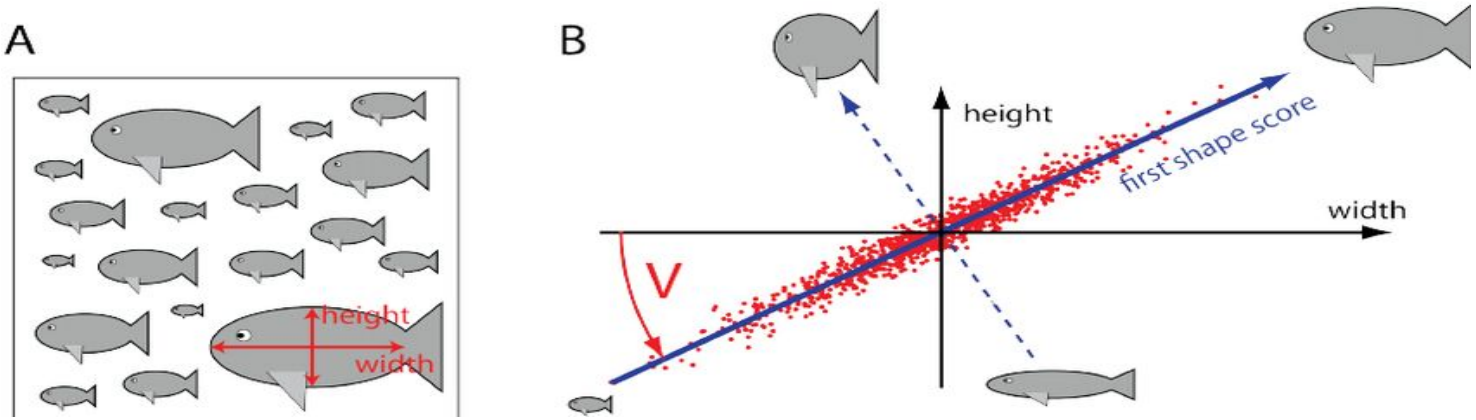
Principal Component Analysis

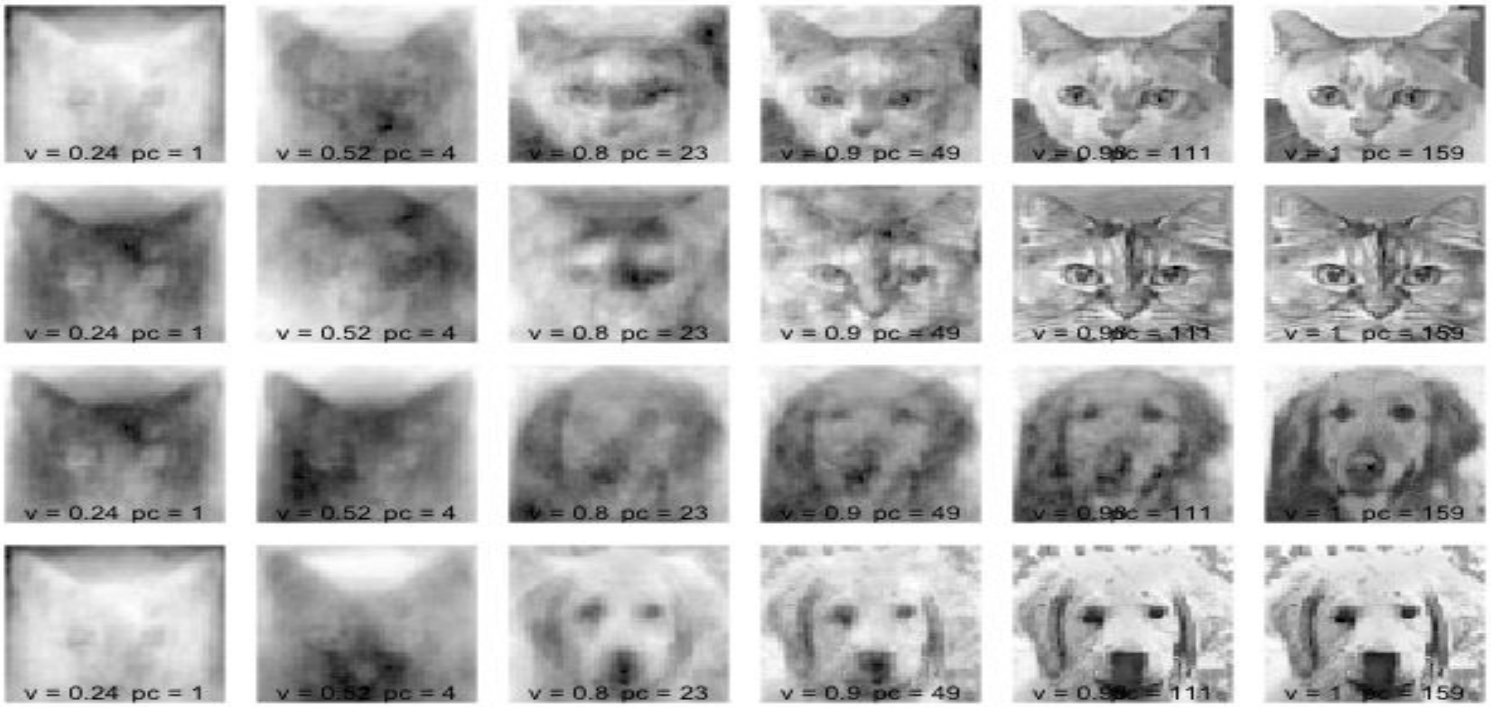
- PCA is a feature extraction method and extracts the most important information.
- This in turn leads to compression because the less important information are discarded. With fewer data points to consider, it becomes simpler to describe and analyse the dataset.
- PCA can be seen a trade-off between faster computation and less memory consumption versus information loss.
- PCA is considered as one of the most useful tools for data analysis.

