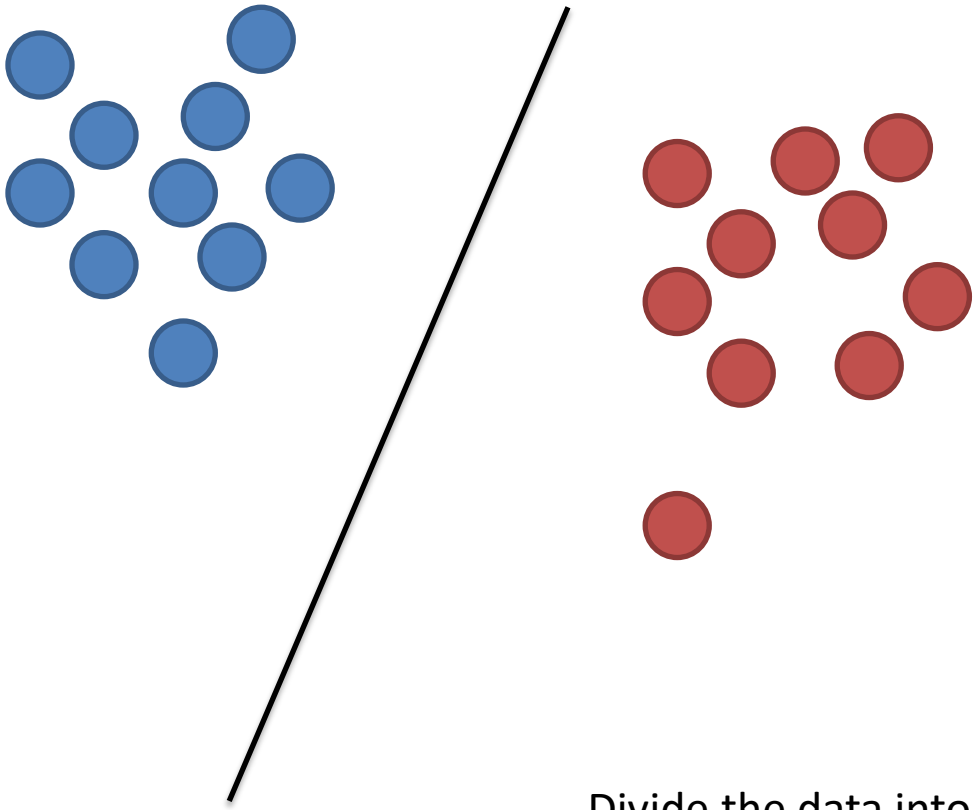
Unsupervised learning

Supervised learning



Divide the data into predefined classes.

Up to this point – supervised learning

- Supervised -> we have labels
- Training data: (x^i, y^i)
- Question:
  - Given x find me the y
  - $P(y|x) = ?$
  - $y = f(x), f = ?$

What we have learned?

- Linear algorithms:
  - $y = g(w^T x + b)$
  - g is transfer (activation) function chosen to match the problem (logistic sigmoid, linear, tanh)
  - w and b are found via optimization:

$$w, b = \arg\max \sum_i Loss\left(y^i, g\left(w^T x^i + b\right)\right) + Reg(w, b)$$

