1. **Logistic Regression:**

1.2.1 Behavior of the algorithm when when there is no regularization:

The behavior of the logistic regression with no regularization tries to fit the training set as much as possible, given the constraints and thus will have big values for w and small classification errors.

When there is no regularization, the weights and error counts on training and validation sets are:

Linearly Separable:

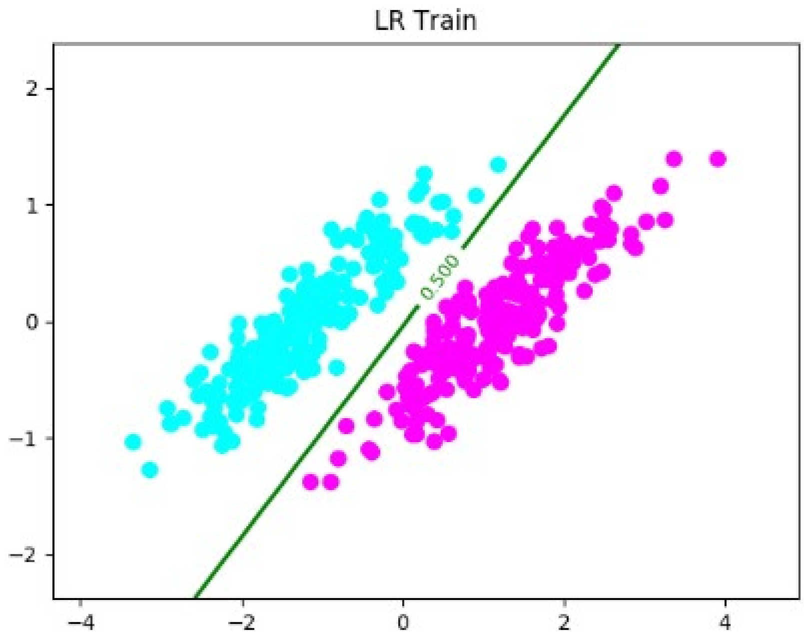
Weights : w0 = 2.59665133

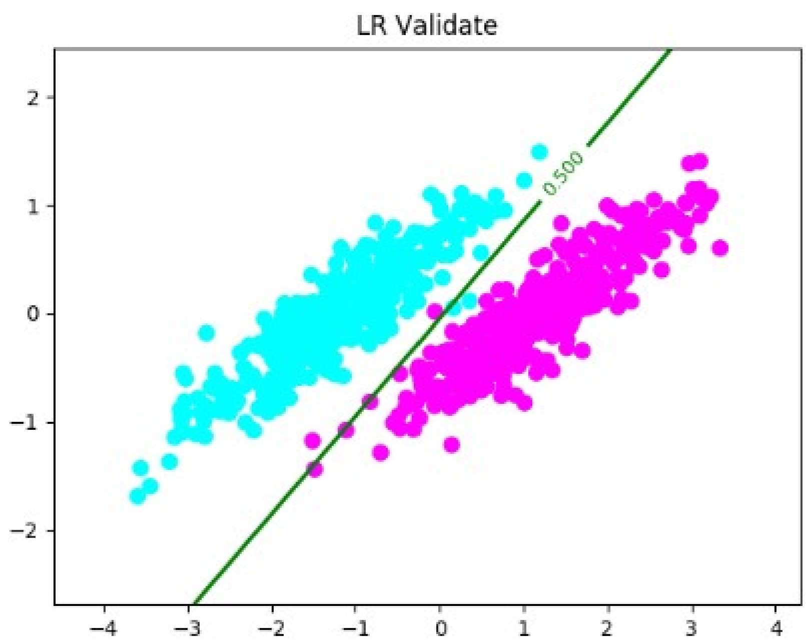
w1 = -47.68478956

w2 = 52.84677869

Training set error count : 0

Validation set error count : 4





Non Linearly Separable:

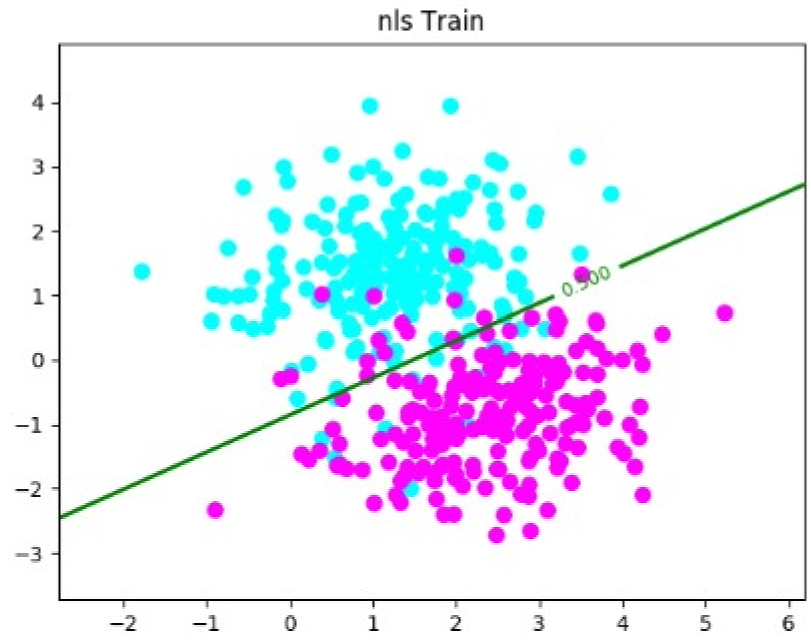
Weights : w0 = 2.09378157

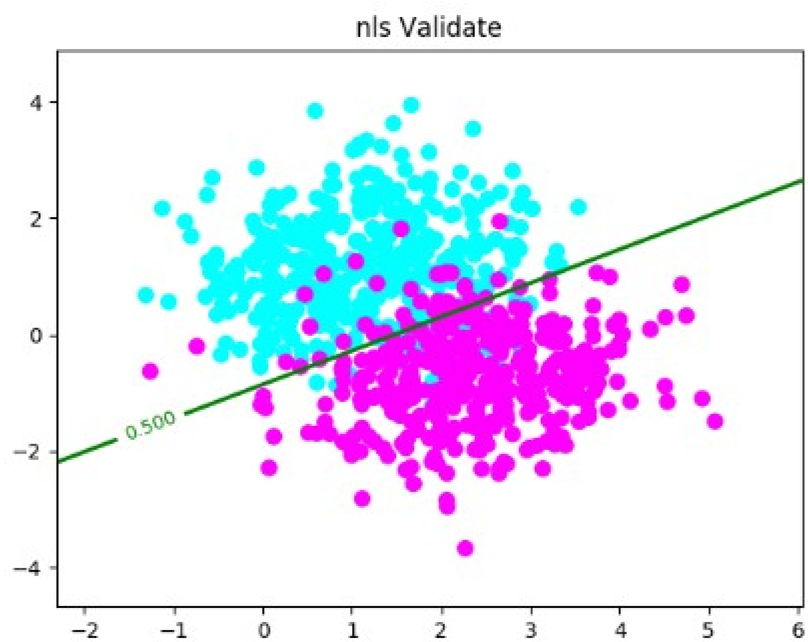
w1 = -1.41668237

w2 = 2.44940736

Training set error count : 30

Validation set error count : 66





Non Linearly Separable(2):

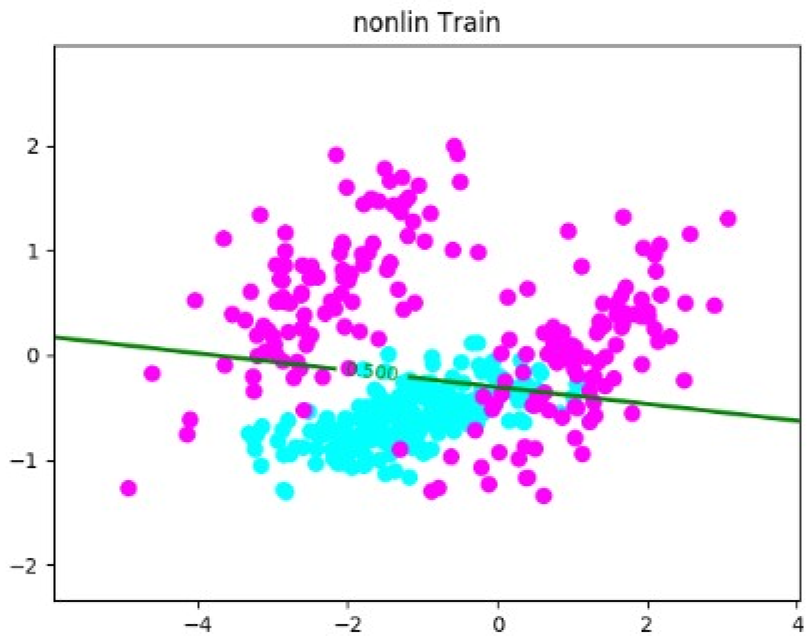
Weights : w0 = -1.01674506

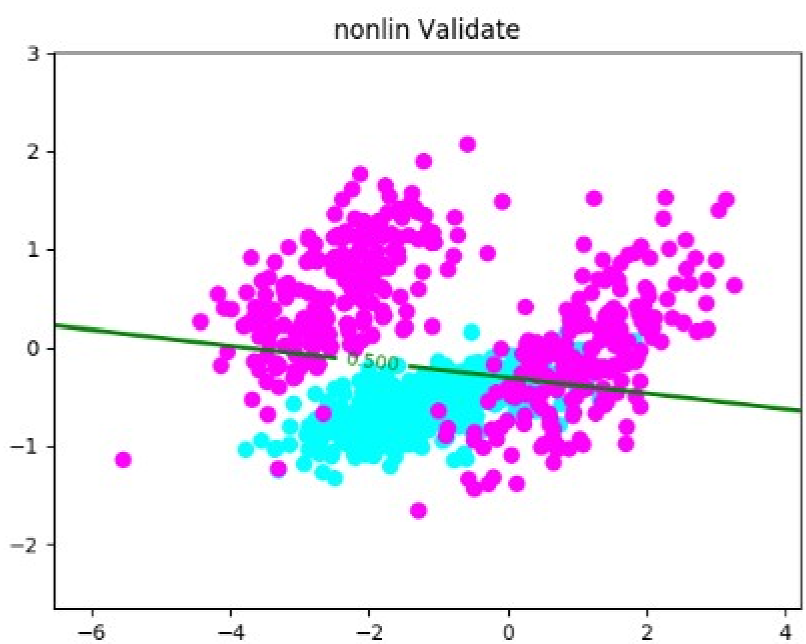
w1 = -0.26996381

w2 = -3.35432192

Training set error count : 69

Validation set error count : 130





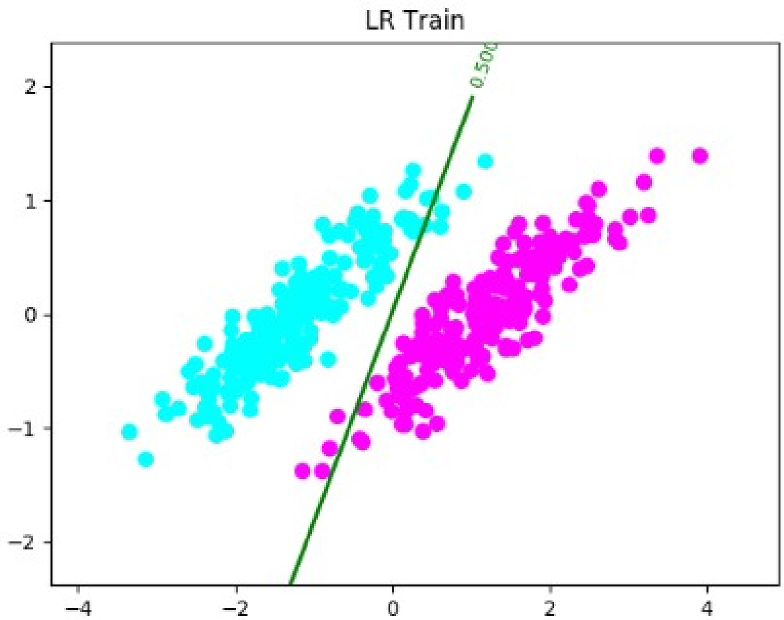
1.2.2 Behavior of the algorithm when lambda increases:

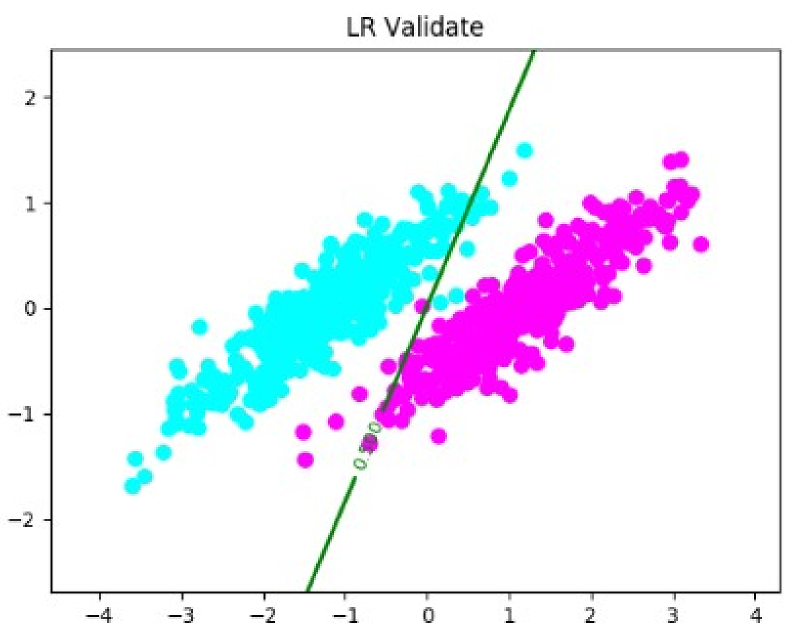
When lambda increases, the model will be regularized and hence the number of errors increases and the weight vector will be smaller. As lambda increases, the slope of the decision boundary increases until a point when it becomes parallel to y axis. (w1, w2 will be heavily regularized that they become equal to 0). When **L = 10,** the graphs and the corresponding weights are:

Linearly Separable:

Weights : w0 = -0.01506527
w1 = -1.72425101
w2 = 0.93339326

Training set error count : 8
Validation set error count : 17





Non Linearly Separable:

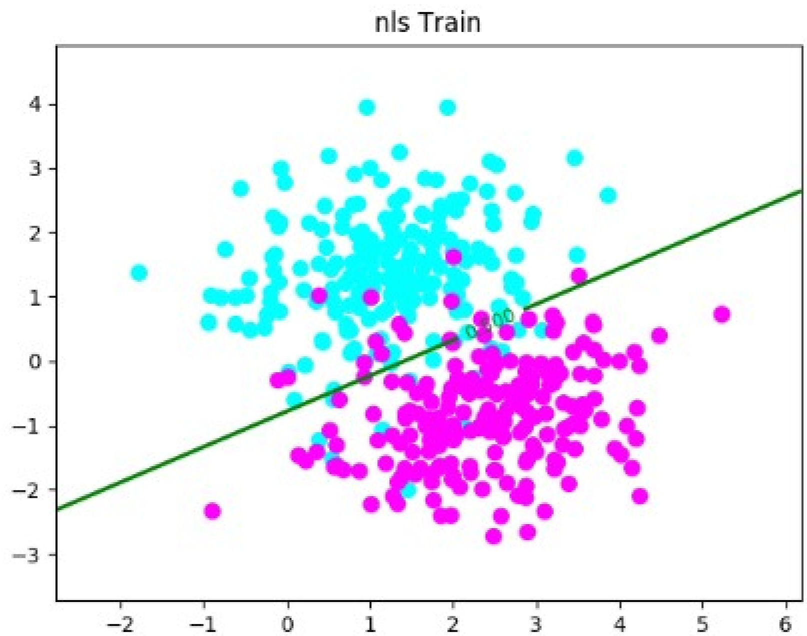
Weights : w0 = 1.08993567

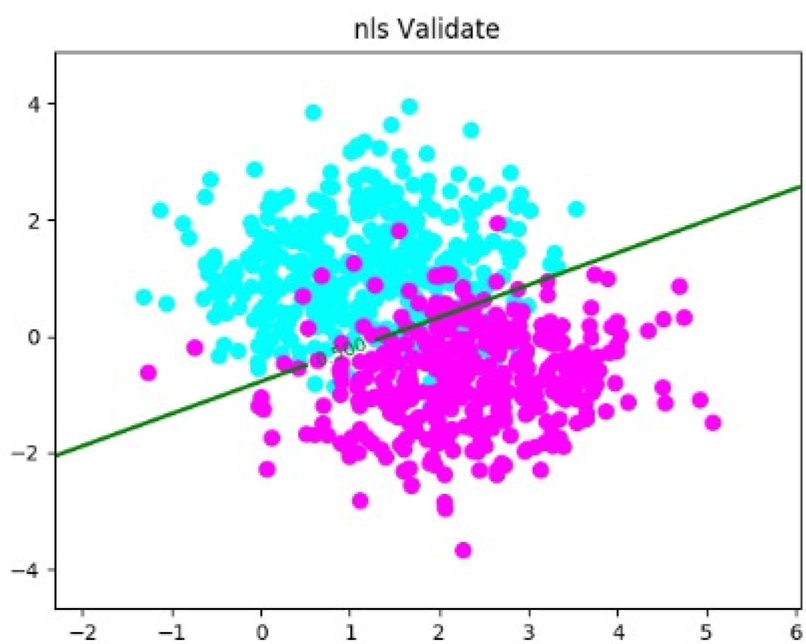
w1 = -0.77421673

w2 = 1.39863589

Training set error count : 30

Validation set error count : 68





Non Linearly Separable(2):

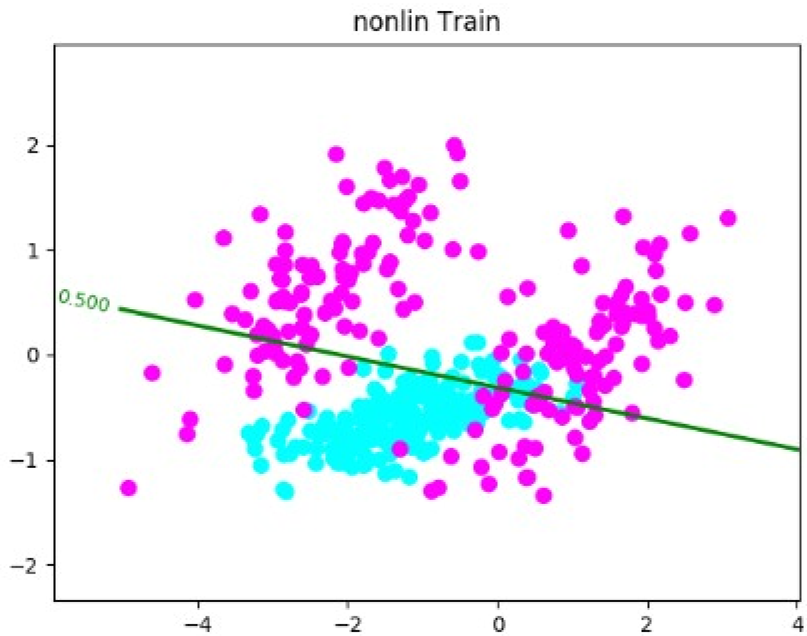
Weights : w0 = -0.45733524

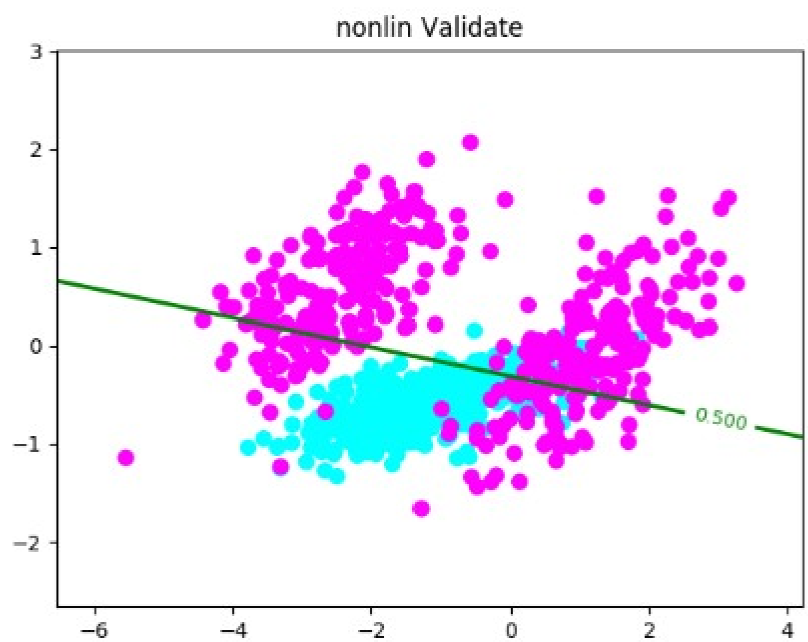
w1 = -0.2172903

w2 = -1.47284982

Training set error count : 73

Validation set error count : 143





When l = 100,

Linearly Separable:

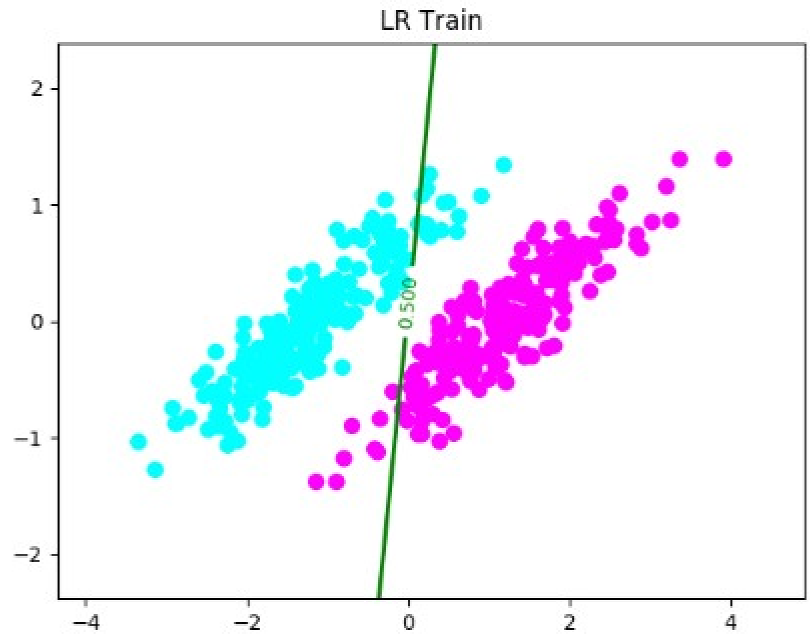
Weights : w0 = -0.00817829

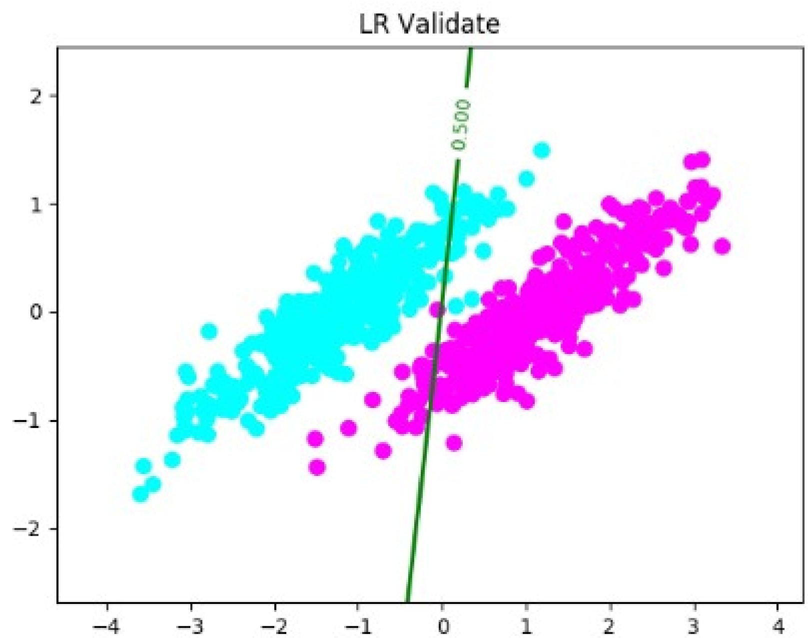
w1 = -0.62546143

w2 = 0.09179153

Training set error count : 23

Validation set error count : 50





Non-Linearly Separable:

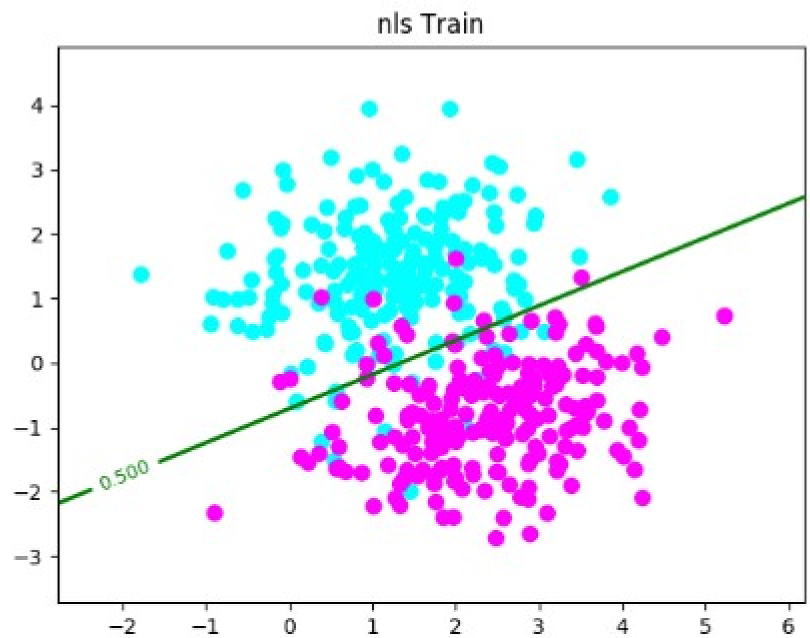
Weights : w0 = 0.38704908

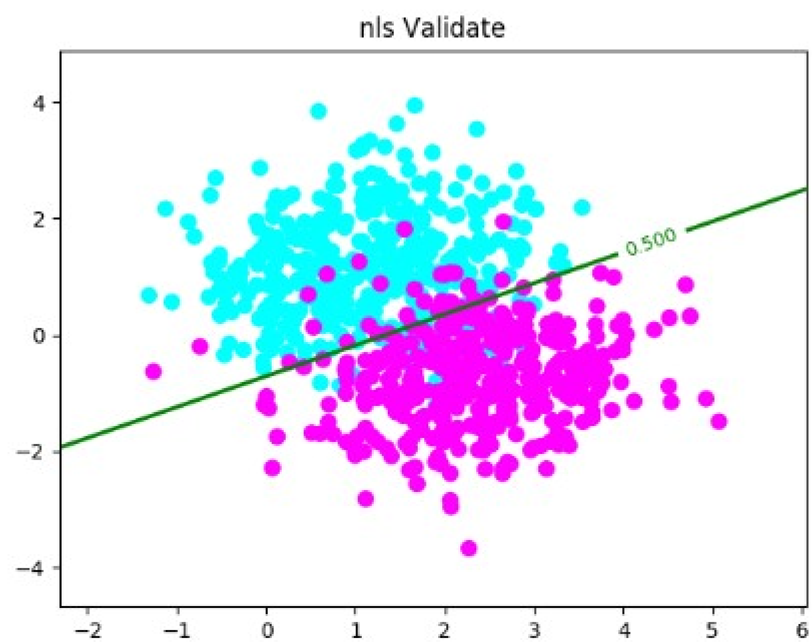
w1 = -0.28980222

w2 = 0.5459234

Training set error count : 29

Validation set error count : 69





Non-Linearly Separable(2):

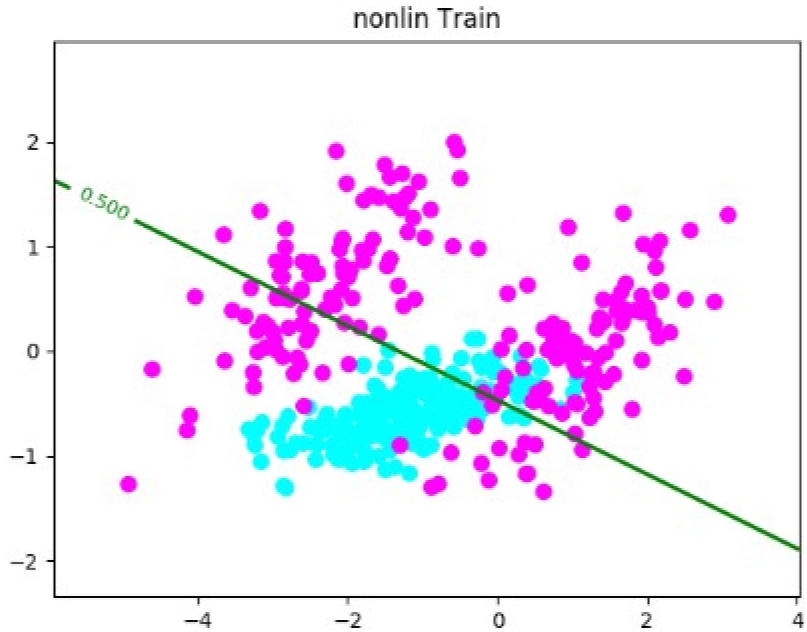
Weights : w0 = -0.16595198

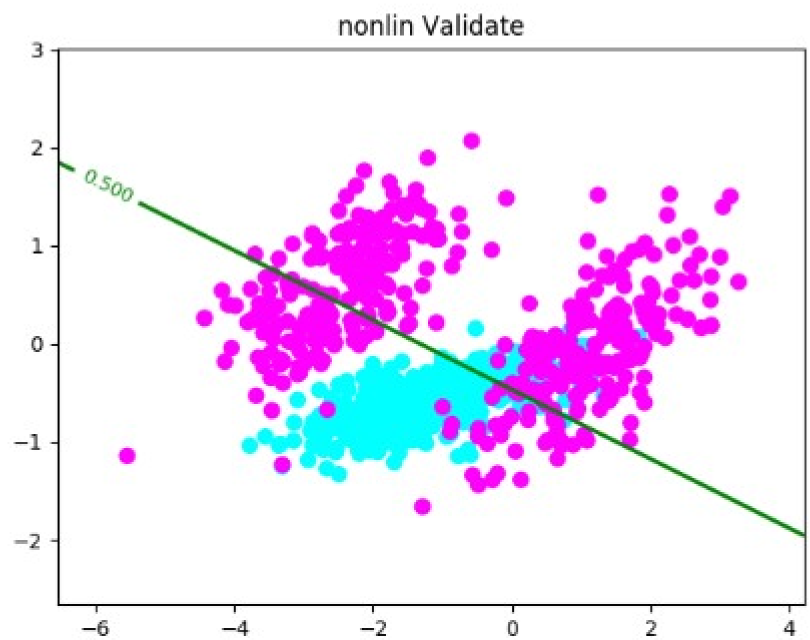
w1 = -0.12572148

w2 = -0.3552611

Training set error count : 86

Validation set error count : 193





1.2.3 General trends:

Since the regularized logistic regression is a for a linearly separable input data, it is not able to accurately classify the data points for non separable data. Hence, the error on non separable data set is very high.