Example of PCA

- We can describe the shape of a fish with two variables: height and width.
- Given the height, we can probably estimate the width; and vice versa. Thus, we may say that the shape of a fish can be described with a single component.
- This doesn't mean that we simply ignore either height or width. Instead, we transform our two original variables into two orthogonal (independent) components
- The first component (blue line) will explain most of the variation in the data.
- The second component (dotted line) will explain the remaining variation.



Example- Facial Recognition



- PCA can be applied for facial recognition. For 90% capture variance, only a third of the components had to be retained. This may be sufficient for Machine Learning applications.

INTERNSHIPSTUDIO

Computation Of PCA

The below steps need to be followed to perform dimensionality reduction using PCA:

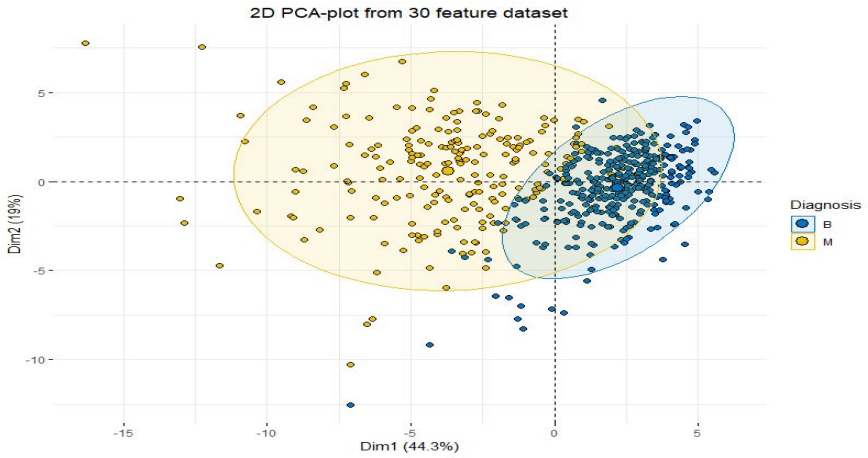
- Standardization of the data
- Computing the covariance matrix
- Calculating the eigenvectors and eigenvalues
- Computing the Principal Components
- Reducing the dimensions of the data set

When should I use PCA?

INTERNSHIPSTUDIO

- Do you want to reduce the number of variables, but aren't able to identify variables to completely remove from consideration?
- Do you want to ensure your variables are independent of one another?
- Are you comfortable making your independent variables less interpretable?

If you answered "yes" to all three questions, then PCA is a good method to use. If you answered "no" to question 3, you **should not** use PCA.



