Assignment No. 7

Title :-
Implement principal component analysis algorithm

Theory :
Implementing PCA in python with scikit learn

Why PCA)
* When there are many i/p attributes, it is difficult to visualize data, there is a very famous term curse of dimension in ML domain
* Basically, it refers to fact that higher number of attributes in dataset adversely affects accuracy and training time of ML.
* principal component analysis is way to address this issue and is used for better data visualization & improving accuracy

How does PCA work

* PCA is an unsupervised pre-processing task that is carried out before applying any ML algorithm. PCA is based on "orthogonal linear transformation which is a mathematical technique to project attributes of data set onto a new co-ordinate system.
* Similarly attribute which stands second in describing variance is called a second principal component and so on. In short complete dataset can be expressed in terms of principal component.

* Principal component analysis as PCA thus convert data from high dimensional space to low dimensional space by selecting most important attributes that capture maximum about dataset.

Python implementation : 1) To implement PCA in scikit learn it is essential to standardize normalize data before applying PCA
2) PCA is imported from sklearn decomposition. we need to
3) usually n- components is chosen to be 2 for better visualization but it matters & depends on data.
4) By fit and transform method, attributes are passed.
5) The value of principal component can be checked using components while expected variance explained by each principal component can be calculated using explained_variance_ratio.

i) Import all libraries
ii) loading data
iii) Apply PCA
iv) check components
v) plot components
vi) calculate variance ratio

conclusion :

Thus, we have implemented principal component analysis algorithm