CS7602 - MACHINE LEARNING

ASSIGNMENT 2

## SUBMITTED BY

JAYASREE LAKSHMI NARAYAN 2016103033

AKHILA G P 2016103503

DATE : 28-03-2019

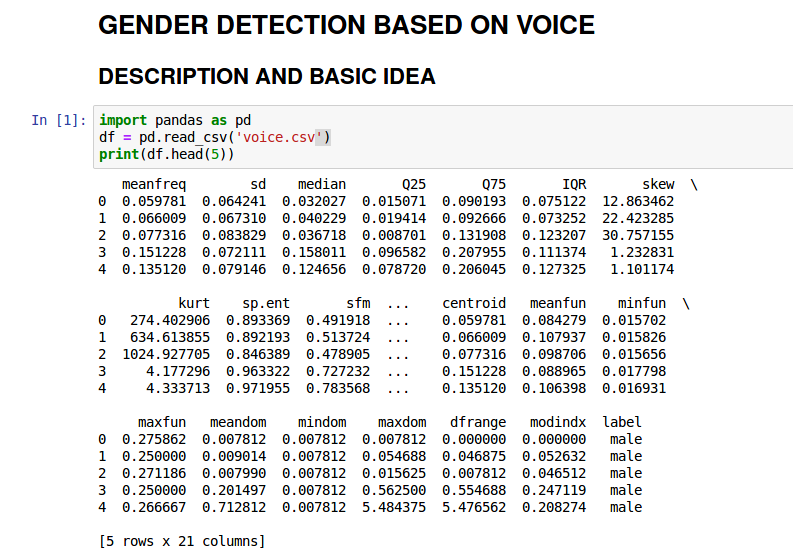
### CONTENTS

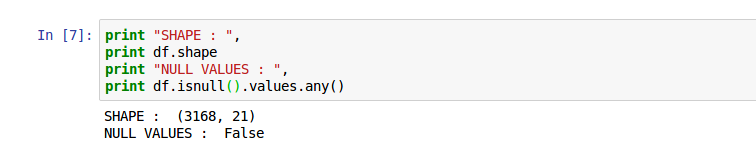
1. SUPPORT VECTOR MACHINES
2. PRINCIPAL COMPONENT ANALYSIS

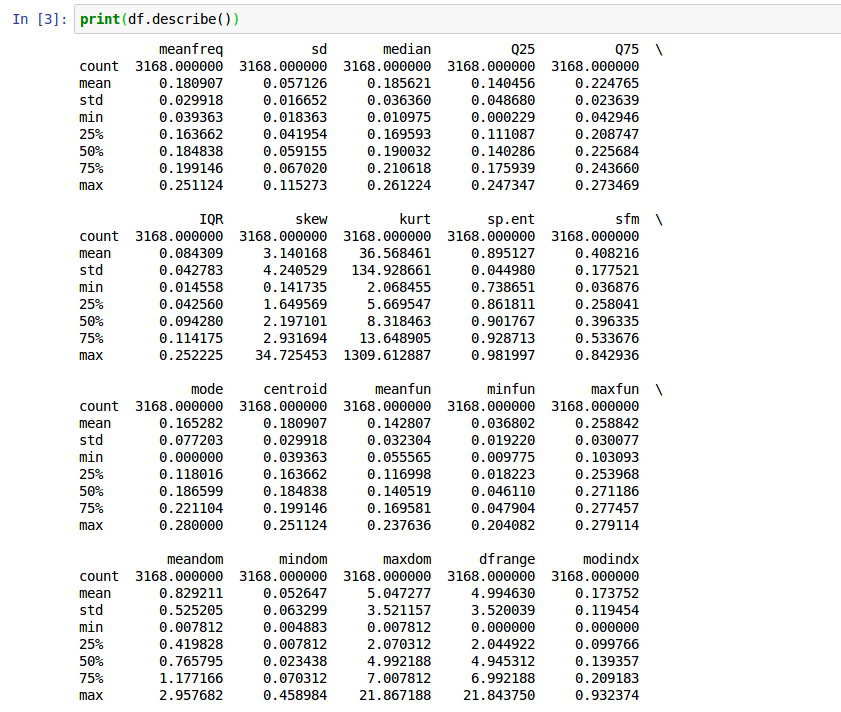
### DATASET USED

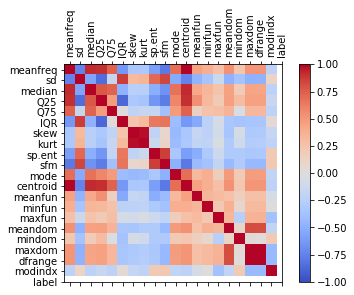
1. GENDER CLASSIFICATION BASED ON VOICE

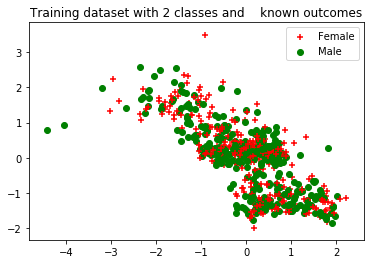
# A DESCRIPTION ON THE DATASET UNDER STUDY







  
Illustration 1: Correlation matrix

  
Illustration 2: Original data distribution

The jupyter notebook with the code is uploaded in Github and the link for the document is <https://github.com/Akhilagp/ML_Assignment>.

**PROCEDURE:**

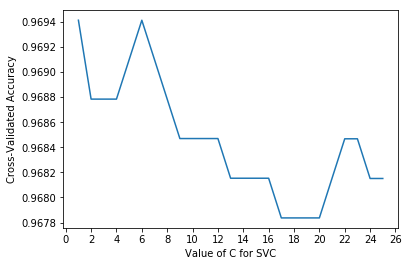
* Support Vectors are the co-ordinates of individual observation. SVM is a frontier which best segregates the two classes (hyper-plane/ line).
* The hyper-plane is selected in such a way that it segregates the two classes better.
* When all the hyper-planes fit well, the best one for classification is chosen by maximizing the distance between nearest data points.
* Kernel trick is used when non linear hyper-planes are needed.
* PARAMETERS VARIED For Understanding

1. Various Kernels
2. Gamma value ( tuning )
3. The ‘C’ parameter – soft margin cost function ( tuning )
