

PDML-Assignment#4 Report

Abdelrahman Matar

SID: 900123206

Date: May 13, 2017

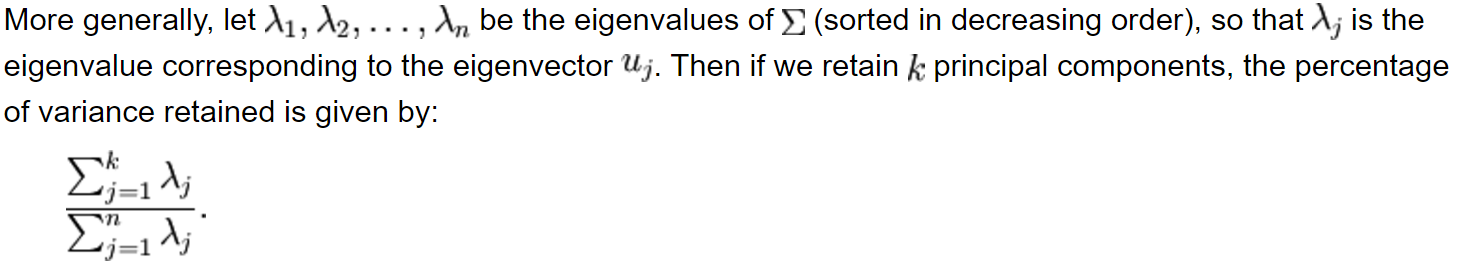


## Report of the results and analysis

1. Cross-validation accuracy of your classifier for each fold and the average cross-validation accuracy without PCA. **[1%]**

|  |  |
| --- | --- |
| Fold number | Accuracy |
| 1 | 75.539% |
| 2 | 79.136% |
| 3 | 77.536% |
| 4 | 77.536% |
| 5 | 78.102% |
| Average cross-validation accuracy = 78.0 +/- 2.0% | |

1. Plot of PCA kept variance in the reduced dimensionality data (yaxis) versus number of used eigenvectors (x\_axis) **[1%]**



I used the previous formula to calculate the variance:

|  |  |
| --- | --- |
| **Num. of used eigenvectors** | **variance** |
| 1 | 0.8885 |
| 2 | 0.9501 |
| 3 | 0.9759 |
| 4 | 0.989 |
| 5 | 0.9965 |
| 6 | 0.9995 |
| 7 | 1 |
| 8 | 1 |

1. Plot of cross-validation accuracy (y-axis) versus number of used eigenvectors (x\_axis) **[2%]**

|  |  |
| --- | --- |
| **Num. of used eigenvectors** | **Avg. cross-valid accuracy (+ std)** |
| 1 | 0.65 (+/- 0.03) |
| 2 | 0.72 (+/- 0.04) |
| 3 | 0.73 (+/- 0.03) |
| 4 | 0.73 (+/- 0.03) |
| 5 | 0.73 (+/- 0.03) |
| 6 | 0.74 (+/- 0.04) |
| 7 | 0.74 (+/- 0.05) |
| 8 | 0.74 (+/- 0.04) |

1. Did PCA help in enhancing the classification accuracy? Comment on either cases. **[1%]**

No it did not help.

Comment: PCA based classification gave less accuracy and this is expected because of the loss of information introduced when using PCA.

\* Please note that I used the following sources for parts of the code:

* I have use used the dataset provided by [Kaggle](https://www.kaggle.com/dssariya/pima-indians-diabetes-data-set) and formatted as csv file.
* I forked this [notebook](https://www.kaggle.com/aconsapart/pima-xgboost-tutorial) from Kaggle and built on top of it for the classifier part.