Loss can be: least squares, cross-entropy, ridge,…

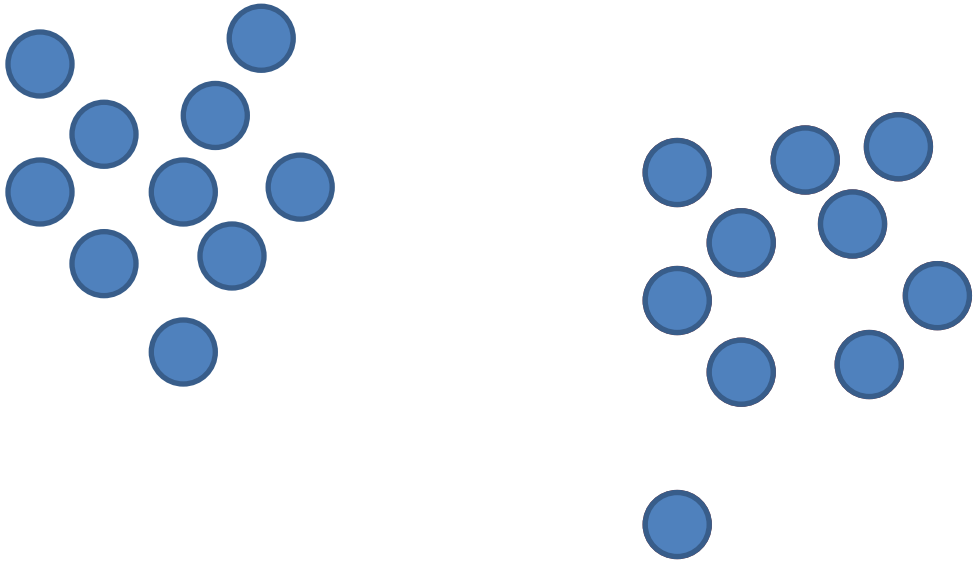
Regularization can: sum of squares, sum of absolute values…

What if the problem is nonlinear?

Transform the data: $x \rightarrow \phi(x)$

- Use a layer (or more) of neurons $x \rightarrow \sigma(W_H^T x + b_H)$
  - The transformation is learned
  - Use the backpropagation algorithm to compute the derivatives with regard to weights and biases
- Use the kernel trick: $K(x, z) = \phi(x)^T \phi(z)$
  - Choose the kernel to match the problem or a good all-purpose one
  - Once chosen, the kernel is fixed (does not adapt!)

Unsupervised learning



In supervised learning we have labels
In unsupervised we don't have them!

Describe the data!:

- Generate samples similar to the data
- Find clusters (distinct groups of points)
- Reduce the dimensionality
- Find good features that describe the data
- …

K-Means – a basic algorithm

Input: m input patterns $x^{(i)}$

1. Initialize K cluster centers $\mu_1 \ldots \mu_k$
   randomly, to some input patterns…
   even better k-means++: sample data points far away from each other: http://theory.stanford.edu/~sergei/slides/BATS-Means.pdf)
2. Loop until convergence:
   1. For all i: set $c^i := \arg\min_j \left\| x^{(i)} - \mu_j \right\|^2$
   2. For all j: set $\mu_j := \frac{\sum_i [c^{(i)}=j] x^{(i)}}{\sum_i [c^{(i)}=j]}$

