# Principal Component Analysis Interview Questions

We believe that you have learned both theoritical and practical knowledge on Naive Bayes classification algorithm through your assignment.

So let's test your knowledge here. This will help you to be prepared for interviews too!

## Best with Quest

#### 1. Why do we need dimensionality reduction? What are its drawbacks?

```
In Machine Learning, dimension refers to the number of features. Dimensionality reduction
is simply, the process of reducing the dimension of your feature set.
Advantages of Dimensionality Reduction:
-Less misleading data means model accuracy improves.
-Fewer dimensions mean less computing. Less data means that algorithms train faster.
-Less data means less storage space required.
-Removes redundant features and noise.
-Dimensionality Reduction helps us visualize the data on 2D plots or 3D plots.

Drawbacks of Dimensionality Reduction are:
-Some information is lost, possibly degrading the performance of subsequent training
algorithms.
-It can be computationally intensive.
--Transformed features are often hard to interpret.
-It makes the independent variables less interpretable.
```

#### *2. Explain the Curse of Dimensionality?*

```
The curse of dimensionality refers to all the problems that arise working with data in the
higher dimensions. As the number of features increase, the number of samples increases,
hence, the model becomes more complex. The more the number of features, the more the
chances of overfitting. A machine learning model that is trained on a large number of
features, gets increasingly dependent on the data it was trained on and in turn
overfitted, resulting in poor performance on real data, beating the purpose. The fewer
features our training data has, the lesser assumptions our model makes and the simpler it
will be.
```

#### 3. What does a PCA do? How is the first principal component axis selected?

```
PCA stands for principal component analysis. It is a dimensionality reduction technique
which summarizes a large set of correlated variables (basically a high dimensional data)
into a smaller number of representative variables, called the principal components, that
explains most of the variability in the original set.
The first principal component axis is selected in a way such that it explains most of the
variation in the data and is closest to all n observations.
```

#### 4. Limitations of PCA?

1.Doesn’t work well for non linearly correlated data.
2. PCA always finds orthogonal principal components. Sometimes, our data demands non-orthogonal principal components to represent the data.
3. PCA always considered the low variance components in the data as noise and recommend us to throw away those components. But, sometimes those components play a major role in a supervised learning task.
4. If the variables are correlated, PCA can achieve dimension reduction. If not, PCA just orders them according to their variances

#### *5. Is it important to standardize before applying PCA?*

PCA finds new directions based on the covariance matrix of original variables. Since the covariance matrix is sensitive to the standardization of variables. Usually, we do standardization to assign equal weights to all the variables. If we use features of different scales, we get misleading directions. But, it is not necessary to standardize the variables, if all the variables are on the same scale.

#### 6. Should one remove highly correlated variables before doing PCA? What will happen when eigenvalues are roughly equal?

No, PCA loads out all highly correlated variables on the same Principal Component(Eigenvector), not different ones.

If all eigenvectors are same then PCA won’t be able to select the principal components because in that case, all principal components are equal.

#### 7. How can you evaluate the performance of a dimensionality reduction algorithm on your dataset?

Intuitively, a dimensionality reduction algorithm performs well if it eliminates a lot of dimensions from the dataset without losing too much information. Alternatively, if you are using dimensionality reduction as a preprocessing step before another Machine Learning algorithm (e.g., a Random Forest classifier), then you can simply measure the performance of that second algorithm; if dimensionality reduction did not lose too much information, then the algorithm should perform just as well as when using the original dataset.

#### *8. Can we use PCA for feature selection?*

No! PCA is not a feature selection technique because if you think, any principal component axis is a linear combination of all the original set of feature variables which defines a new set of axes that explain most of the variability in the data. So while it performs well in many practical settings, it does not result in the development of a model that relies upon a small set of the original features.

#### *9. Can we use PCA for regression? When is it advisable to use PCA for regression?*

Yes, we can use principal components for regression setup. PCA would perform well in cases when the first few principal components are sufficient to capture most of the variation in the predictors as well as the relationship with the response. The only drawback to this approach is that the new reduced set of features would be modeled ignoring the response variable Y when applying a PCA and while these features may do a good overall job of explaining the variation in X, the model will perform poorly, if these variables don’t explain the variation in Y.

***10. What do the coefficients of a principal component tell you? What does a principal component in a PCA represent?***

If we project all the points on the principal component, they tell us that variable 2 is N times as important as variable 1

Principal component represents a line or an axis along which the data varies the most and it also is the line that is closest to all n observations. OR
It is the linear combination of observed variables that results in an axis or a set of axes, that explain/s most of the variability in the dataset.
Mathematically speaking, it is the eigenvector of the first principal component. The sum of the squared distances is the eigenvalue for PC1 and the square root of the eigenvalue is the singular value for the PC1

# Now Rest with this Quest :)