

Aim: Implement the principal component analysis algorithm.

Task:

1. Write a Python Program to perform dimensionality on any common dataset such as Iris dataset and a dataset of any domain with high dimension.
2. Choose a supervised machine learning model, train and test the model by giving
 - a. Dataset without applying PCA on both the datasets
 - b. Apply PCA on both the dataset and then train the model
3. Note the accuracy of the model – with PCA and without PCA
4. Analyse the result. Write in brief a paragraph on your observations.

Submission:

1. Synopsis of Dataset description of high dimensionality
2. Theory of PCA with example
3. Python code for implementation – python file for Common dataset and you dataset.

Output:

1. Variance
2. Accuracy

References:

1. <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
2. <https://www.turing.com/kb/guide-to-principal-component-analysis>

Important Instructions: Total files to be submitted: 4

Synopsis of Dataset (in 1 document), Theory of PCA with example and inferences from the output on your dataset (1 document), Codes – common dataset (1 document) and your dataset (1 document).