The K-Means optimization problem

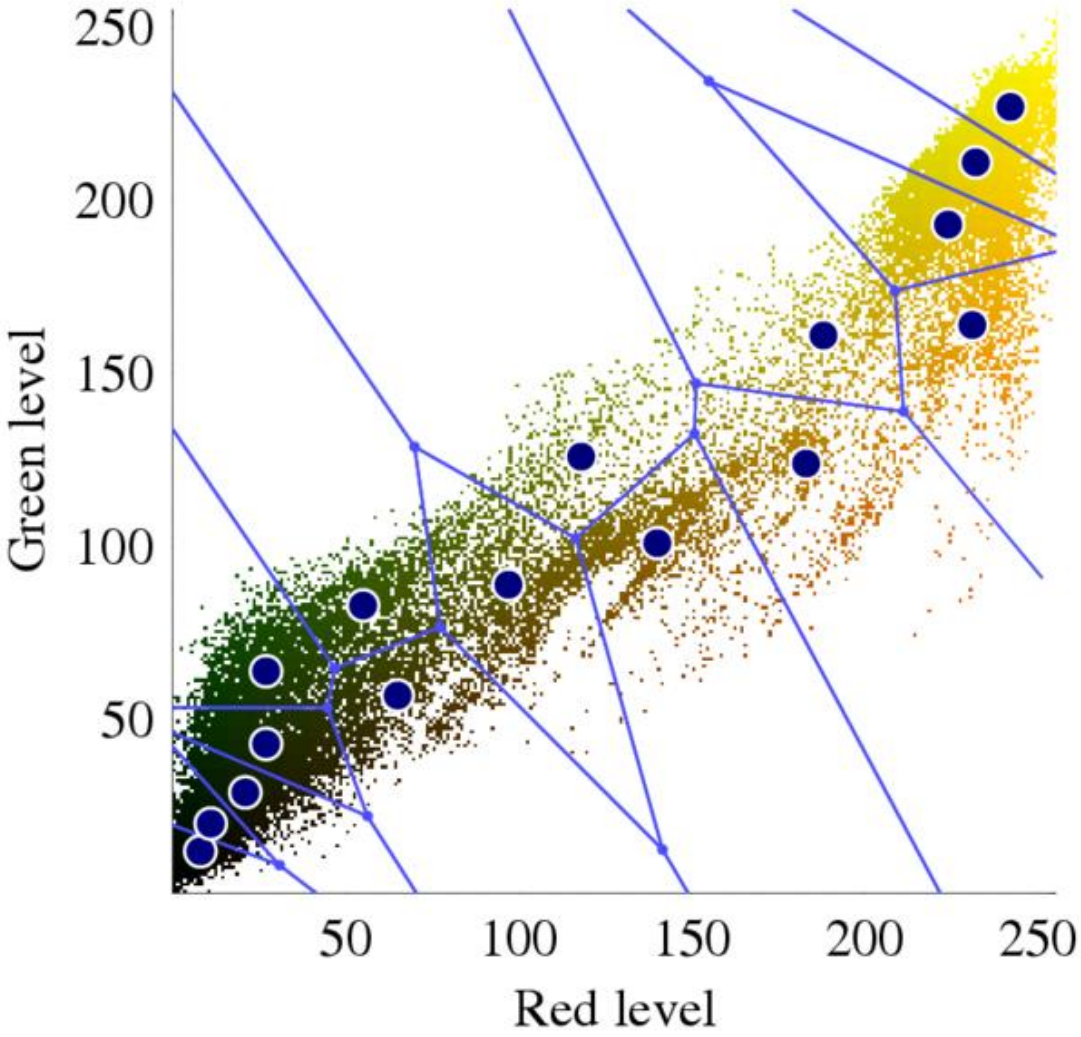
- $J(c, \mu) = \sum_i \left\| x^{(i)} - \mu_{c^{(i)}} \right\|^2$
- The K-means algorithm repeatedely minimizes this over c, then over μ mu etc.

- Important: understand how this ensures that we will terminate training.
- Auto-encoding interpretation:
  - we encode a point as the ID of its cluster
  - decode by simply returning the coordinates of the cluster.
- Energy-based interpretation:
  - Energy is distance
  - Assigning a point to the closest cluster minimizes energy