4. Degree of polynomial kernel.
5. The number of Principal Components

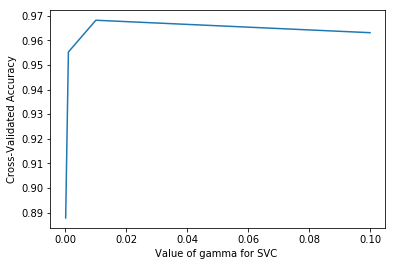
**OUTPUT:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Classifer** | **Accuracy**  **with default parameters** | **Hyper-parameters tuning** | |
| **C** | **Accuracy** |
| **SVC with Linear kernel** | 0.9694 | 0.1 | 0.9700 |
| 0.2 | 0.9691 |
| 0.3 | 0.9690 |
| 0.4 | 0.9690 |
| 0.5 | 0.9694 |
| **SVC with RBF kernel** | 0.9659 | **Gamma** | **Accuracy** |
| 0.01 | 0.96815 |
| 0.02 | 0.9678 |
| 0.03 | 0.9678 |
| 0.04 | 0.9668 |
| 0.05 | 0.9659 |
| **SVC with polynomial kernel** | 0.9457 | **Degree** | **Accuracy** |
| 2 | 0.8506 |
| 3 | 0.9457 |
| 4 | 0.8312 |
| 5 | 0.8659 |
| 6 | 0.7747 |

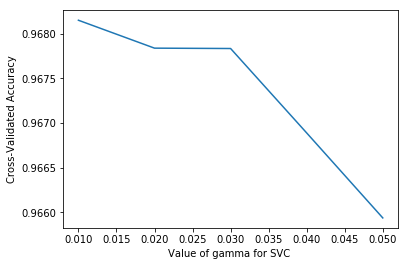
**INFERENCE:**

  
Illustration 3: Tuning soft-cost parameter

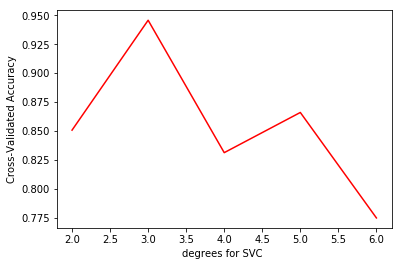
From illustration 1, it is evident that around C=0-1 and C=7-8, the accuracy hits the peak and at C=2 it falls. Varying the the values for C between 0 and 1, the exact value of C where the accuracy is the highest is found (C=0.1)

  
Illustration 4: Tuning gamma value

From illustration 2, around gamma = 0.0 and 0.02, the accuracy increases and starts to fall after 0.02. In turn, tuning the parameter with values between 0 and 0.02 the highest accuracy is obtained at gamma = 0.01.

