

# Develop a Byte-Size Learning module to help a learner implement Principal Component Analysis In Lord Ganesh Style

By Aviral Bhardwaj

PGD-Big Data Analytics

CDAC Pune

**1-Maharaj, Where Are you, Our Bhakta's  
are puzzled With Principal Component  
Analysis ,Help Them**

2-Aha PCA ,I  
need Modak  
Before  
explaining It.



4-Modak..Modak



3-Lijiye  
Modak.



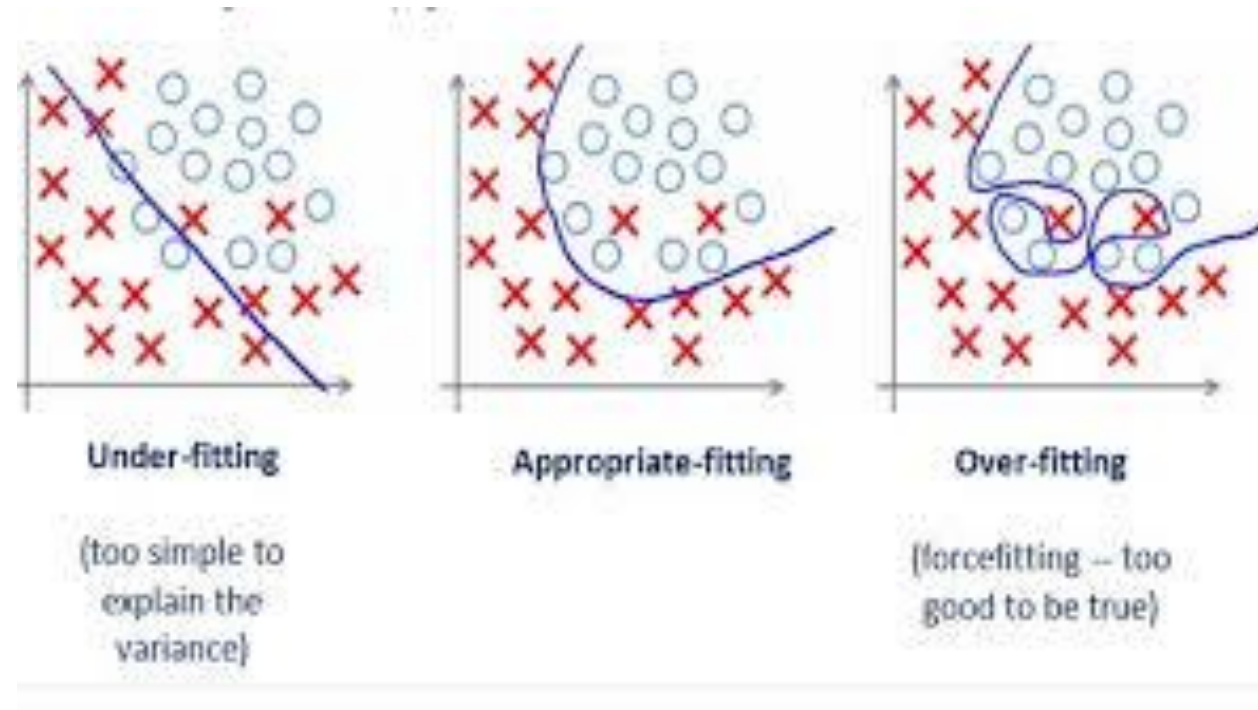
2-Ok Musak Raj ! Now, listen my Bhakta's are doing their machine learning project but most of the time they forgot to tune their ML Algo,so the given data is going to be biased and it is known as Overfitting, then dad gave work to Karl Pearson for discovering PCA for reducing overfitting(dimensions reduction) and the name is known as Principle Component Analysis.

1-Maharaj,Please  
Start from start



3-Still Not Understand Prabhu.

1-Aha please explain More..



4-It's not for you ,it's for the Analytics Vidya Team to select my another Bhakta for their team.

2-Hope you understand the Problem here



Karl also made new term name is “Views” means how to see your data from different angle or analyzing their model differently.  
e.g-3D→2D(PC1) from another angle 2D(PC2) or simply reduce complicity.