K-Means for compression

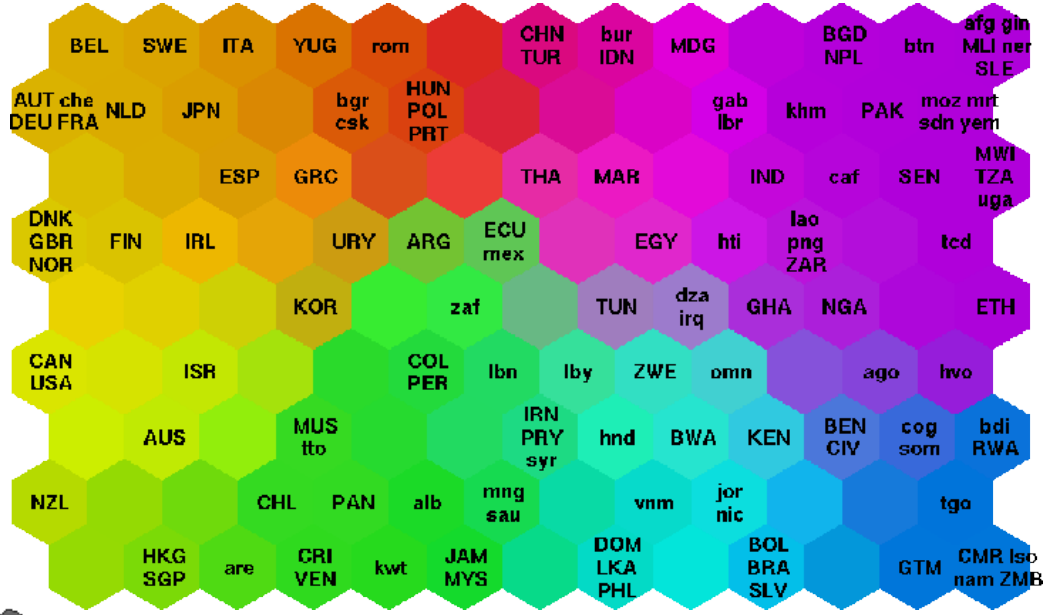
Original: each pixel has two colors, or 16 bits/pixel

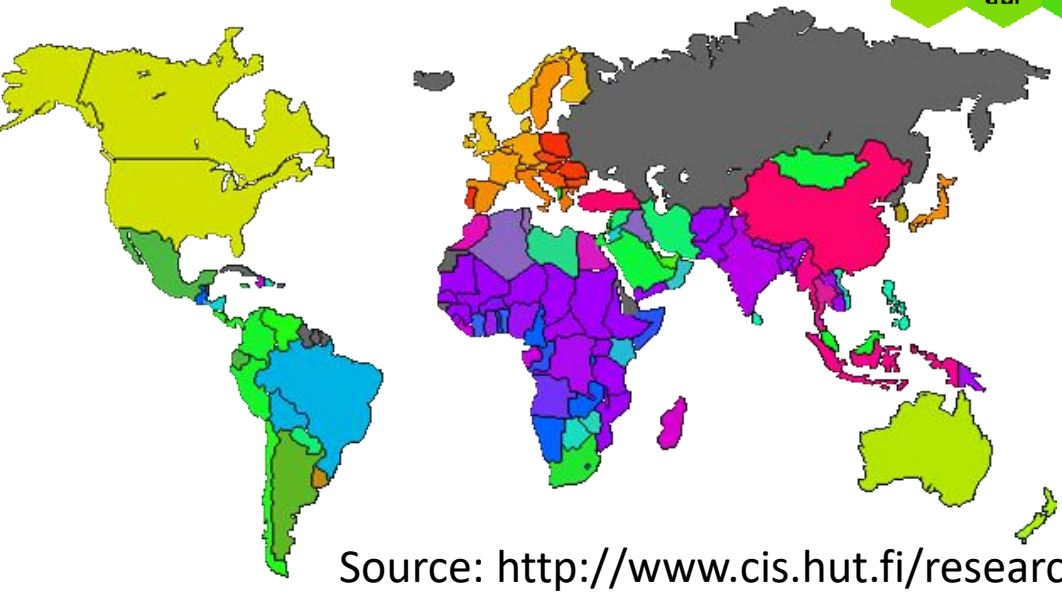
Compressed: each pixel is one of 16 centers, or 4 bits/pixel



Self Organising Maps

A 2D embedding of countries based on 32 life quality indicators made in 1992:





Source: http://www.cis.hut.fi/research/som-research/worldmap.html

Self Organising Maps algorithm

- We have units whose weights point into the data-space
- The units have a topology
- Training:
  - Pick a data point x
  - Find the closest unit (compare the weights to x)
  - Modify x and its neighbors in the chosen topology

