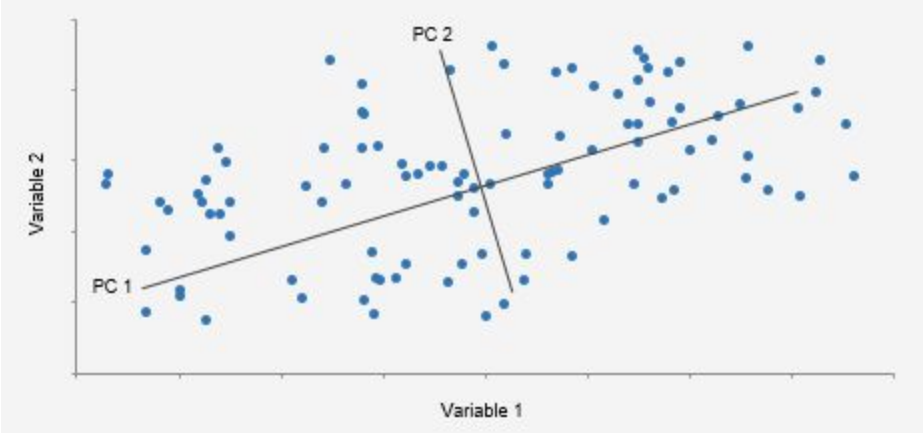
1. What does a PCA do? How is the first principal component axis selected?

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.



2. What does a principal component in a PCA represent?

It 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

3. What do the coefficients of a principal component tell you?

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

4. 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.

5. 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.**

6. What are some caveats of PCA?

Make sure data is on the same scale

Make sure data is centered

The maximum number of principal components equals the number of variables. There is a principal component or PC for each variable in the data set. However, if there are fewer samples than variables, the number of samples puts an upper bound on the number of PCs with eigenvalues greater than 0