1.3: Regularization on polynomial basis functions:

Since we are mapping the input to higher order, it is possible for the logistic regression to handle second order polynomial basis. Hence without regularization, the model over-fits. Regularization causes the weights to be small and the weights will be smaller as we increase l. For non-linear and non separable cases, the error is less since the features are mapped in higher dimensions. For linearly separable data, mapping to higher order feature space provided worse results but for nls and nonlin data sets, there was a better performance because mapping into higher space allowed the decision boundary to linearly separate the data in the new feature space.

As we can see from the above graphs and data, as we increase the regularizing term l, over-fitting reduces and the number of errors decreases. The rate of change of error rate in validation set is less and this shows us that when using regularization, over-fitting is reduced and hence the model is a little flexible but the number of errors in the training data increases since the model is not tightly fit to it anymore.

The result of using polynomial basis functions setting **l = 0 (No regularization)** is given below. The weights and the number of errors are also specified.

Linearly Separable:

Weights : w0 = 0.04729134

w1 = -2.42724697

w3 = 0.1450316

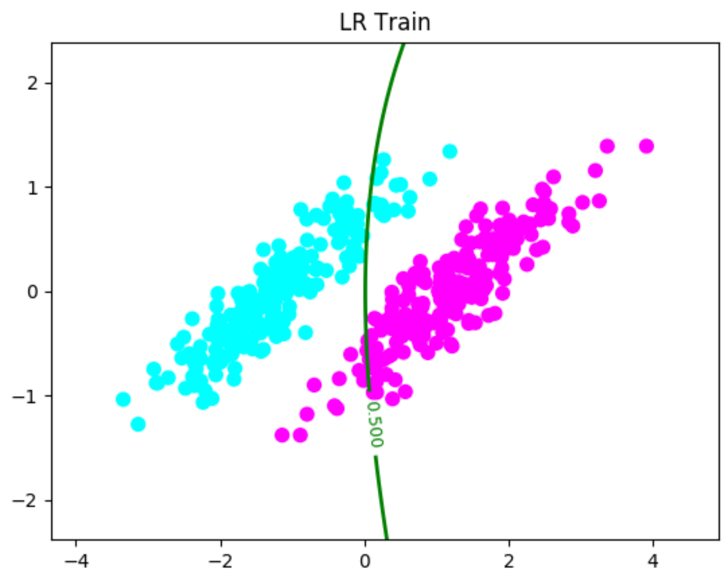
w4 = 0.14503517

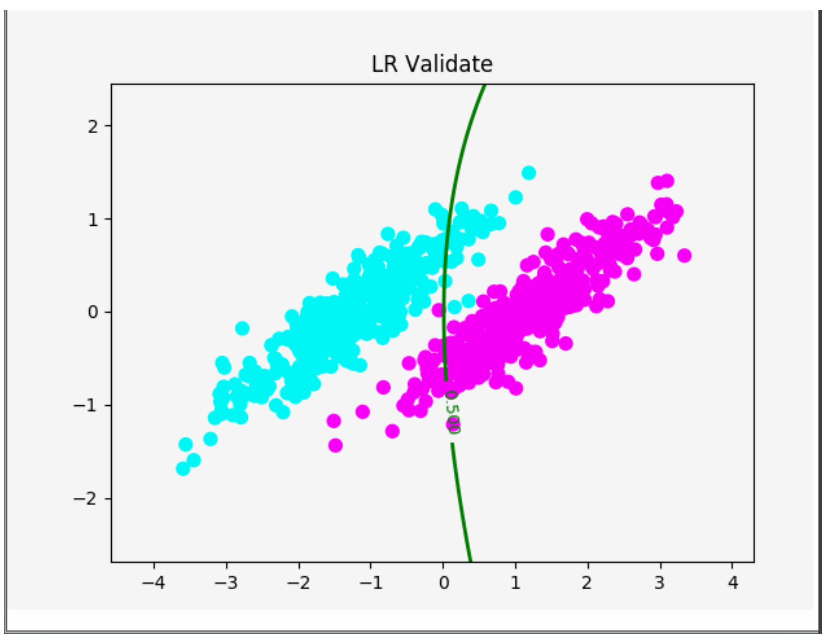
w5 =-0.14854224

w6 = 0.23866279

Training set error count : 27

Validation set error count : 59





Non Linearly Separable:

Weights : $w0 = 1.93069782$

$w1 = -0.76020259$

$w3 = 0.46761887$

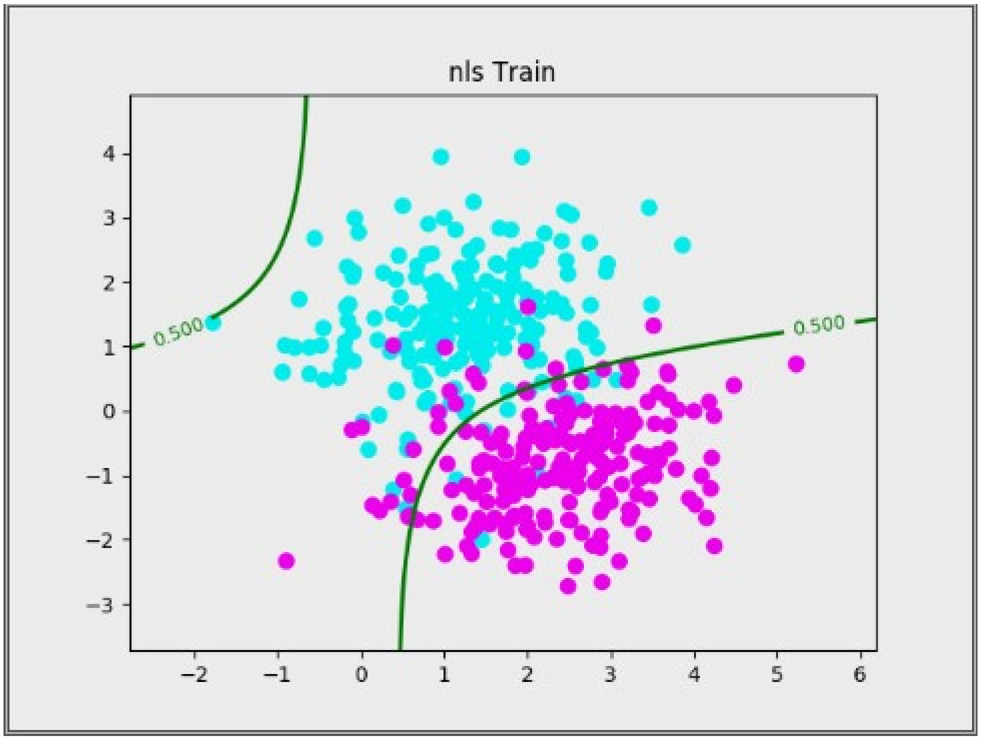
$w4 = 0.46760018$

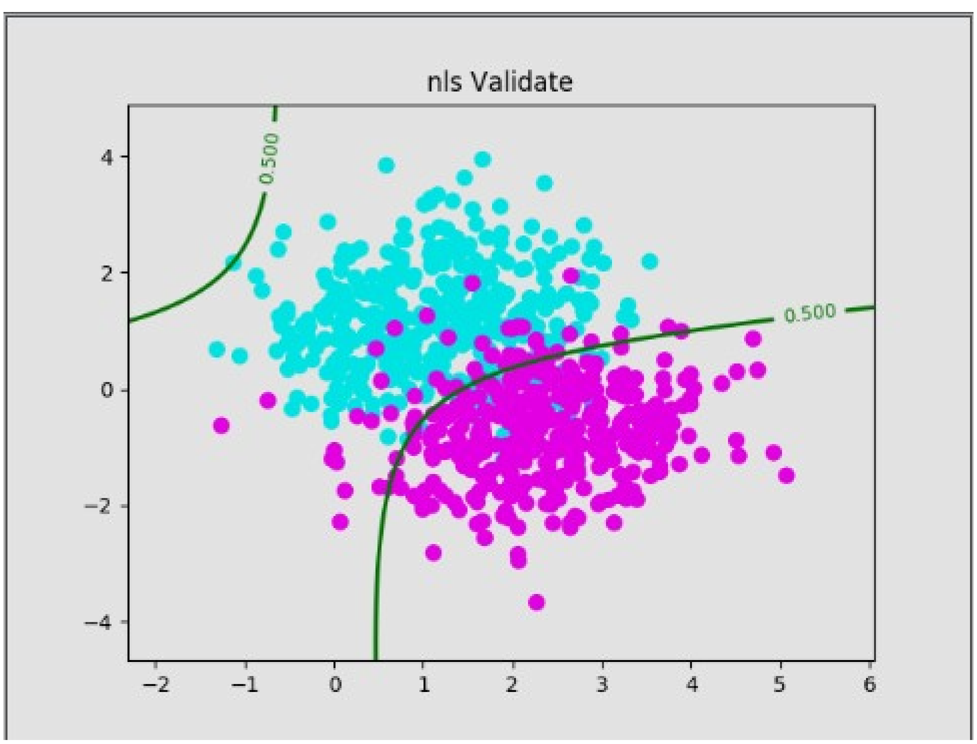
$w5 = -0.18408145$

$w6 = 0.06980724$

Training set error count : 35

Validation set error count : 81



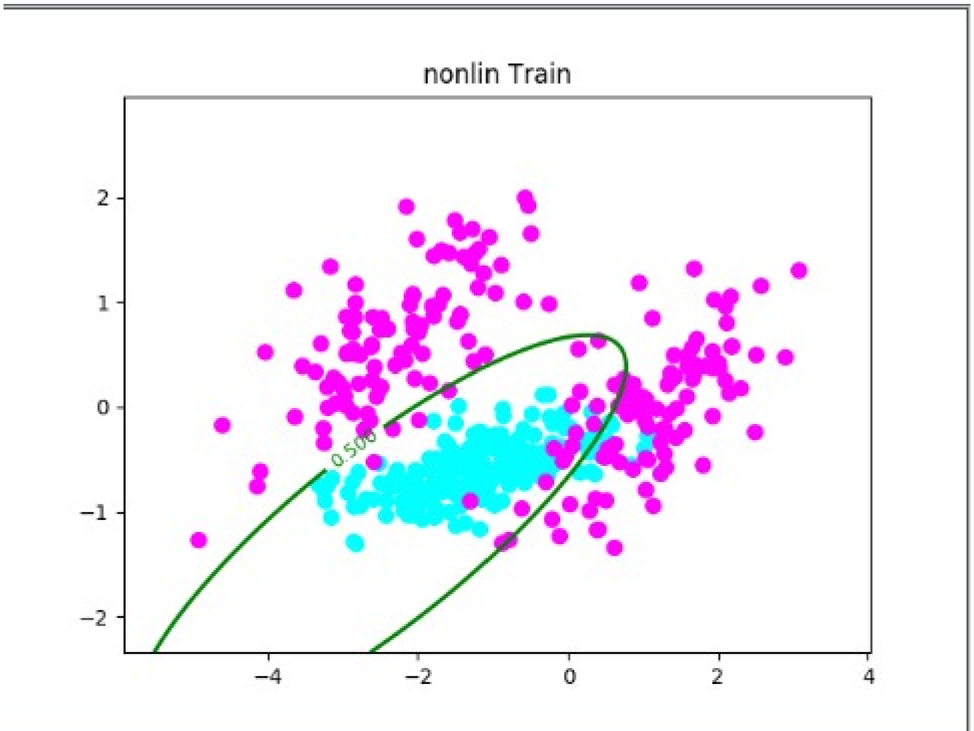


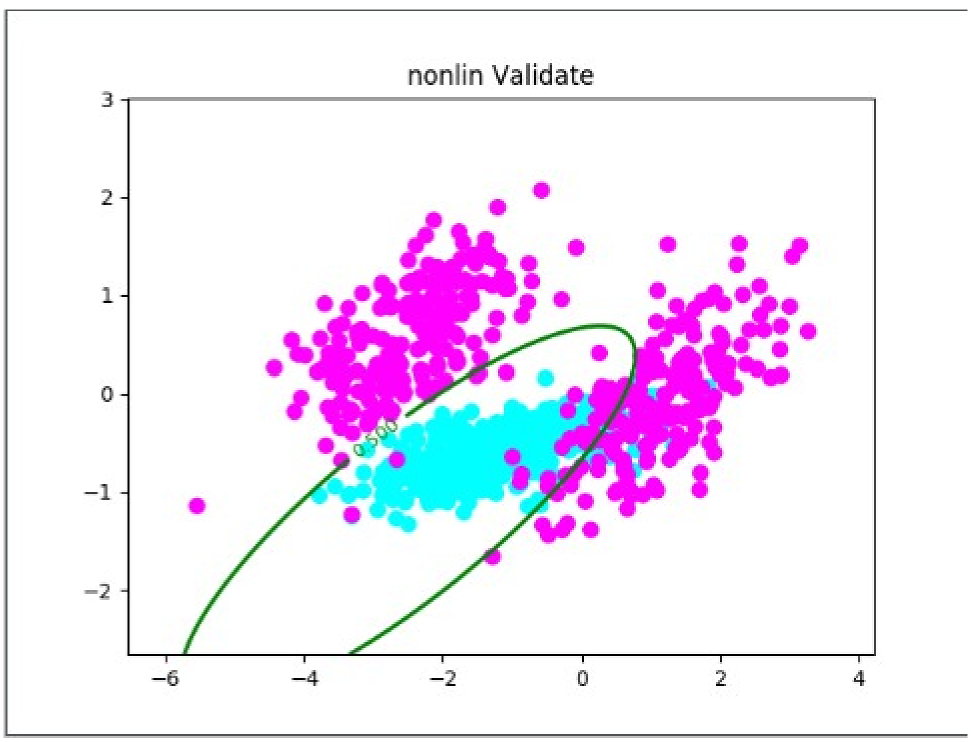
Non Linearly Separable(2):

Weights : w0 = 1.73782384
w1 = -1.37659095
w3 = 1.3825534
w4 = 1.38255344
w5 = -1.36161498
w6 = -3.92934765

Training set error count : 35

Validation set error count : 72





When l = 10,

Linearly Separable:

Weights : w0 = -0.02418571

w1 = -1.4304912

w3 = 0.03984169

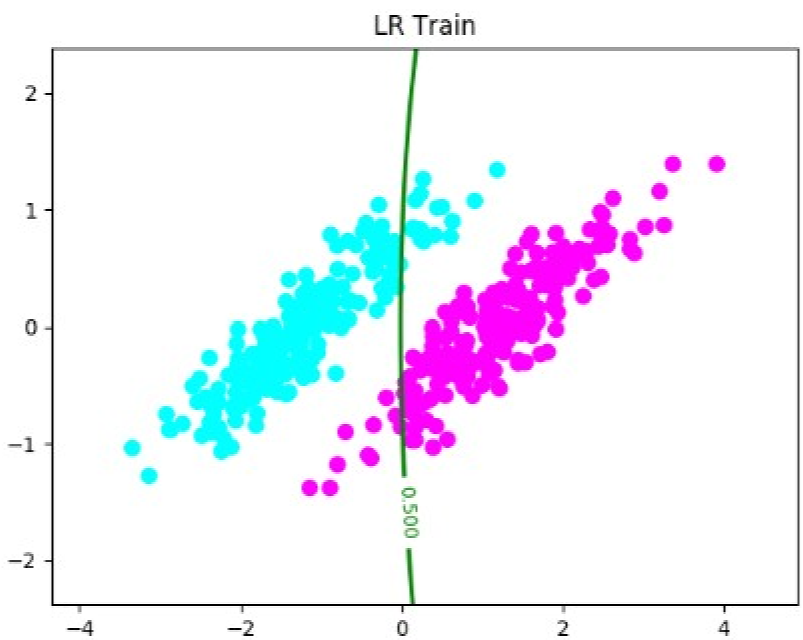
w4 = 0.03984173

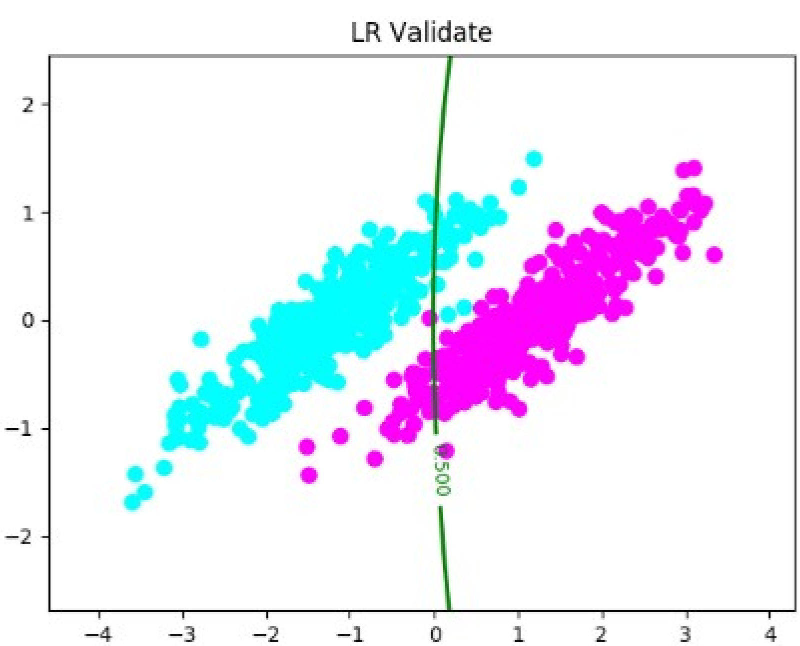
w5 = -0.00655184

w6 = 0.06099804

Training set error count : 25

Validation set error count : 60





Non Linearly Separable:

Weights : w0 = 1.53472266

w1 = - 0.38436523

w3 = 0.4076053

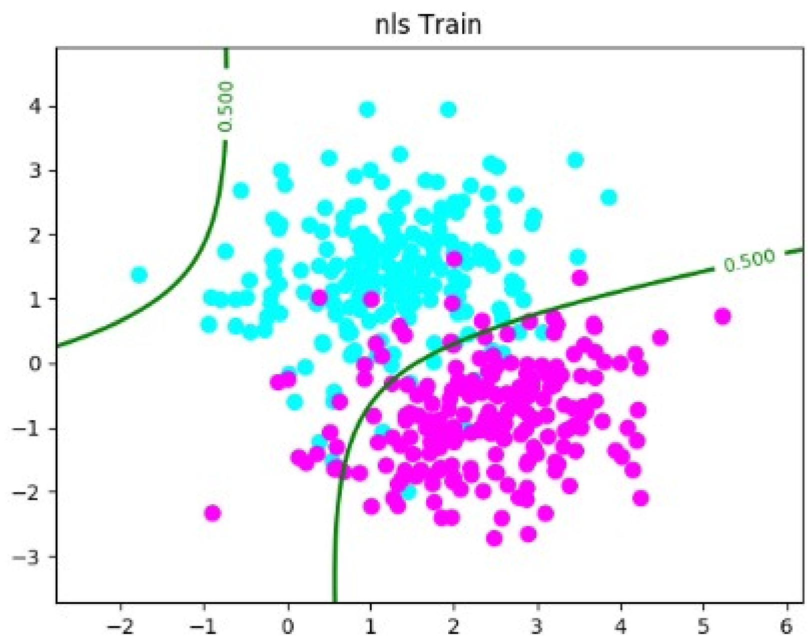
w4 = 0.40760526

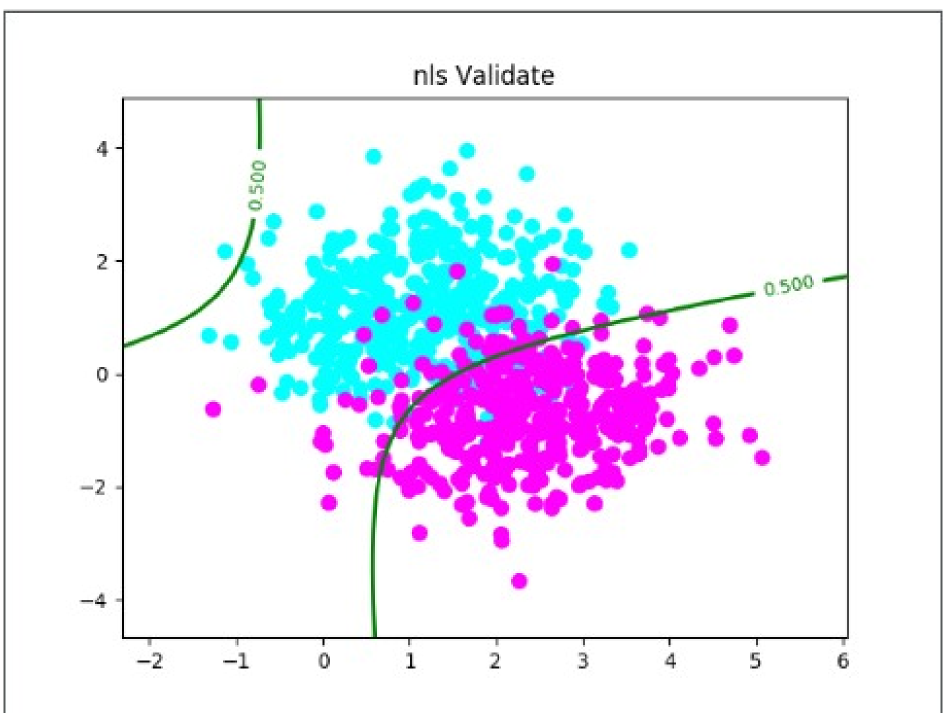
w5 = -0.28855404

w6 = 0.0977048

Training set error count : 37

Validation set error count : 81





Non Linearly Separable(2):

Weights : w0 = 0.67617748

w1 = -0.81251499

w3 = 0.6192013

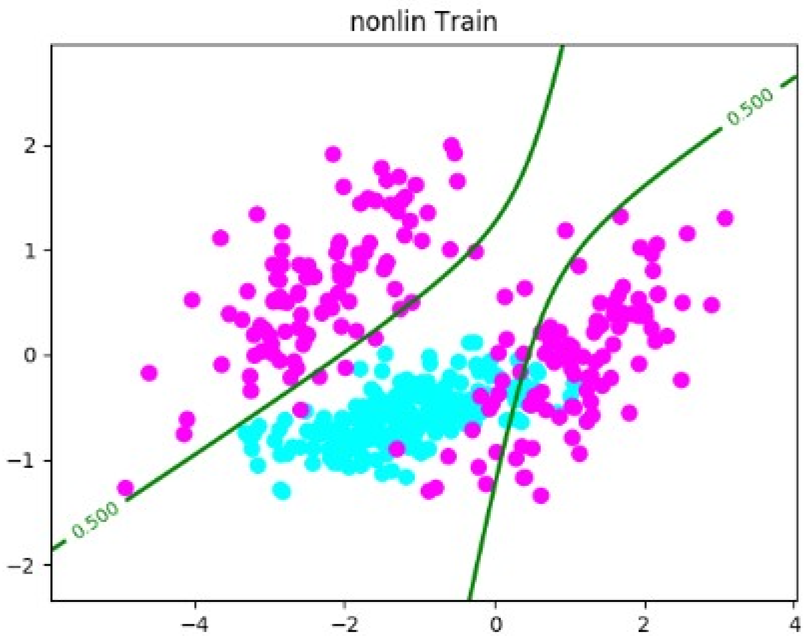
w4 = 0.61920118

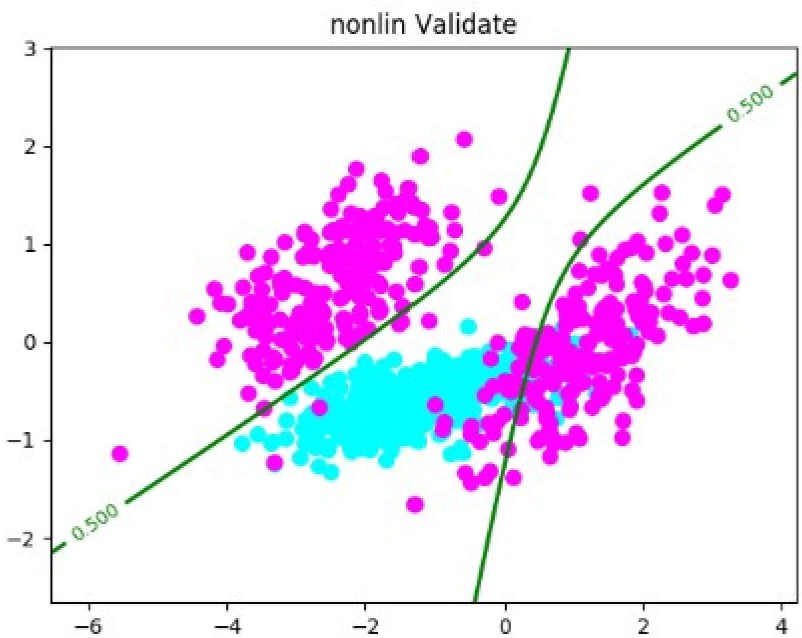
w5 = -0.72197768

w6 = -0.43022213

Training set error count : 40

Validation set error count : 74





2. **Support Vector Machine:**

2.2 SVM Implementation report:

The results of implementing SVM is depicted in the graphs below. Since nls and nonlin are not linearly separable, the upper bound of c on alpha was required to produce a classification. The primal form of SVM is also implemented and the graphs corresponding to **c = 50** is given below for both primal and dual on all the data sets. The weights and the misclassification errors on the training and validation set is also listed. As we can see from the graphs below, as we increase c, the model over-fits and the number of misclassification error reduces. The effect of c in primal s more noticeable that that of dual

Primal:

Linearly Separable:

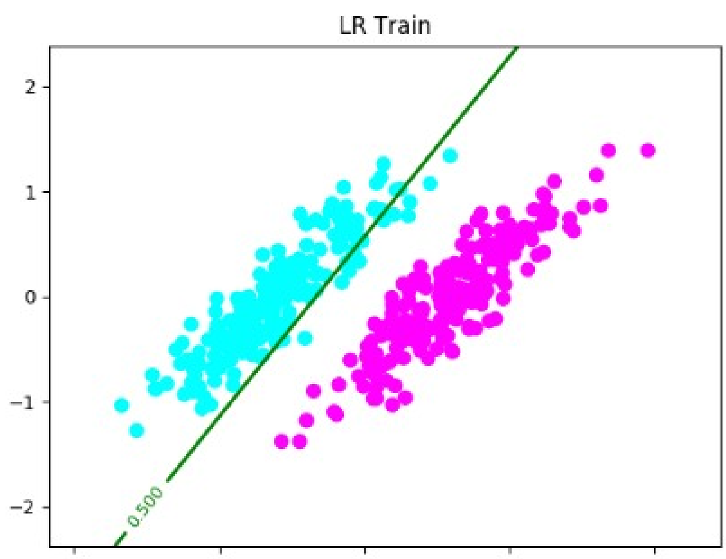
Weights : w0 = -0.24505618

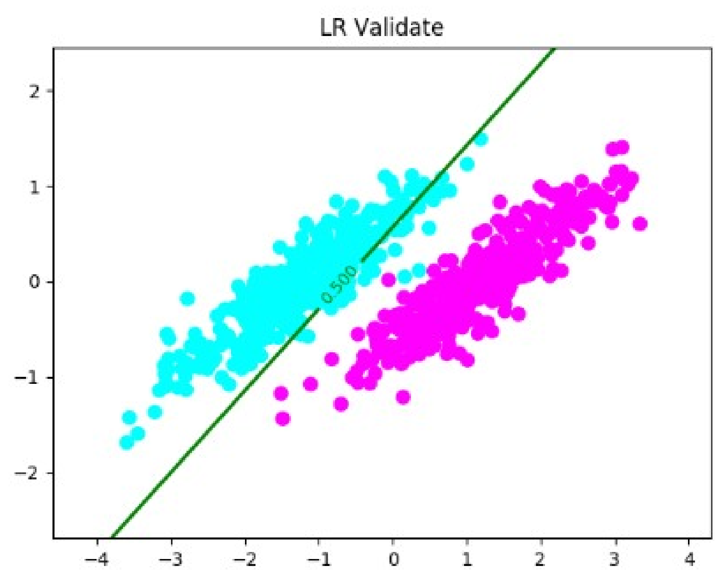
w1 = -1.12852829

w2 = 1.31886421

Training set error count : 0

Validation set error count : 3





Non Linearly Separable:

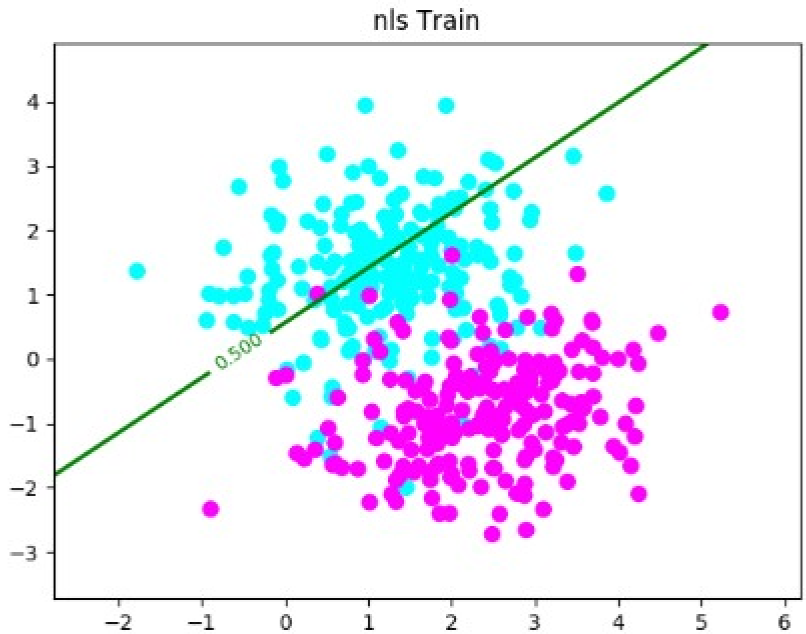
Weights : w0 = -0.48839257

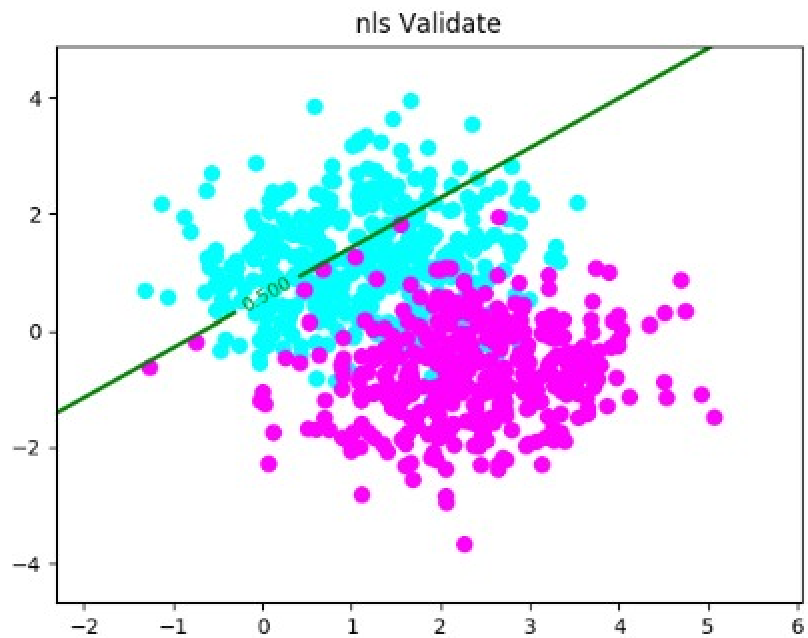
w1 = 0.19219624

w2 = 1.95725041

Training set error count : 91

Validation set error count : 214





Non Linearly Separable(2):

