

## Which of these are TRUE? - The purpose of Dimensionality Reduction is:

1. reduce the number of observations
2. identify a set of fewer and new features to represent the data
3. avoid overfitting
4. saves on computational resources
5. all of the above

# Which of these are TRUE? - The purpose of Dimensionality Reduction is:

1. reduce the number of observations

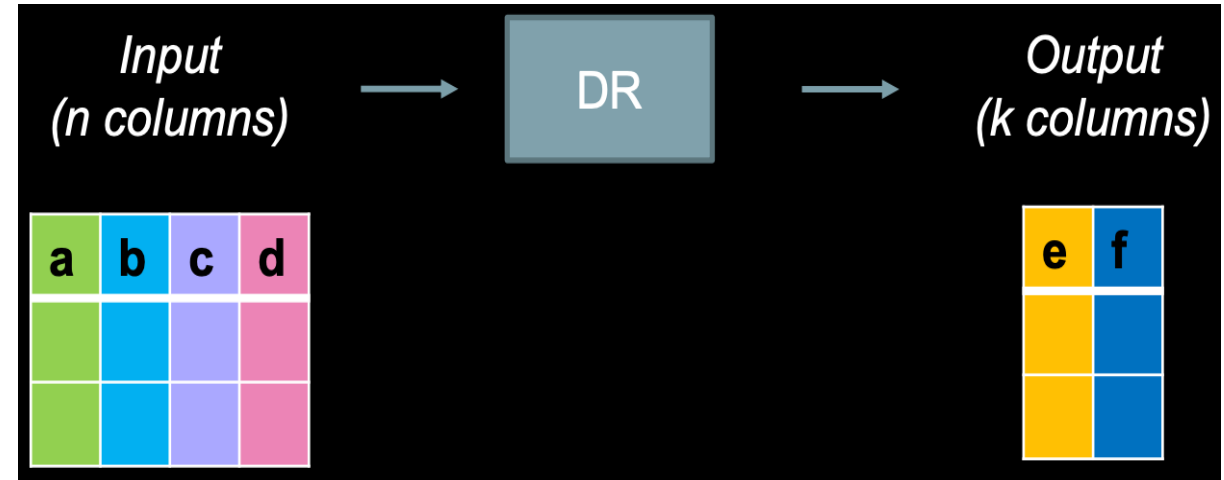
False, DR works on variables (columns) not observations (rows)

2. identify a set of fewer and new features to represent the data **True**

3. avoid overfitting **True**

4. saves on computational resources **True**

5. all of the above **False** – since 1. is False




Which of these does NOT describe the 'Curse of Dimensionality' ?

1. the more features we have, the more data points we need in order to fill the space
2. as dimensions approach infinity, distance between points becomes the same
3. increased dimensions creates more dense data
4. a data set with more variables than observations

# Which of these does NOT describe the 'Curse of Dimensionality' ?

1. the more features we have, the more data points we need in order to fill the space
2. as dimensions approach infinity, distance between points becomes the same
3. increased dimensions creates more dense data **does NOT describe CofD Increased dimensions create more SPARSE data**
4. a data set with more variables than observations

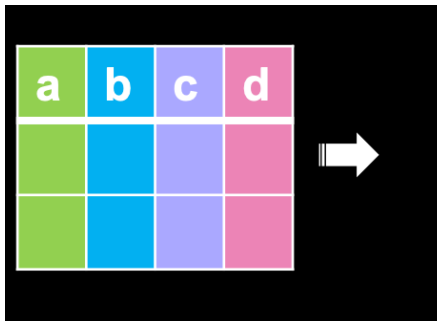
	Red	Maroon	Pink	Flamingo	Blue	Turquoise	Seaweed	Ocean
	1	0	0	0	0	0	0	0
	0	1	0	0	0	0	0	0
	0	0	1	0	0	0	0	0
	0	0	0	1	0	0	0	0
	0	0	0	0	1	0	0	0
	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	1

# Following the steps to run a PCA algorithm, why is it so important to standardise your data?

1. makes the training time faster
2. helps find the features with higher variance
3. standardising allows other people to understand your work
4. it is good practice in data wrangling

# Following the steps to run a PCA algorithm, why is it so important to standardise your data?

1. makes the training time faster
2. helps find the features with higher variance **True**



Name	New dim	combination	Variance
1 <sup>st</sup> component	e	$0.5a + 0.6b + 0.1c + 0.3d$	0.8
2 <sup>nd</sup> component	f	$0.6a + 0.1b + 0.2c + 0.5d$	0.1
	g		0.05
	h		0.05

