**Dimensionality Reduction – PCA vs LDA**

* While working on A problem of Multi-Class classification, I observed following behaviour of PCA.

*With increase in variance coverage for PCA (taking more and more Principal components into consideration for model building), Classification Accuracy was also increasing.*

Ex. 98 % of variance covering PCA was giving me, 57 % Classification Accuracy.

100 % variance covering PCA was giving me, around 60% Classification Accuracy.

And, this behaviour triggered a question to me about PCA and its importance in Machine Learning.

* **If 100% variance covering PCA is giving me best possible Classification Accuracy, why would I go for 98 % variance covering PCA?**

What is the significance of Dimensionality Reduction? Why to use Principal Component Analysis at the cost of classification accuracy?

To understand PCA and clarify my doubts, I did research on this topic and following are my findings:

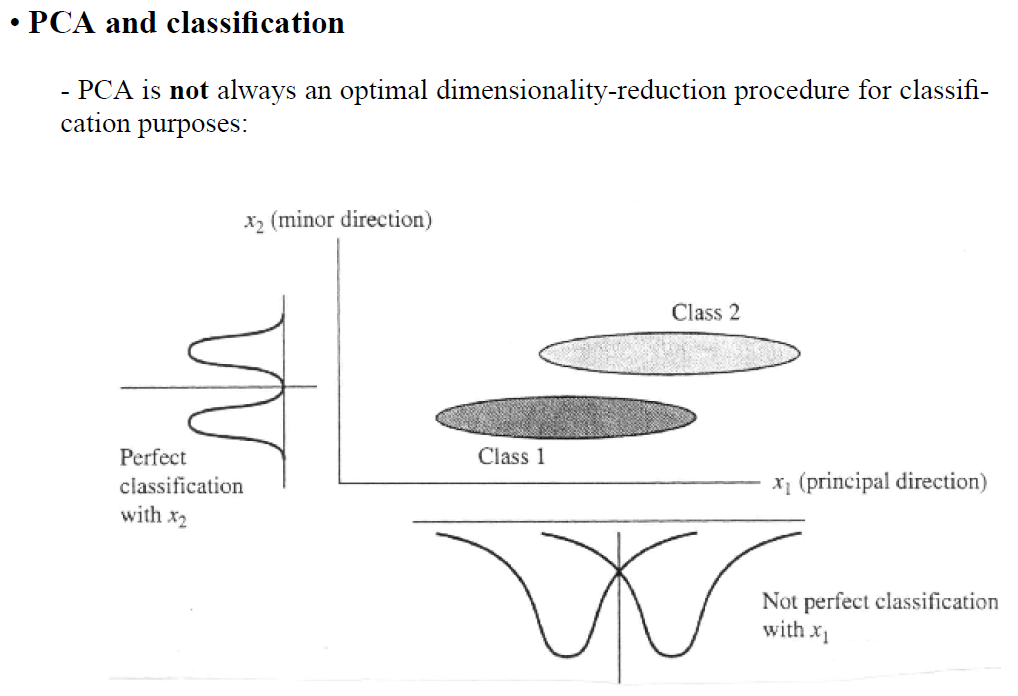
1. **Principal Components do not necessarily have any correlation to classification accuracy.**

There could be a 2-variable situation where 99% of the variance corresponds to the first PC but that PC has no relation to the underlying classes in the data. Whereas the second PC (which only contributes to 1% of the variance) is the one that can separate the classes. If we only keep the first PC, then we lose the feature that provides the ability to classify the data. In practice, smaller (lower variance) PCs often are associated with noise so there can be benefit in removing them, but there is no guarantee of this.

Example: Consider we have two variables: a person's weight (in kilograms) and body temperature (in degrees Celsius). And, we want to predict which person has the flu and which do not.

In this case, weight has a much greater variance (i.e. its value can vary between a large range) but probably no correlation to the labelled variable flu, whereas temperature, which has low variance (i.e. its value can vary between a smaller range usually 98-104 degree Celsius), has a strong correlation to the labelled variable flu. After the Principal Components transformation, the first PC will be strongly aligned with weight (since it has much greater variance) so if we drop second PC, we would be losing almost all our classification accuracy.

It is important to remember that Principal Components is an **unsupervised transformation** of the data. It does not consider **labels** of your training data when calculating the transformation (as opposed to **Fisher's Linear Discriminant**).



The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original dataset. PCA finds a vector that **"best represents"** a data set in a much lower dimension.

In contrast to that, **Linear Discriminant Analysis** (LDA) finds a vector that **"best discriminates"** between our data set classes. PCA only looks into the dimensions, not into categories. And, hence we can say that PCA is a dimensionality reduction technique, useful to choose top k principal components, but not advisable to be used for classification purpose.

For Classification challenges, it is preferred to use Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM). And, For Regression challenges, PCA will be a preferred option.

1. **There is always a Trade-Off between Classification Accuracy and Model’s Training Cost.**

Dimensionality Reduction may help in improvement of the accuracy of the predictive models. This is possible because features selection enables us to reduce the noise from the data and select the most useful features to be used during the training session. This greatly helps the model to become less overfit to the noises from the training data. The model will have a better generalization ability when tested with unseen data points.

However, sometimes features selection also results in the loss of predictive accuracy as it happened in my case.

Example: 100 PC features = 95% accuracy

50 PC features = 92% accuracy

based on the difference between 2 results, we can say that removing 50 less significant features from a total of 100 features gives us 92% accuracy.

Features selection reduced the dimensionality of the predictive model. And, As the dimensionality of the model decreases, the complexity of the model also decreases. This eventually leads to a faster model training time and convergence.

Here, we achieve great reduction in model complexity, training time and overfitting with increase in the generalization ability by sacrificing 3% of predictive accuracy.

1. **Significance of Number of Dimensions of Dataset**

Effect of Dimensionality Reduction techniques like PCA or LDA on Machine Learning problems, can be effectively seen when Number of Dimensions of Dataset is high.

On dataset with small number of dimensions, dimensionality reduction sparsely gives positive and useful results.

**References:**

1. <https://stackoverflow.com/questions/30779009/why-does-classifier-accuracy-drop-after-pca-even-though-99-of-the-total-varian>
2. <https://www.quora.com/What-is-the-motivation-around-a-PCA-Is-it-computational-or-is-it-to-increase-accuracy>
3. <https://stats.stackexchange.com/questions/244677/how-to-decide-between-pca-and-logistic-regression/244680#244680>
4. <https://www.researchgate.net/post/PCA-consider_all_principal_components>
5. <http://www.cs.princeton.edu/courses/archive/fall08/cos429/CourseMaterials/lecture2/PCA_handout.pdf>
6. <https://www.quora.com/Can-feature-selection-improve-accuracy-or-only-reduce-training-time>