

# What is PCA

PCA is a un-supervised Machine Learning Algorithm. PCA is mainly used in dimensionality reduction in a dataset consisting of many variables that are highly co-related or lightly co-related with each other while retaining variations present in the dataset upto a maximum extent. Typically used in EDA and Predictive Analytics

In [ ]:

1

## Working

PCA performs a linear transformation on the data so that most of the variance or information in the dataset is captured by the principal components.

The first principal component will capture the most variance, followed by the second

```
1 In ML Variance is one of the most imp factors that directly affect the accuracy of the output
2 When ML model becomes too sensitive for independent variable it tries to find out the relationship
3 between
4 every feature which gives rise to the problem like 'overfitting' or high variance
5 Too much noise enters the data set because of high variance and thus the results are affected.
```

## Normalization / Scaling

```
1 Normalization is necessary to make every variable in propotion with each other.
2 we have seen in the previous example as how we had scaled (normalize) our iris (training & testing
3 data)
4 we know that models which are not scaled properly in accordance with each other tends to perform poorly
5 in comparison to those scaled well.
```

In [ ]:

1

In [1]:

```
1 import numpy as np
2 import pandas as pd
```

In [2]:

```
1 df = pd.read_csv('Wine.csv')
```

In [3]:

```
1 df.head()
```

Out[3]:

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	Nonflavanoid_Phenols	Proanthocyanins	Color_Intensity	Hue
0	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05
2	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03
3	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86
4	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04



In [ ]:

1

In [ ]:

1

In [ ]:

1

In [ ]:

1