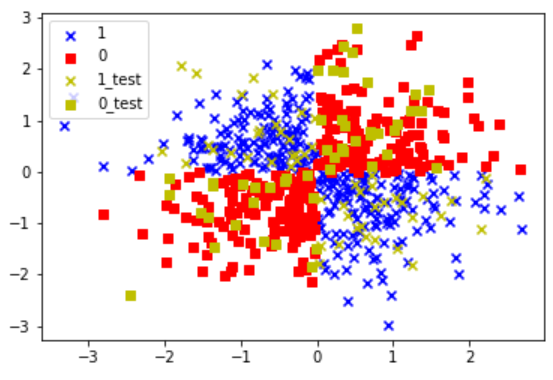
Q. Try out different kinds of SVM Kernel Functions on non-linearly separable datasets. You may create your own datasets that are non-linearly separable, i.e., can be separated by a non-linear structure, and see which all Kernels can be suitable to classify them.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Non-linear datasets are the ones that can not be separated on the basis of linear classification. The classification plane either is some nonlinear curve or exists in various planes. The dataset created had a definite relation of xor logic for simple depiction.

Train samples: 550

Test samples: 100



Various SVM kernels had been experimented with, along with the tunning of their hyperparameters. The various arguments of SVM:

kernel: RBF (default), poly or sigmoid, your own kernel.

C: this is the regularization parameter described in the Tuning Parameters section

gamma: this was also described in the Tuning Parameters section

degree: it is used only if the chosen kernel is poly and sets the degree of the polynomial

* Linear SVM:
  + Train: 62.36
  + Test: 67
* Sigmoid SVM:
  + Train: 51.82
  + Test: 55
* Polynomial:

|  | degree 1 | **degree 2** | degree 3 | degree 4 | degree 5 | degree 6 | degree 7 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| train | 62.36 | **98.73** | 52.91 | 89.27 | 52.73 | 83.82 | 51.27 |
| test | 66 | **99** | 51 | 88 | 52 | 83 | 46 |

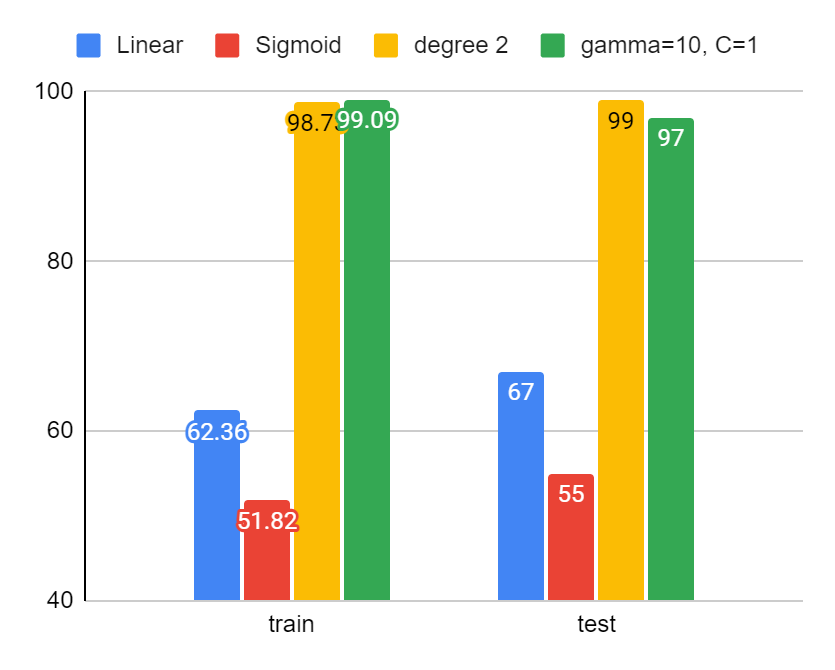
* RBF:

Iterating for gamma

| gamma | 1/100 | 1 | **10** | 100 | 1000 |
| --- | --- | --- | --- | --- | --- |
| C | 1 | 1 | **1** | 1 | 1 |
| train | 50.18 | 97.82 | **99.09** | 99.82 | 100 |
| test | 51 | 95 | **97** | 96 | 73 |

Iterating for C

| gamma | 10 | **10** | 10 | 10 |
| --- | --- | --- | --- | --- |
| C | 1/10 | **1** | 10 | 100 |
| train | 50.18 | **99.09** | 99.45 | 100 |
| test | 51 | **97** | 96 | 96 |



link to tabulated results and graphs

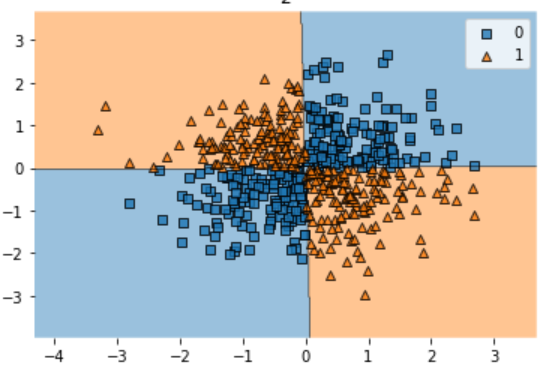
<https://docs.google.com/spreadsheets/d/1FouA17WZe1dyi7TTi-dSFHSZ2rw7y52jRmwei44A8PA/edit#gid=0>

After experimenting on all the various types of kernels and tunning their hyperparameters it’s observed that a polynomial kernel of degree 2 serves as the best fit which provides great performance on the train as well as test dataset. It fits generlistically with low bias in contrast to the next best performing kernel of RBF with gamma=10, C=1, which slightly overfits the training dataset but serves well on the test set.

1. poly-deg 2

accuracy on train set 98.73 %

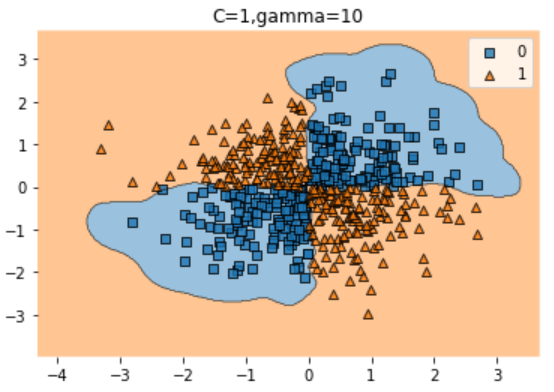
accuracy on the test set 99.0 %



1. RBF kernel-gamma=10, C=1

accuracy on train set 99.09 %

accuracy on the test set 97.0 %



NOTE: that due to the small size of the training dataset and the definite relation between the prediction and features, the model tends to overfit the training samples as observed in RBF kernel with gamma as 10 and C as 100 where the accuracy of 100% is achieved in the train set. In a practical situation, such direct relations are not found and anomalies are even present which makes model training harder.