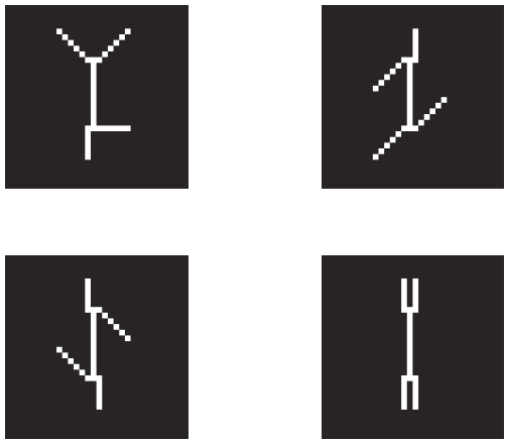
Start with a grid of units.

Each unit points to a location in data space

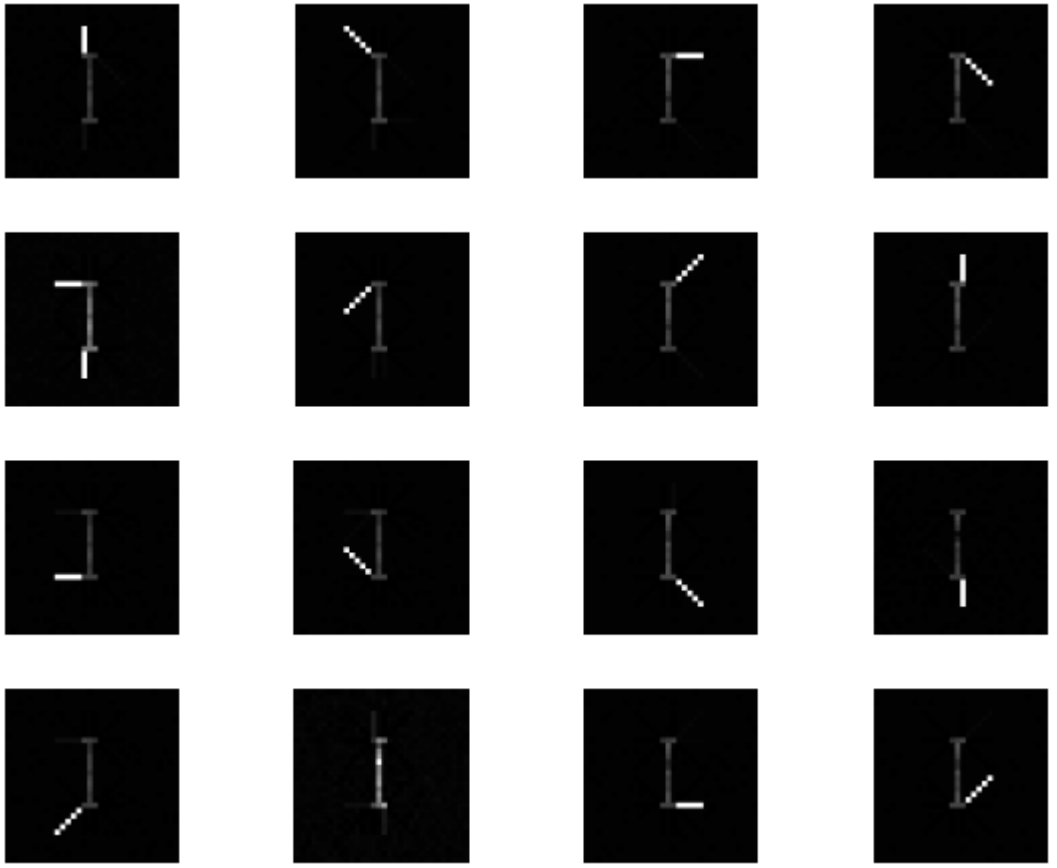
Update a unit and its neighbors in the grid