Maharaj ,But how to find PC in PCA



Ok.Maharaj



Another Important thing is → Number of PC can be less then or equal to Number of attributes in your dataset for the building the model .PCs are depend upon how many features you have .Ok  
Mushak raj





2-It's very simple Musakraj, if I give you 3 Modak bigger one<less bigger<very small Modak ,then which one you will choose?



4-That's it ,it is as simple as.

1-But Maharaj, if we have more PCs like PC1,PC2,PC3 .. Then which value we should consider?

3-Of course Bigger one Maharaj.



1-It can't be so simple their  
should be catch isn't it  
Maharaj



2-Aha Now you are getting things another  
property is orthogonal property means PC1 and  
PC2 should not be dependent each other .They  
should be totally independent.





Ok if we have X and Y variable  
in dataset the first we will take  
their mean known as X(mean)  
and Y (mean)



X	Y
3	6
6	5
X(mean)	Y( mean)

Maharaj  
Numerical  
Example  
Please



$$\text{Covariance}(x,y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

1-That's it  
Maharaj



4-Still  
thinking  
which x,y I  
will follow

3-Where  $x_1$  and  $y_1$  is first row  
value and  $\bar{x}$ ,  $\bar{y}$  is means of  
the dataset and  $n$  is number of  
rows calculation

2-No ,my pal after  
that we have another  
concept known as  
Covariance



$$C = \begin{bmatrix} \text{Cov}(x, x) & \text{Cov}(x, y) \\ \text{Cov}(y, x) & \text{Con}(y, y) \end{bmatrix}$$

2-we have to  
calculate covariance  
of above MATRIX



1-Still thinking which  
 $x, y$  I will follow



I want to  
learn by  
numerical  
example



For (X,X),  $X(\text{mean})=4.5$  and  $n=2$  then  $n-1=1$

X	$X-X(\text{mean})$	$(X-X(\text{mean}))^*$ $(X-X(\text{mean}))$
3	-0.75	0.5625
6	1.5	2.25

Sum=2.812

For (Y,Y),  $Y(\text{mean})=5.5$  and  $n=2$  then  $n-1=1$

Y	$Y-Y(\text{mean})$	$(Y-Y(\text{mean}))^*$ $(Y-Y(\text{mean}))$
3	-0.75	0.5625
6	1.5	2.25

Sum=0.5

Ok .As you  
wish



Now for (x,y) and (y , x)  $n=2$  then  $n-1=1$

X	Y	$A=X-X(\text{mean})$	$B=Y-Y(\text{mean})$	$A*B$
3	6	-0.75	0.5	-0.375
6	5	1.5	-0.5	-0.75

Now for (x,y) sum= -1.125

Now for (y , x) sum = -1.125



# Now Put them into matrix and find the Eigen value and Eigen vector

- [|C-Lemda I|=0](#)

2.812	-1.125
-1.125	0.5

- [And find the Eigen value and vector from below link](#)
- <https://www.emathhelp.net/calculators/linear-algebra/eigenvalue-and-eigenvector-calculator/>



## ANSWER

$$\text{Eigenvalue: } -\frac{\sqrt{2601961}}{1000} + \frac{207}{125}, \text{ eigenvector: } \begin{bmatrix} -\frac{1156}{1125} + \frac{\sqrt{2601961}}{1125} \\ 1 \end{bmatrix}$$

$$\text{Eigenvalue: } \frac{\sqrt{2601961}}{1000} + \frac{207}{125}, \text{ eigenvector: } \begin{bmatrix} -\frac{\sqrt{2601961}}{1125} - \frac{1156}{1125} \\ 1 \end{bmatrix}$$

2-Right Question ,Now which one have greater Eigen value then it's Eigen vector Will be our Principle Component and the whole procedure is known as Principle Component Analysis. That's it ..Simple isn't it?

1-Now Maharaj we Got 2 Eigen value and Eigen Vector, which one will be my PC??



Thank you Musakraj, Anyone Can write Code but not everyone can understand, I will tell him to Write A code also he will submit both, And I believe that He will selected.

And by Chance If Not then Dad Must have different Plan for him.

Maharaj You taught me a lot,I learned Whole PCA In just 16 Slides. Awesome

But I am thinking you invested lot of time for your Bhakta Aviral Bhardwaj ,but what if he will not get selected because he didn't write single line of code yet?



# Thank You Analytics Vidya Team For This Opportunity

- Name-Aviral Bhardwaj
- CDAC Pune
- Big Data Analytics(Completed)

