**Assignment\_23**

1.What are the key reasons for reducing the dimensionality of a dataset? What are the major disadvantages?

* **It may lead to some amount of data loss.**
* **PCA tends to find liner correlations between variables, which is sometimes undesirable.**
* **PCA fails in cases where mean and covariance are not enough to defines datasets.**

2. What is the dimensionality curse?

**Ans: The curse of dimensionality, indicates that the number of samples needed to estimate an arbitrary function with a given level of accuracy grows exponentially with respect to number of input variables of the function.**

3. Tell if its possible to reverse the process of reducing the dimensionality of a dataset? If so, how can you go about doing it? If not, what is the reason?

**Ans: No, dimensionality reduction is not reversible. It loses information.**

4. Can PCA be utilized to reduce the dimensionality of a nonlinear dataset with a lot of variables?

**Ans: PCA can be used to significantly reduce the dimensionality of most datasets, even if they are highly nonlinear because it can at least get rid of useless dimensions.**

5. Assume you're running PCA on a 1,000-dimensional dataset with a 95 percent explained variance ratio. What is the number of dimensions that the resulting dataset would have?

**Ans: hard to say, it depends on dataset.**

6. Will you use vanilla PCA, incremental PCA, randomized PCA, or kernel PCA in which situations?

* **Vanilla PCA: the dataset fit in memory**
* **Incremental PCA: larget dataset that don't fit in memory, online taks**
* **Randomized PCA: considerably reduce dimensionality and the dataset fit the memory.**
* **kenrl PCA: used for nonlinear PCA**

7. How do you assess a dimensionality reduction algorithm's success on your dataset?

Ans:

8. Is it logical to use two different dimensionality reduction algorithms in a chain?

**Ans: Yes, it makes sense to combine two DR methods.**