PCA, Your Best Friend(?)

Boston useR Meetup

https://bit.ly/3uFat9u

Matthew Curcio
matt.curcio.ri[@]gmail.com
April 13, 2022

Why is it your ‘best’ friend?

Why is it your ‘best’ friend?

It reveals the (spooky) hidden structure in your dataset.

1. Identify important and/or similar features,
2. Reduce the dimensionality of the data,
3. Decrease redundancy in the data,
4. Filter out of the noise from the data,
5. Compress the data,
6. Prepare the data for further analysis using other techniques.
7. Quick & relatively efficient operation on ‘large’ datasets.

The Good, The Bad & The Ugly

The Good

1. PCA preserves the global structure among the data points,
2. It is efficiently applied to large data sets,
3. PCA may provide information on important features found in your data.

The Bad

1. PCA can easily suffer from scale complications,
2. PCA is susceptible to outliers,
   - For example: If the sample size is small this can influence scaling and relative point placement,

The Ugly

1. Intuitive understanding can be tricky.

What should you do now?

“Why don't you tell the world - eh!?” *Dr Strangelove*