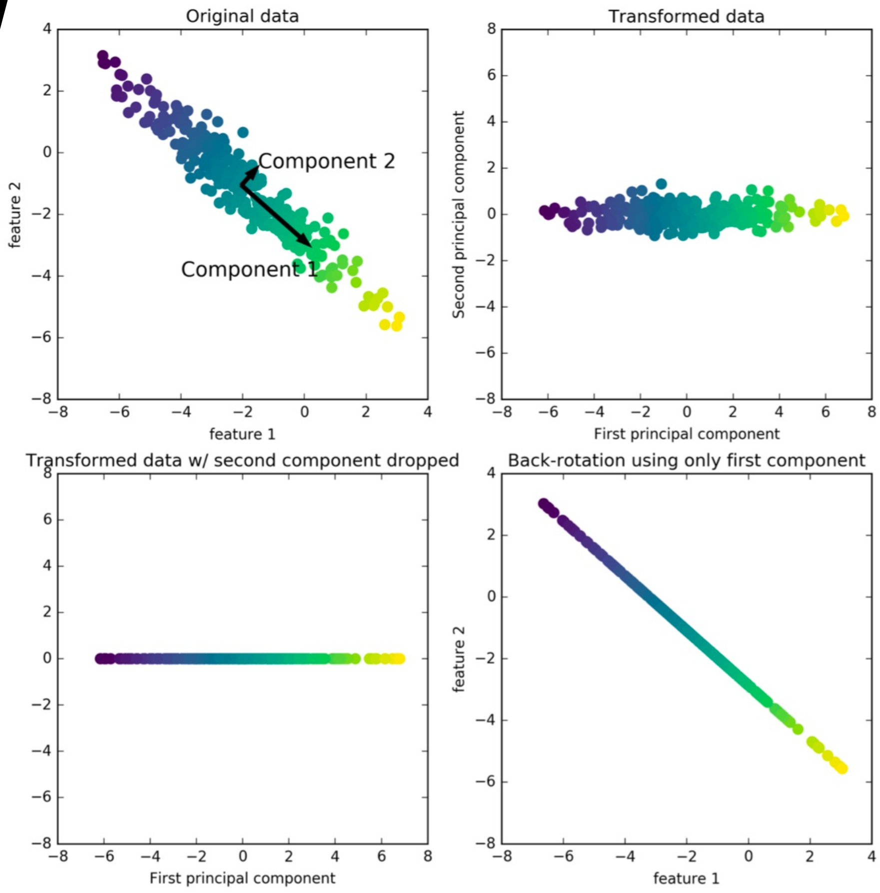
$$Var(z_2) = \frac{1}{N} \sum_{z_2} (z_2 - \bar{z_2})^2 \quad \|w^2\|_2 = 1$$

$$w^1 \cdot w^2 = 0$$

PCA: Part III


Second principal component
First principal component
Feature 2
Feature 1
Rotation: first and second principal components become the new coordinate axes.
Second principal component
First principal component

PCA: Part IV



Original data
Component 2
Component 1
feature 1
feature 2
Transformed data
First principal component
Second principal component
Transformed data w/ second component dropped
First principal component
Back-rotation using only first component
feature 1
feature 2