Dimensionality Reduction

Kenneth (Kenny) Joseph

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

What is dimensionality reduction?

- We take our big data matrix and reduce it down to a smaller size
- There are a few reasons we might want to use dimensionality reduction. Can you think of any?

What is dimensionality reduction?

- We take our big data matrix and reduce it down to a smaller size
- There are a few reasons we might want to use dimensionality reduction. Can you think of any?
  - Visualization (why?)
  - To shrink the size of our feature set (why?) efficiency
  - To understand and get rid of correlations between our features… find the "intrinsic dimensionality"

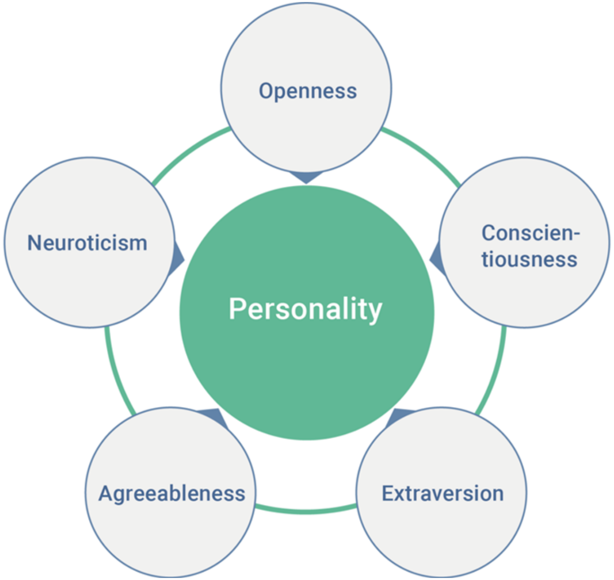
The utility of dimensionality reduction - visualization

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

https://setosa.io/ev/principal-component-analysis/

https://s3-us-west-2.amazonaws.com/lab-apps/pix-plot/index.html#MES25713

The utility of dimensionality reduction – learning about our features



The utility of dimensionality reduction – improve learning

Model (training time)	Redmond	Havel	ninjutsu	graffiti	capitulate
Collobert (50d) (2 months)	conyers lubbock keene	plauen dzerzhinsky osterreich	reiki kohona karate	cheesecake gossip dioramas	abdicate accede rearm
Turian (200d) (few weeks)	McCarthy Alston Cousins	Jewell Arzu Ovitz	- - -	gunfire emotion impunity	- - -
Mnih (100d) (7 days)	Podhurst Harlang Agarwal	Pontiff Pinochet Rodionov	- - -	anaesthetics monkeys Jews	Mavericks planning hesitated
Skip-Phrase (1000d, 1 day)	Redmond Wash. Redmond Washington Microsoft	Vaclav Havel president Vaclav Havel Velvet Revolution	ninja martial arts swordsmanship	spray paint grafitti taggers	capitulation capitulated capitulating

Table 6: Examples of the closest tokens given various well known models and the Skip-gram model trained on phrases using over 30 billion training words. An empty cell means that the word was not in the vocabulary.

https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

How do we do dimensionality reduction?

- Lots and lots and lots of ways
- I am going to introduce three:
  - Principle Component Analysis (PCA)
  - Singular Value Decomposition (SVD)
  - Uniform Manifold Approximation and Prediction (UMAP)
- The first two are intimately related (you can use SVD to solve PCA, and vice versa)
- The last one is neat and a relative newcomer.
  - You do not need to understand the math.

What is PCA? Explanation 1

The goal of principal component analysis is to identify the most meaningful **basis** to re-express a data set. The hope is that this new basis will filter out the noise and reveal hidden structure.

Put another way:

Is there another basis, which is a linear combination of the original basis, that best re-expresses our data set?