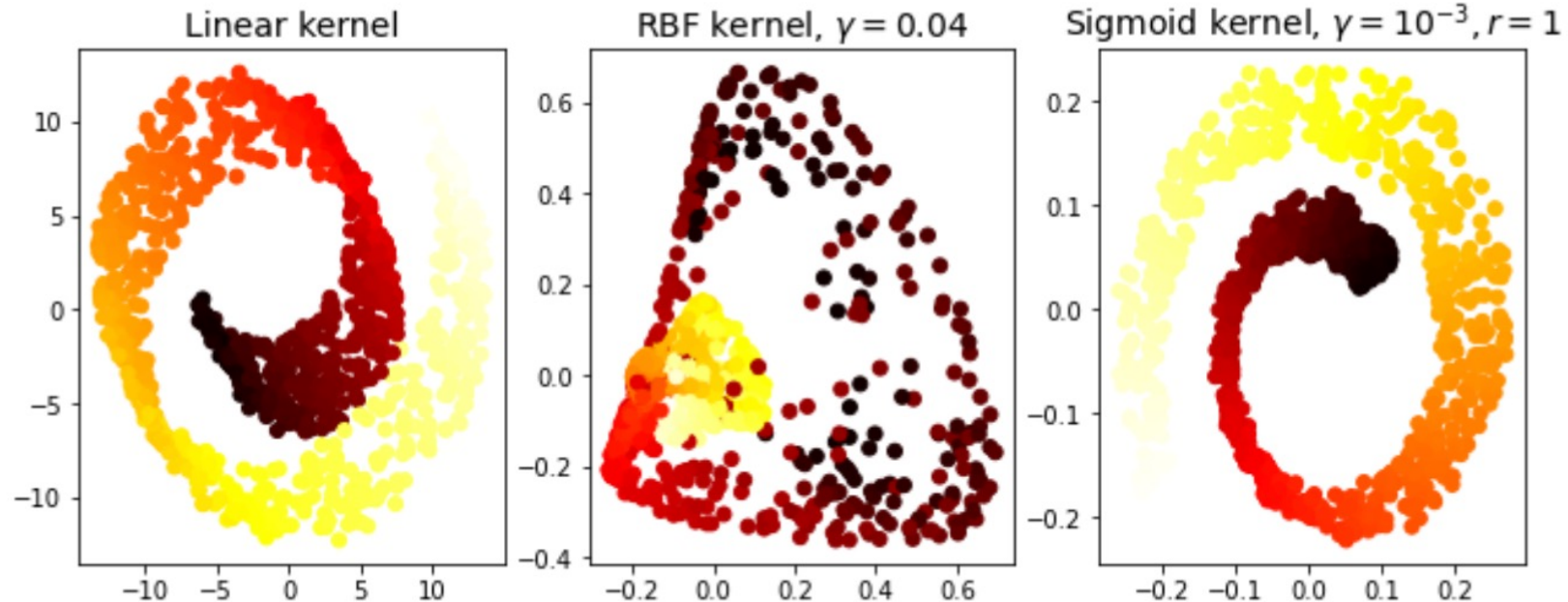
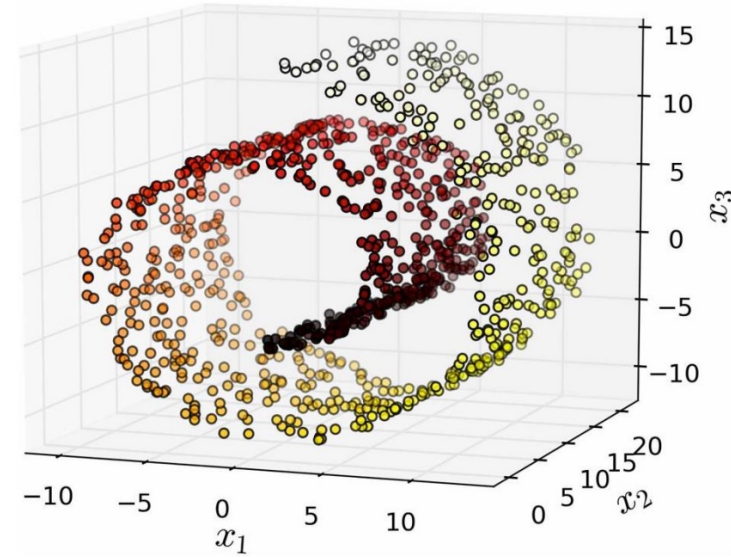
3. standardising allows other people to understand your work
4. it is good practice in data wrangling

## Which of these is TRUE about Kernel PCA?

1. transforms non-linear data into a linearly separable form
2. increases the dimensionality of data
3. uses PCA to reduce the number of features
4. All of the above

# Which of these is TRUE about Kernel PCA?

1. transforms non-linear data into a linearly separable form
2. increases the dimensionality of data
3. uses PCA to reduce the number of features
4. All of the above **True**





# Which of these is FALSE about Locally Linear Embedding?

1. measures how each instance relates to its closest neighbours
2. looks for low-dimensional representation where local relations are best preserved
3. preserves global variance
4. None of the above

# Which of these is FALSE about Locally Linear Embedding?

1. measures how each instance relates to its closest neighbours **True**

Uses a KNN approach to find the k nearest neighbours of every data point.

2. looks for low-dimensional representation where local relations are best preserved **True**

Find the positions of all the points in the new lower-dimensional space

3. preserves global variance **False, LLE preserves LOCAL variance**

4. None of the above **Also False ... since 3. is False**