Dictionary learning

Data

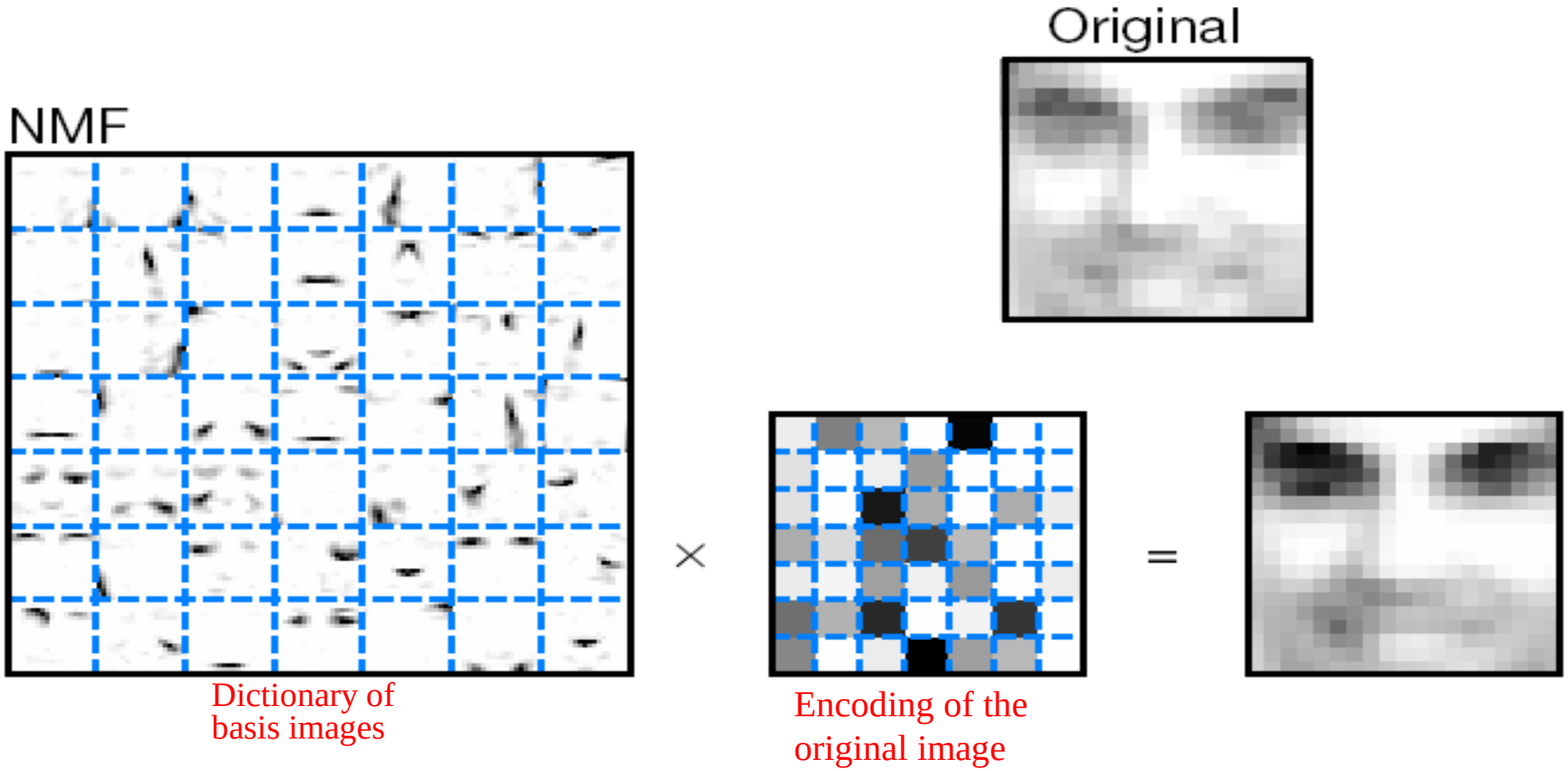


Dictionary



$$x = \mu c$$

Dictionary Learning



dx.doi.org/10.1038/44565

Text dictionaries

Small artificial dataset in bag-of-words format
(Topics: ANIMAL d1 and d4, RELIGION d2, FOOD d3)