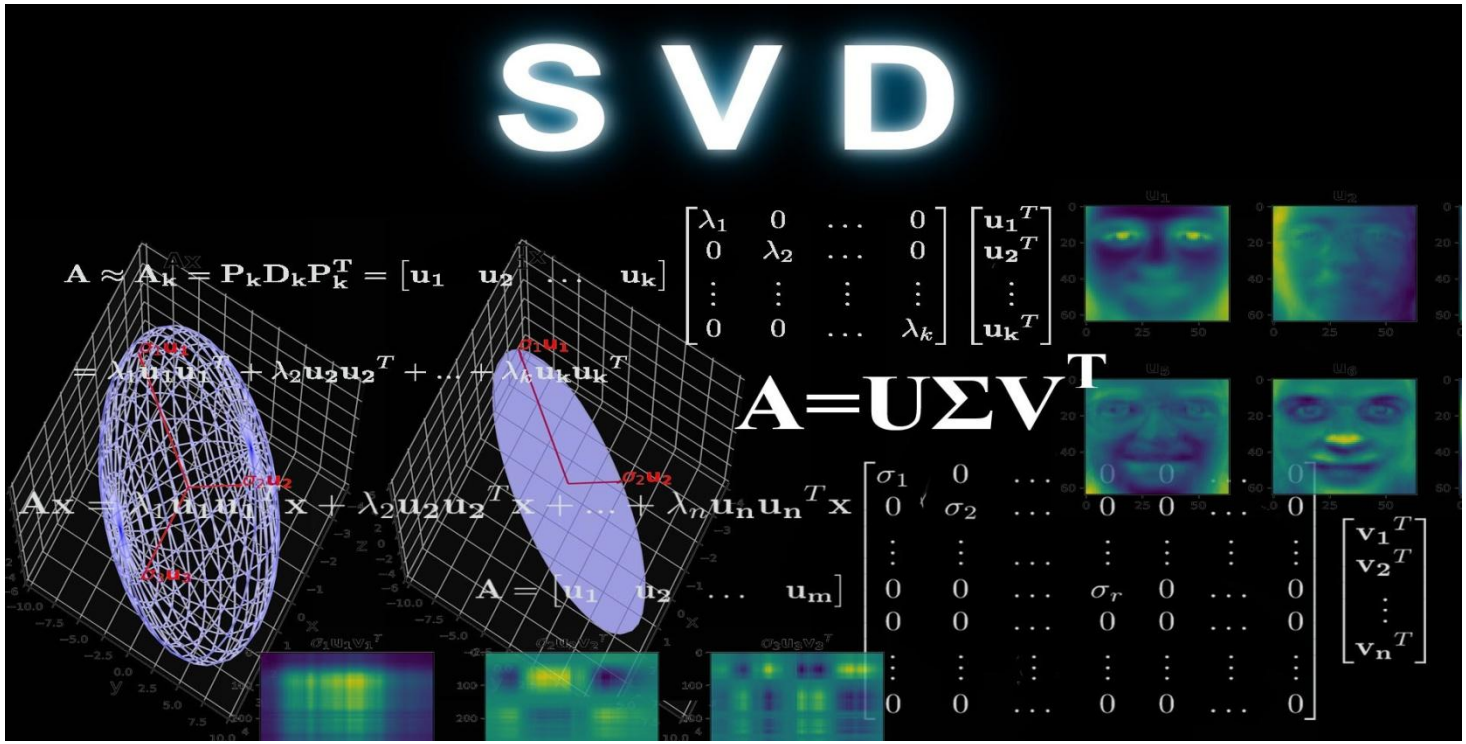
PCA- Advantages/Disadvantages

DISADVANTAGES

- Usefulness depends on the researchers' ability to develop a complete and accurate set of attributes - If important attributes are missed, precision of the procedure is reduced accordingly.
- Naming of the factors (independent variables) can be difficult - multiple attributes can be highly correlated with no apparent reason.
- If the observed variables are completely unrelated, PCA analysis is unable to produce a meaningful pattern.

Singular Value Decomposition

- Singular Value Decomposition (SVD) is another dimensionality reduction technique in data science.
- SVD allows us to extract and untangle information.



INTERNSHIPSTUDIO

- Q.1 What are the properties of KNN?
- Q.2 What do you mean by PCA?
- Q.3 Explain Euclidean Distance?
- Q.4 What do you mean by SVD?
- Q.5 What is the role of PCA?

Thank You