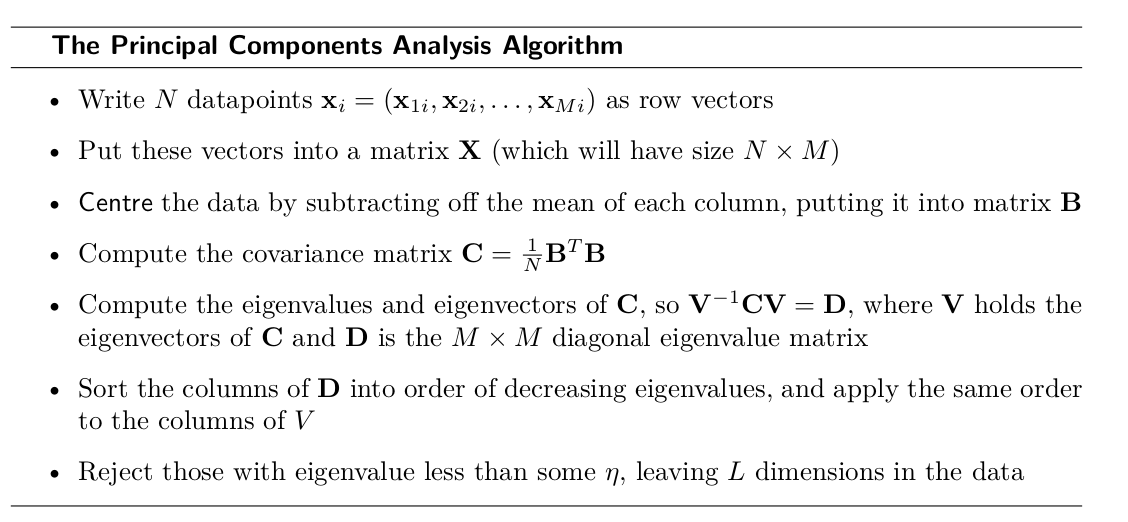
  
Illustration 5: Finest tuning

From illustration 4, it is evident that the 3rd degree polynomial ( default value ) gives the maximum accuracy.

  
Illustration 6: Degree tuning

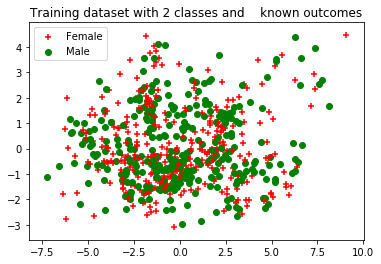
**PRINICIPAL COMPONENT ANALYSIS**

**ALGORITHM**

****

**OUTPUT**

|  |  |  |
| --- | --- | --- |
| **Classifer** | **PCA** | |
| **No. of Principal Components** | **Accuracy** |
| **SVC with Linear kernel** | 6 | 0.88328 |
| 8 | 0.9526 |
| 10 | 0.9763 |
| 12 | 0.9842 |
| 14 | 0.9668 |
| **SVC with RBF kernel** | 6 | 0.8880 |
| 8 | 0.9258 |
| 10 | 0.9495 |
| 12 | 0.9637 |
| 14 | 0.9558 |
| **SVC with polynomial kernel** | 6 | 0.9274 |
| 8 | 0.9479 |
| 10 | 0.9558 |
| 12 | 0.9495 |
| 14 | 0.9463 |



**INFERENCE:**

* For various kernels, the number of principal components are varied, and the highest accuracy is found.
* For Linear SVM, when using PCA and reducing the dimensionality, the accuracy has increased by 1% ( Number of PC = 12 ).
* For polynomial SVM, when using PCA and reducing the dimensionality, the accuracy has increased by 2% ( Number of PC = 10 ).