$\boldsymbol{X}$ =

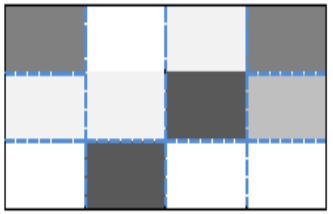
doc_id	d1	d2	d3	d4
dog	5	1	1	6
bible	0	4	1	0
pizza	1	1	5	2
cat	6	0	1	5
tomato	1	1	6	1
god	0	4	0	0

Data = Dictionary x Encoding

d1	d2	d3	d4
Dog	dog	dog	dog
Bible	bible	bible	bible
pizza	pizza	pizza	pizza
cat	cat	cat	cat
tomato	tomato	tomato	tomato
	god		

t1	t2	t3
dog	dog	dog
bible	bible	bible
pizza	pizza	pizza
cat	cat	cat
tomato	tomato	tomato
god	god	god

X



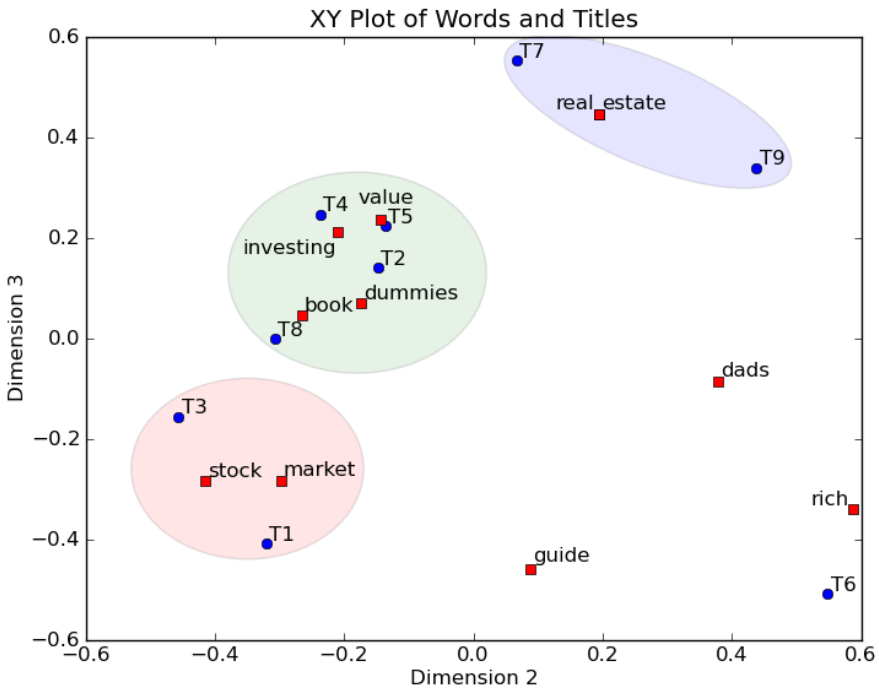
Interpretation: words for topics

Interpretation: documents in topics

SVD example

- 9 documents (keywords underlined)

1. The Neatest Little Guide to Stock Market Investing
2. Investing For Dummies, 4th Edition
3. The Little Book of Common Sense Investing: The Only Way to Guarantee You Fair Share of Stock Market Returns
4. The Little Book of Value Investing
5. Value Investing: From Graham to Buffett and Beyond
6. Rich Dad's Guide to Investing: What the Rich Invest in, That the Poor and the Middle Class Do Not!
7. Investing in Real Estate, 5th Edition
8. Stock Investing For Dummies
9. Rich Dad's Advisors: The ABC's of Real Estate Investing: The Secrets of Finding Hidden Profits Most Investors Miss



Source: http://www.puffinwarellc.com/index.php/news-and-articles/articles/33-latent-semantic-analysis-tutorial.html?start=1