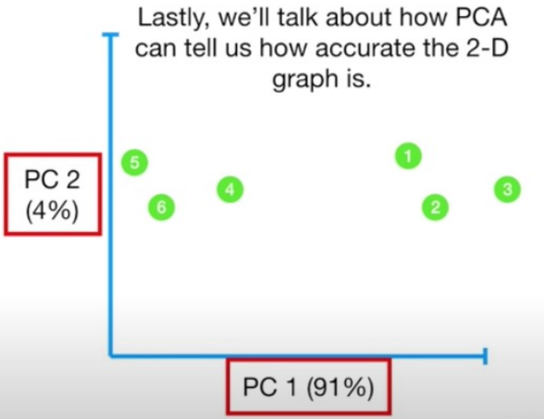
From: https://arxiv.org/pdf/1404.1100.pdf

What is PCA? Explanation 2

An algorithm that aims to **minimize reconstruction error** of the data with a **fixed number of dimensions** (where that number is much smaller than the number of dimensions in the matrix)

How does PCA work?

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1
Gene 3	12	9	10	2.5	1.3	2
Gene 4	5	7	6	2	4	7



https://www.youtube.com/watch?v=FgakZw6K1QQ&ab_channel=StatQuestwithJoshStarmer

Key Points

- PCA assumes
  - **the right goal is to minimize reconstruction error**
  - **Linearity**
  - **~~Orthogonality~~ of the PCs**

max. Σ(distances) along PC.

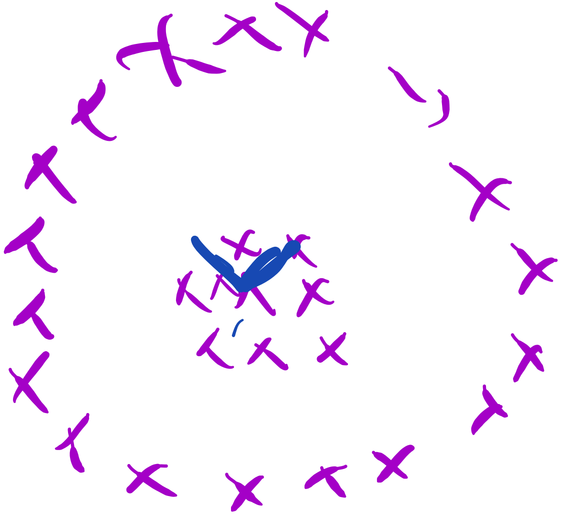
- The **loadings** of features onto each PC represent …

the importance of each feature on each component

- The **skree plot** tells us …

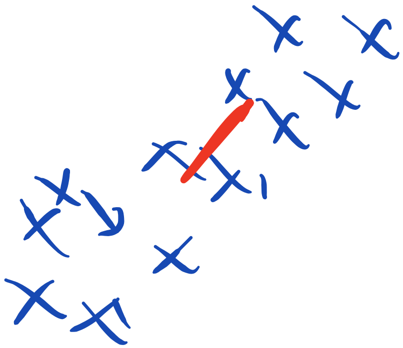
how much variation in the data each PC explains

Will PCA help us on this data? Example 1



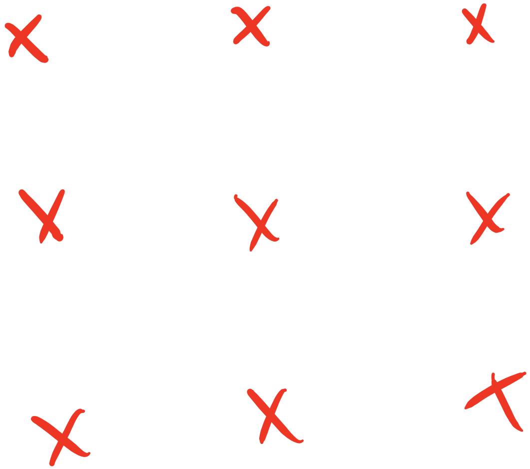
No,
non-linear

Will PCA help us on this data? Example 2



Will PCA help us on this data? Example 3

No,
uniform
dist'n



What will PC1 be...

Draw a dataset where PC1 explains maximal variance

Another visual analysis

https://setosa.io/ev/principal-component-analysis/

PCA vs. SVD

- PCA is SVD after you have **centered the data**
- Why might you **not** want to center data?
- Why might you **want** to center data? ∅

go from sparse → dense ; bad for large datasets

it works slightly better

Code demo

d_1 the dog and the cat.

d_2 I am a dog.

	the	dog	and	cat	i	am	a
d_1	2	1	1	1	0	0	0
d_2	0	1	0	0	1	1	1