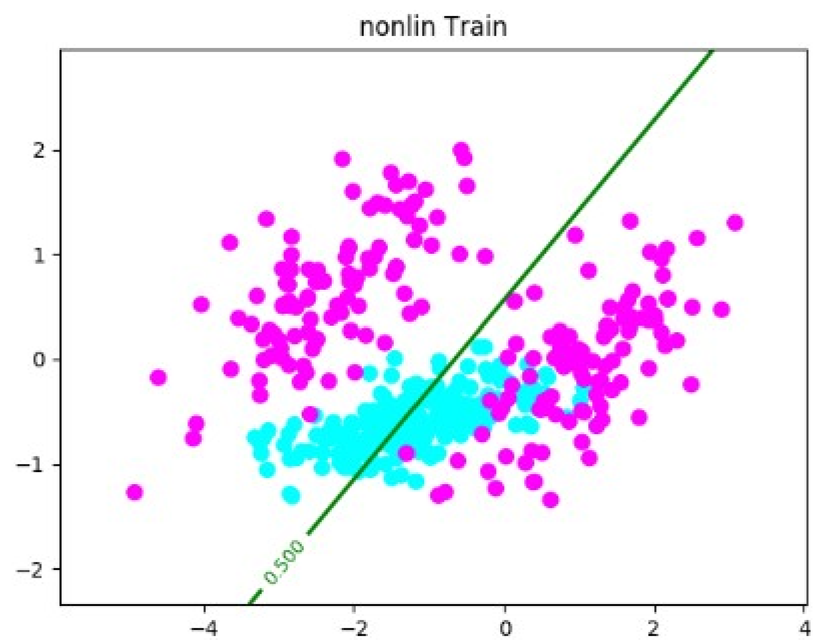
Weights : w0 = -0.50677567

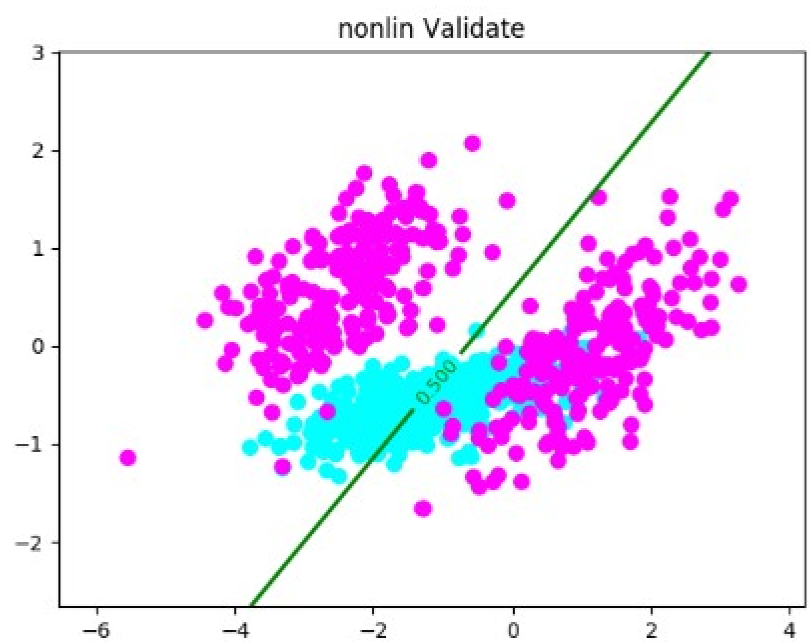
w1 = -0.21785751

w2 = -1.90695535

Training set error count : 180

Validation set error count : 352





When c = 10, for primal we get the following output

Linearly Separable:

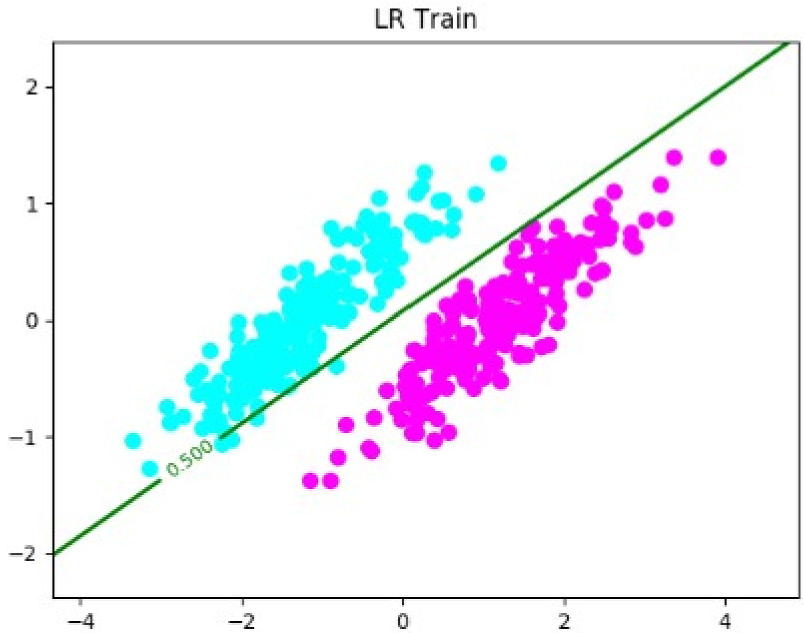
Weights : w0 = 0.30644623

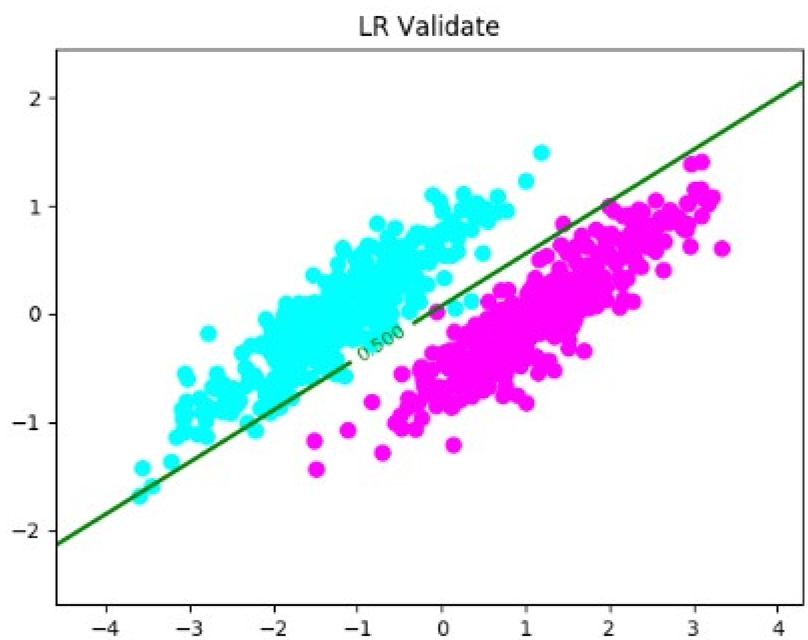
w1 = -1.27487651

w2 = 2.65057583

Training set error count : 4

Validation set error count : 11





Non Linearly Separable:

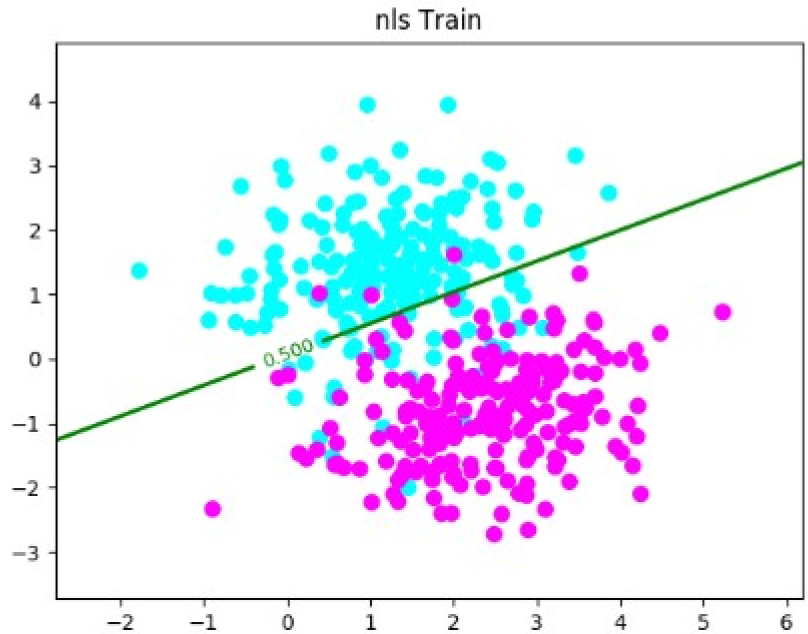
Weights : w0 = -0.48839257

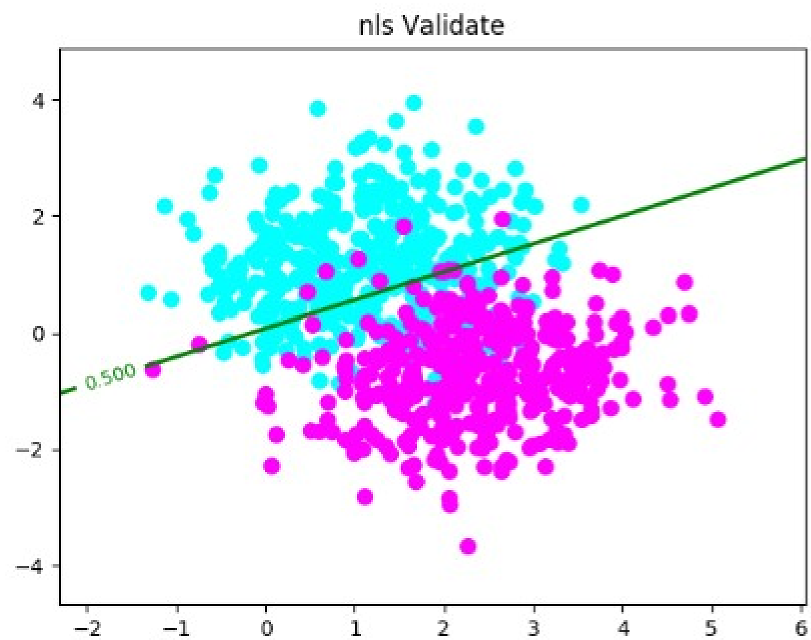
w1 = 0.19219624

w2 = 1.95725041

Training set error count : 42

Validation set error count : 103





Non Linearly Separable(2):

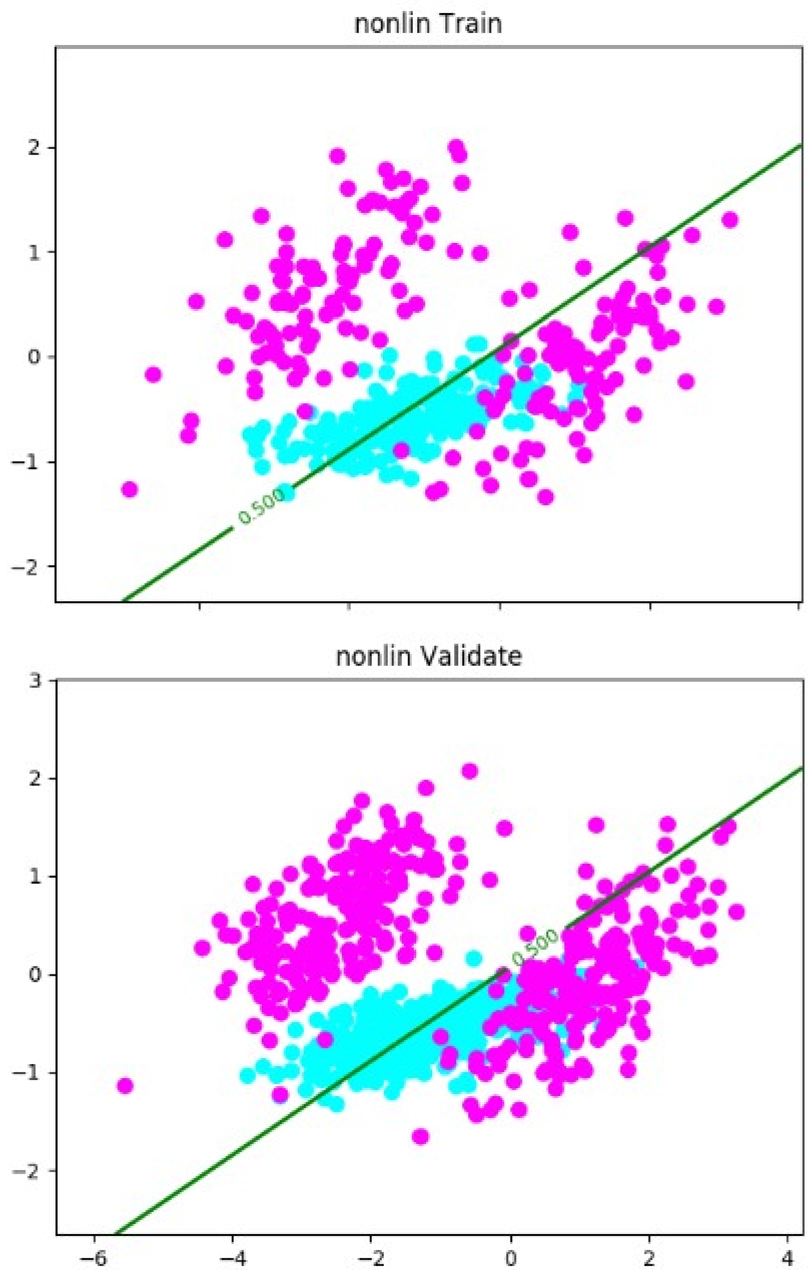
Weights : w0 = -0.50677567

w1 = -0.21785751

w2 = -1.90695535

Training set error count : 194

Validation set error count : 397



For dual, when C = 50

Linearly Separable:

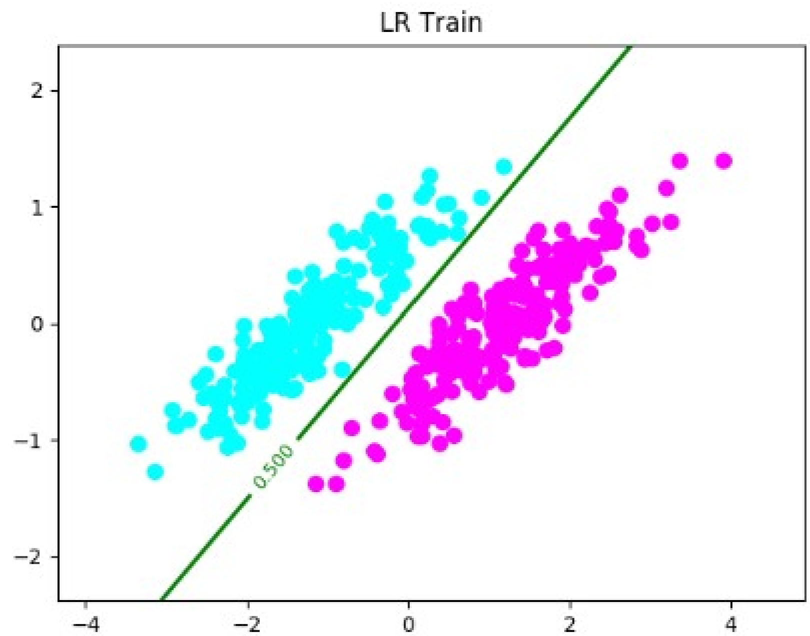
Weights : w0 = 0.09222021

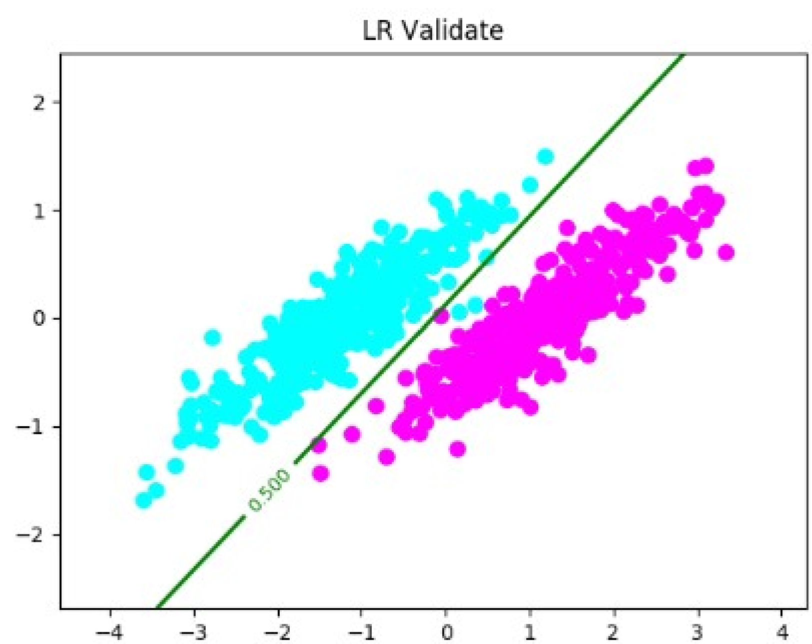
w1 = -2.76438112

w2 = 3.38479058

Training set error count : 0

Validation set error count : 4





Non Linearly Separable:

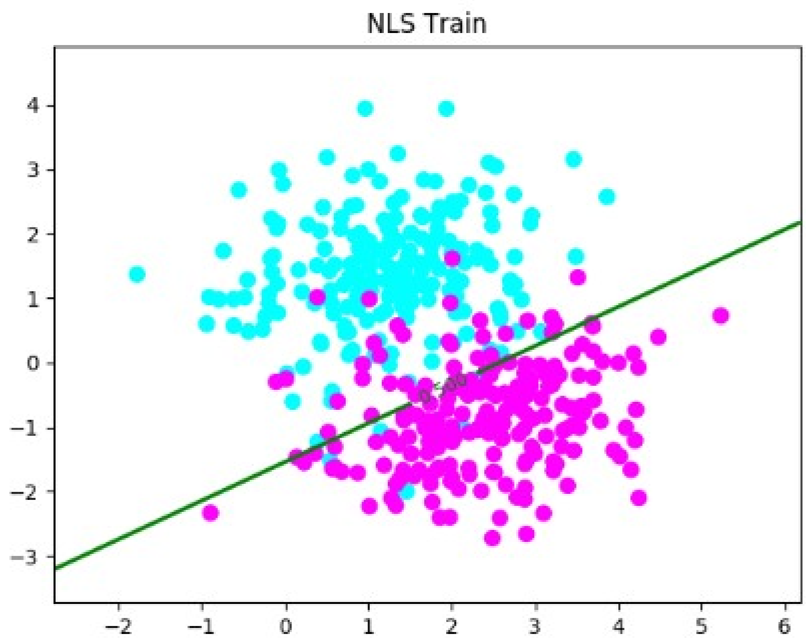
Weights : w0 = 2.68168858

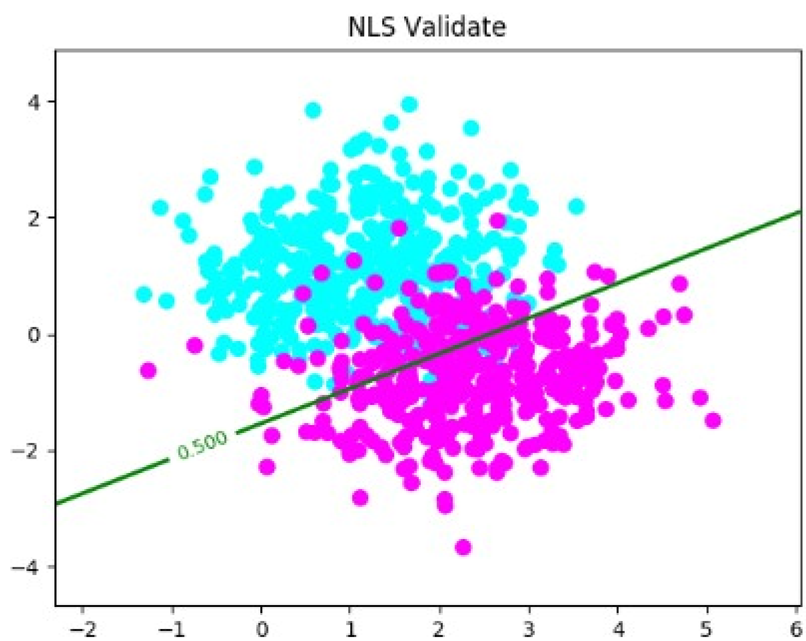
w1 = -0.8509694

w2 = 1.41568194

Training set error count : 67

Validation set error count : 156





Non Linearly Separable(2):

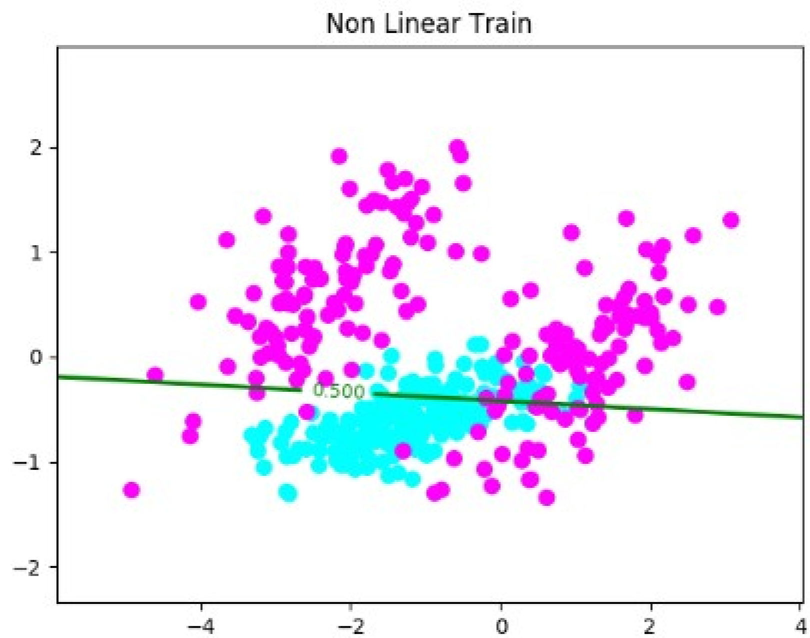
Weights : w0 = -0.6156045

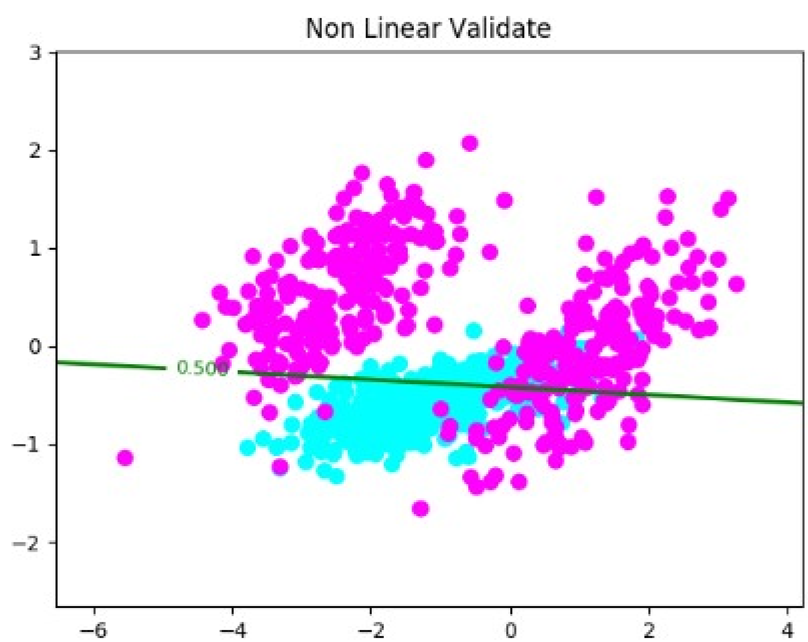
w1 = -0.10330625

w2 = -2.6558326

Training set error count : 64

Validation set error count : 126





When C = 10,

Linearly Separable:

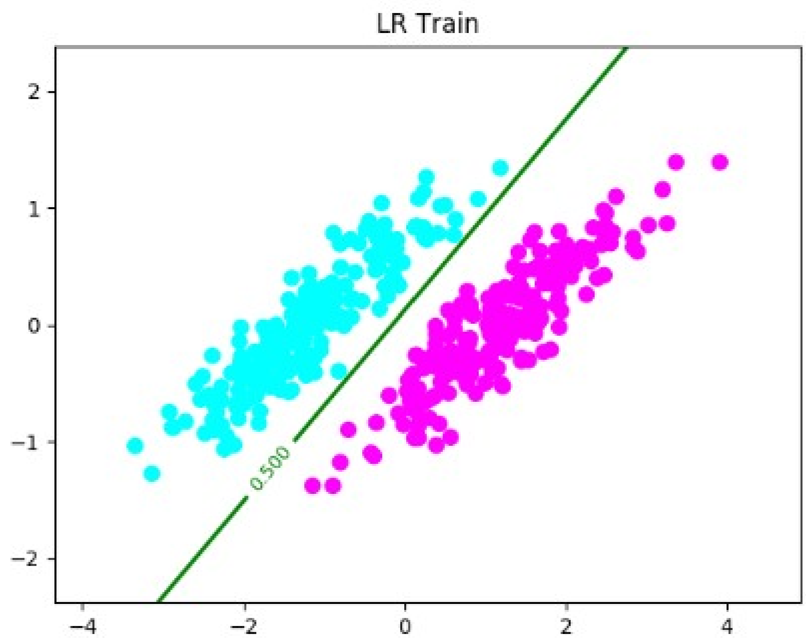
Weights : w0 = 0.09222021

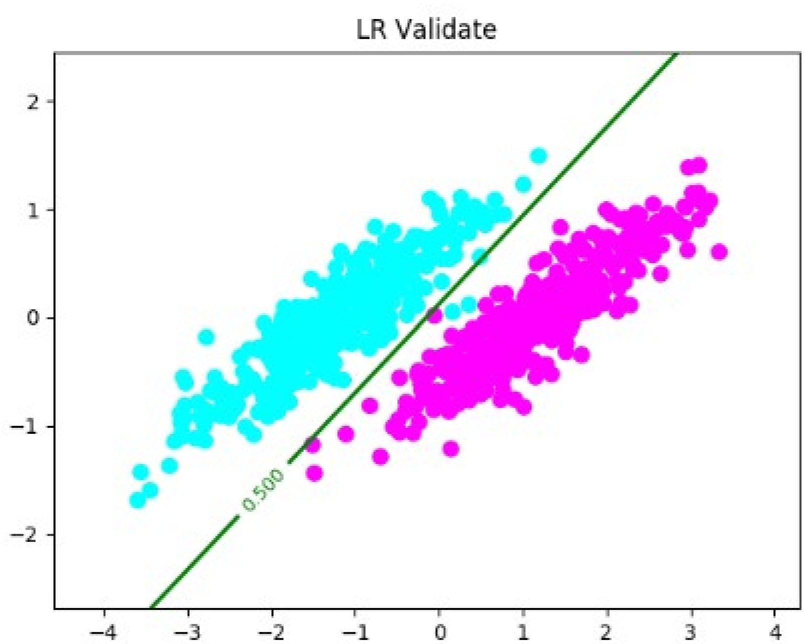
w1 = -2.76438112

w2 = 3.38479058

Training set error count : 0

Validation set error count : 4





Non Linearly Separable:

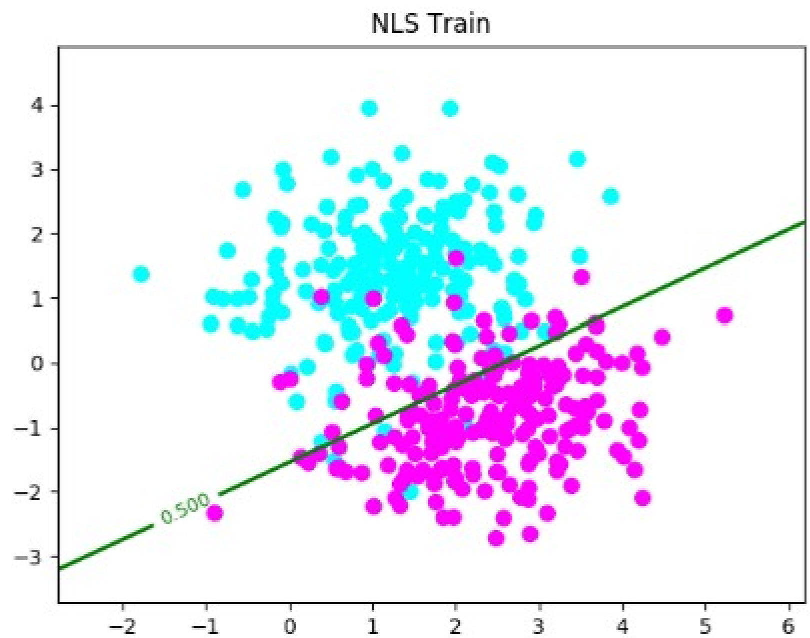
Weights : $w0 = 2.68168855$

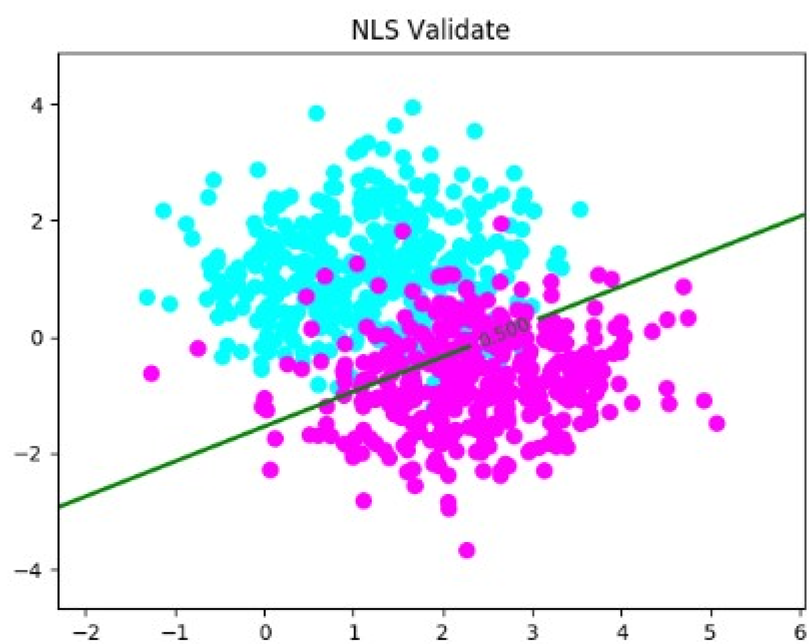
$w1 = -0.85096936$

$w2 = 1.41568182$

Training set error count : 67

Validation set error count : 156





Non Linearly Separable(2):

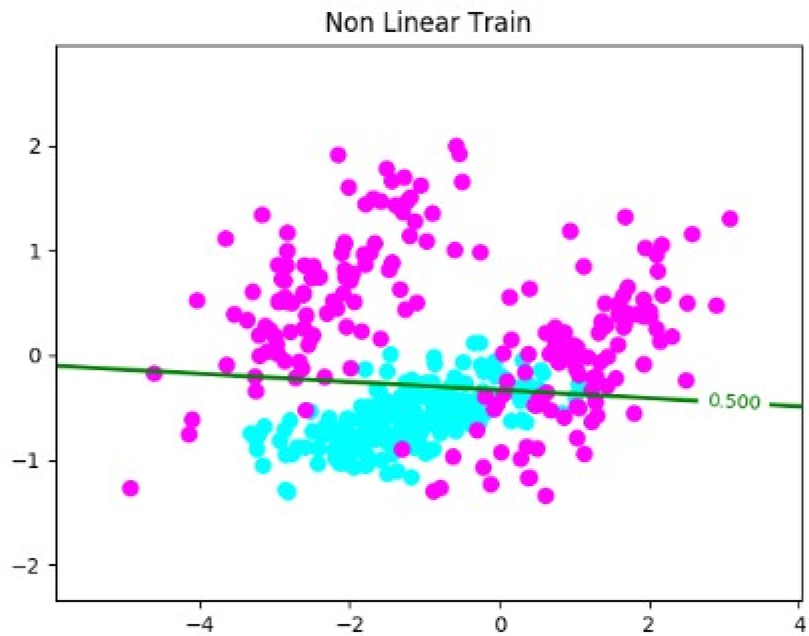
Weights : w0 = -0.37877077

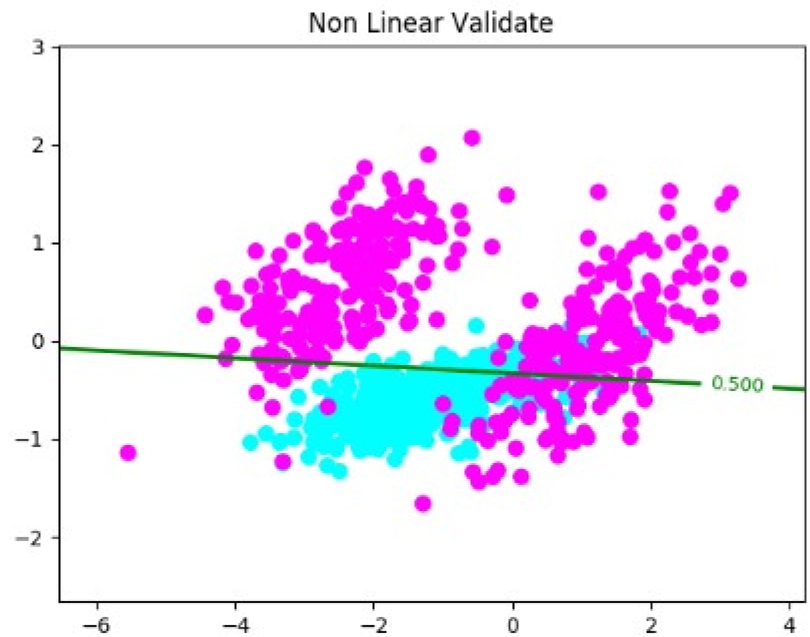
w1 = -0.10330611

w2 = -2.65583341

Training set error count : 67

Validation set error count : 127





3. **Kernel Support Vector Machine:**

Some of the graphs obtained by applying a polynomial kernel are:

When c = 50,

Linearly Separable:

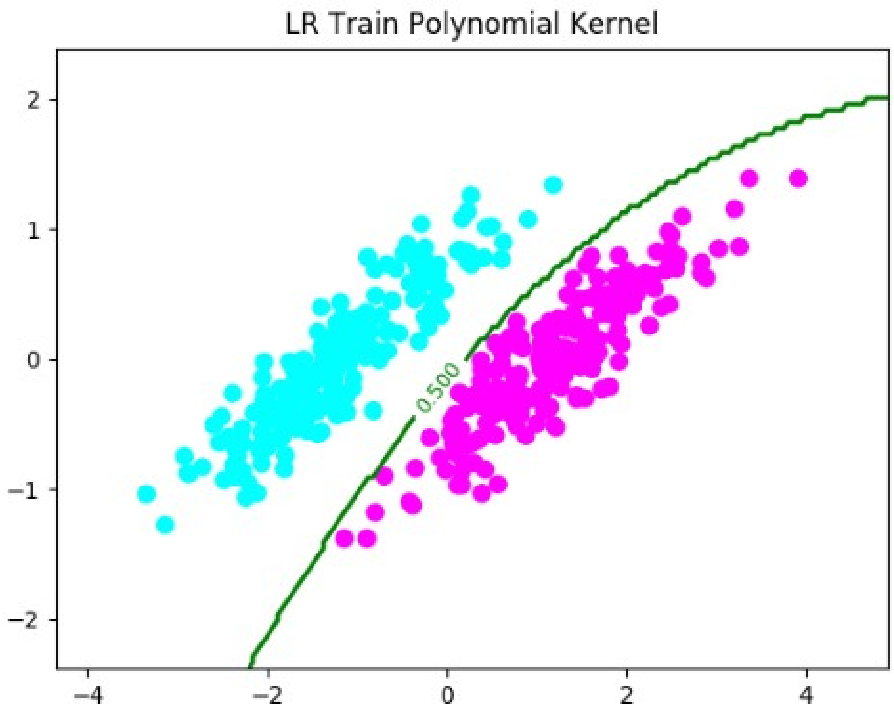
Weights : w0 = 0.57428486

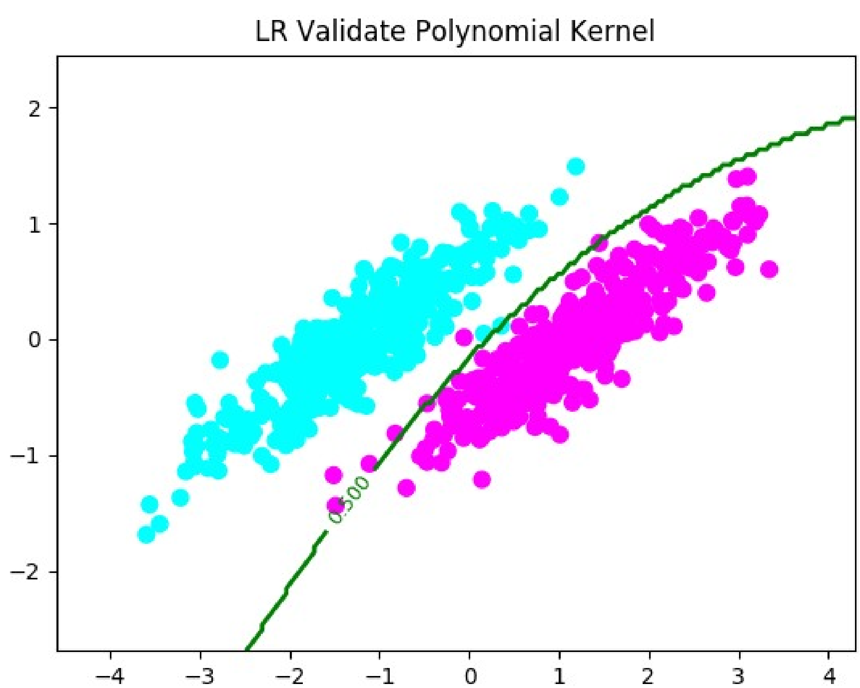
w1 = -1.32673392

w2 = 1.64958729

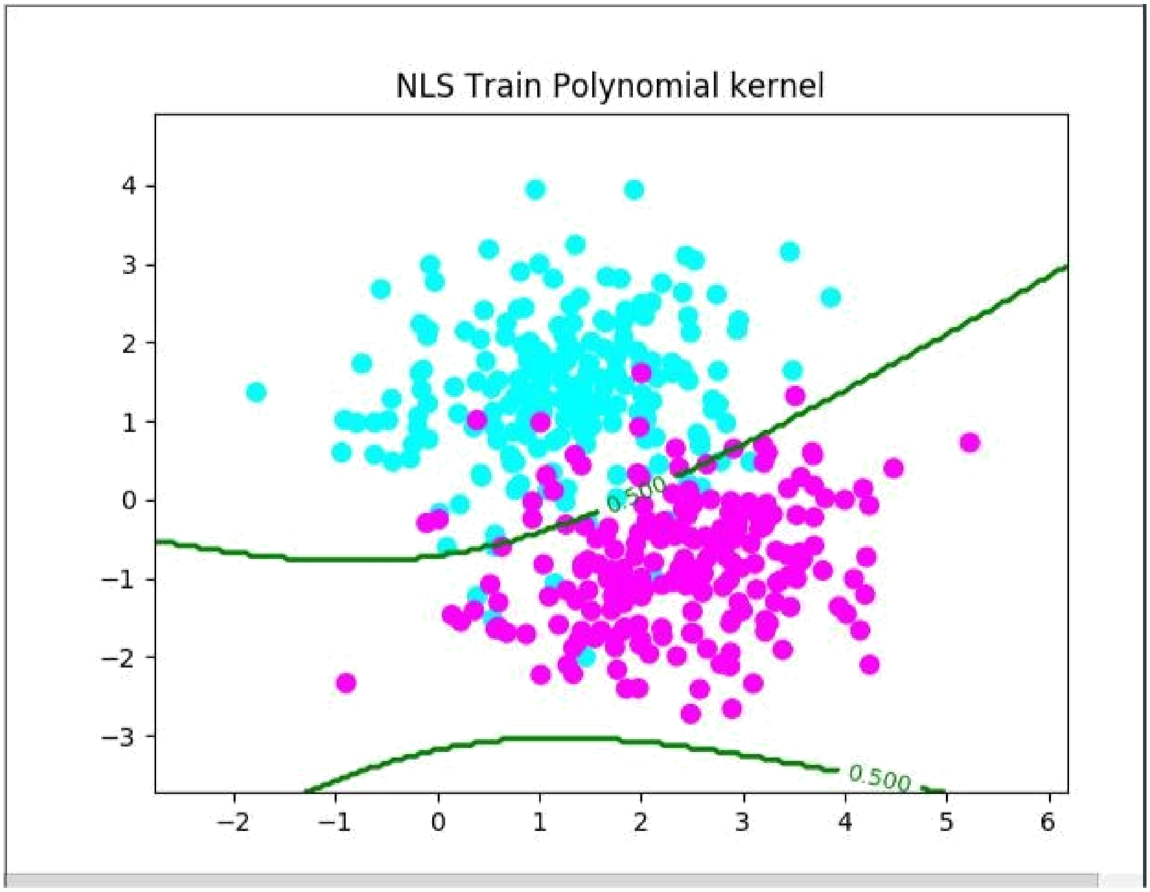
Training set error count : 0

Validation set error count : 7

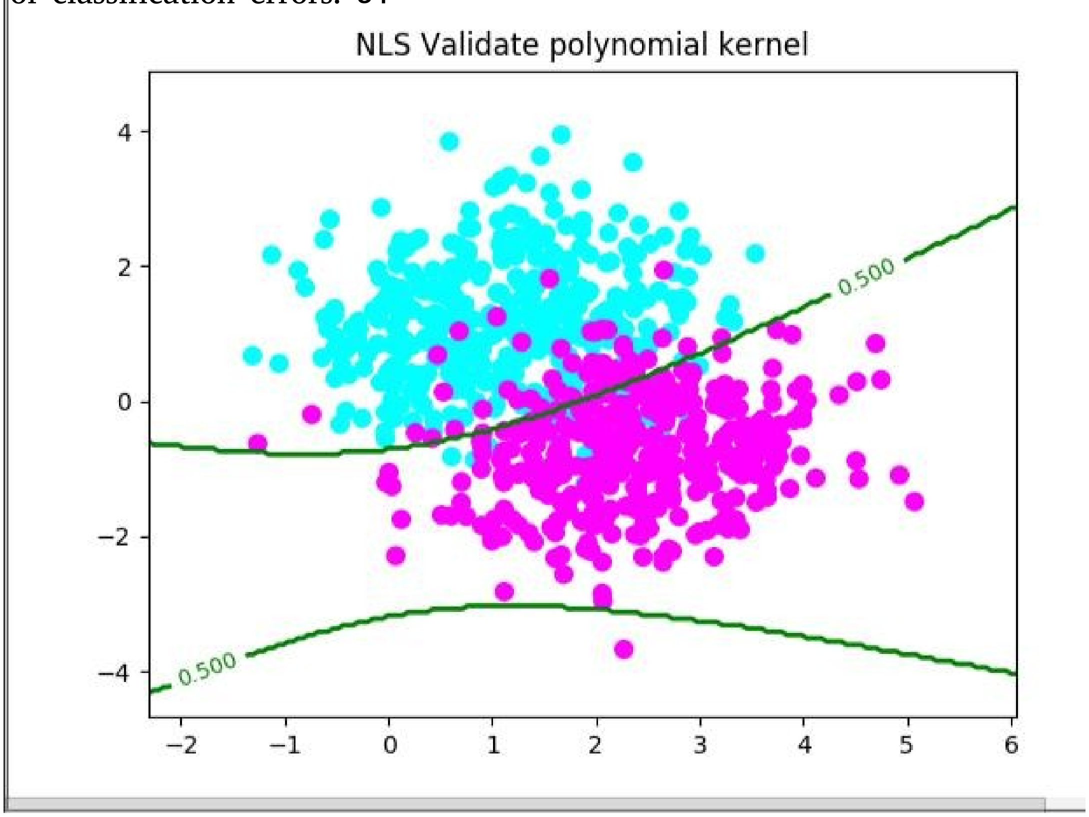




When C = 0,



Number of classification errors: 31

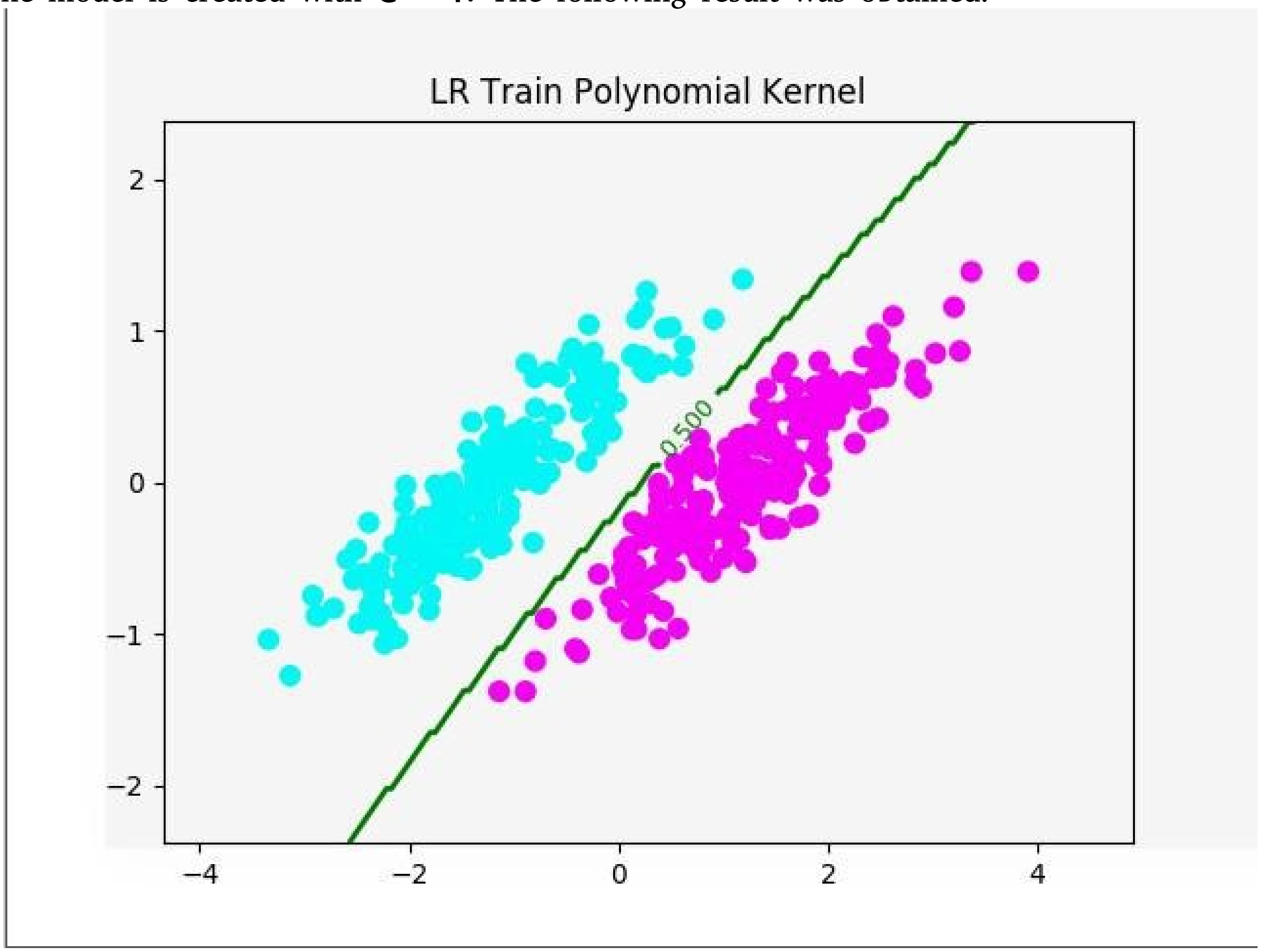


Number of errors: **78**

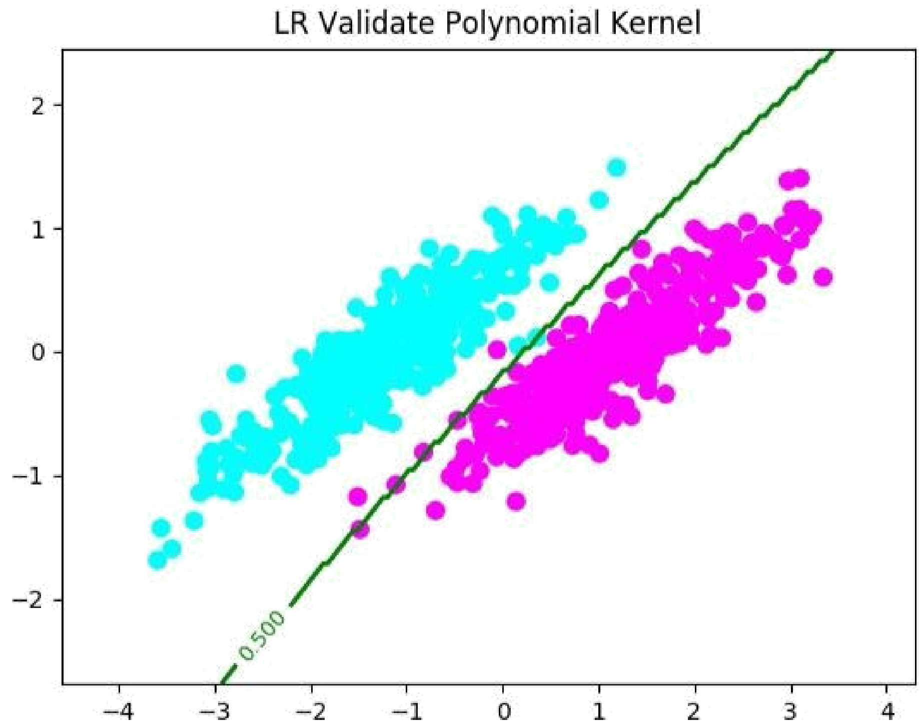
The Gaussian kernel did not produce results for linearly separable and non linearly separable data.

3.1 Effect of C

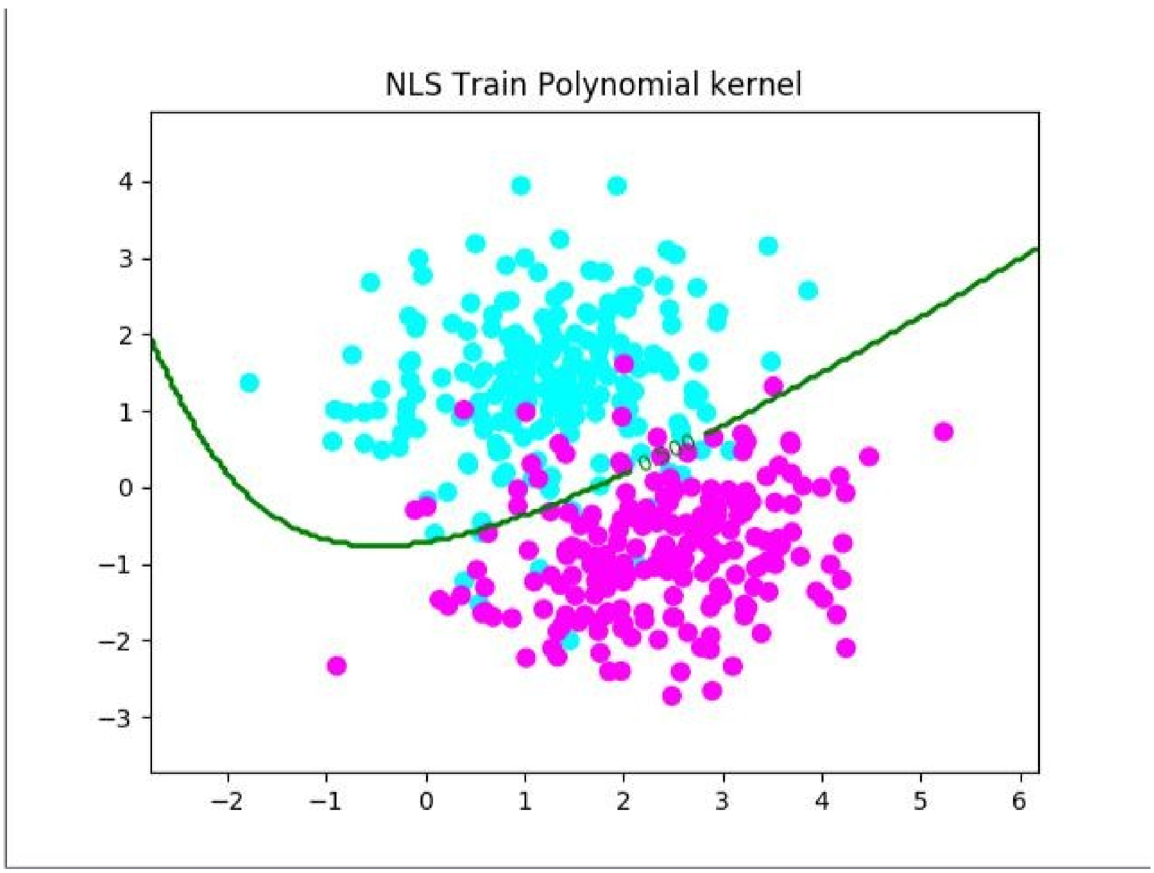
The model is created with **C = 1.** The following result was obtained:



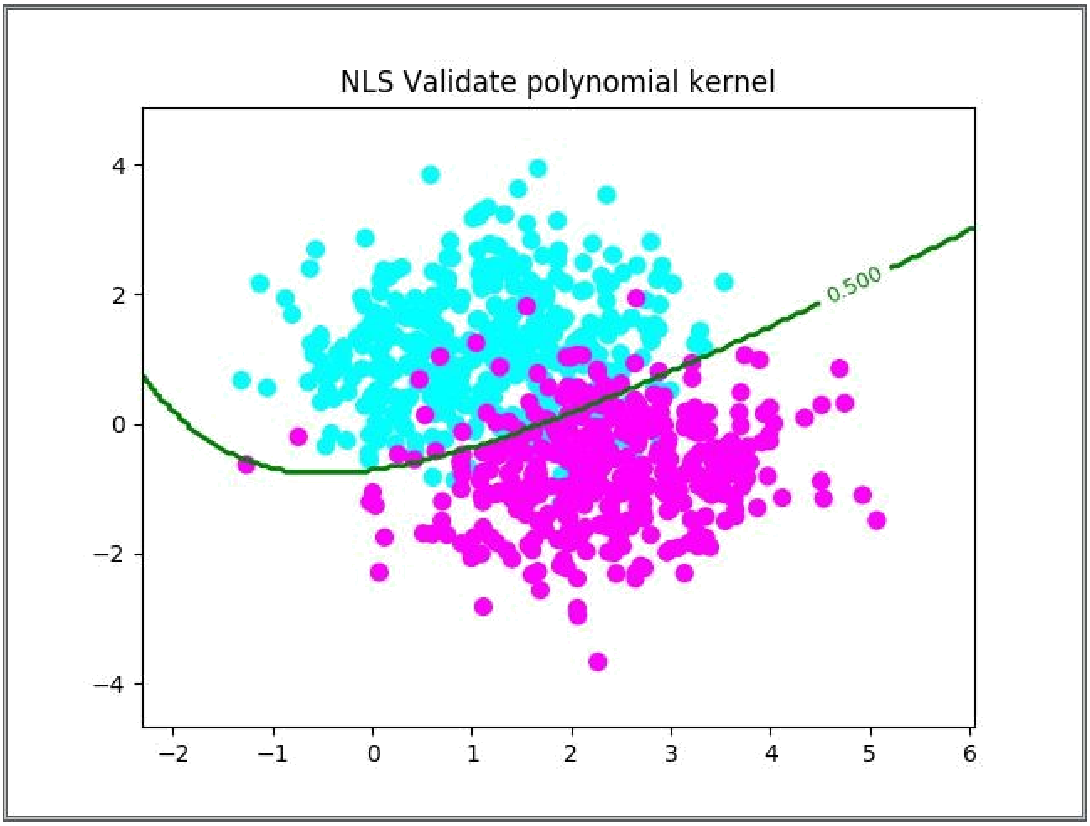
Number of classification errors: **0**



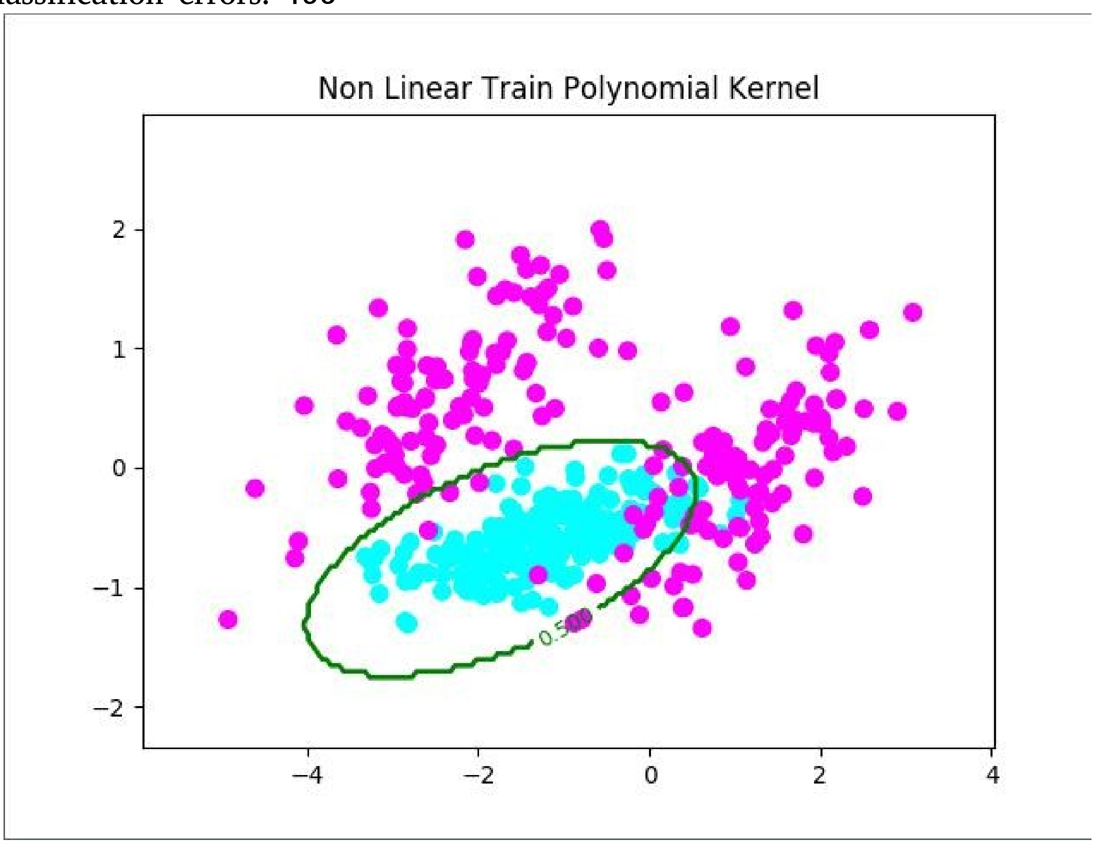
No. of classification errors: **4**



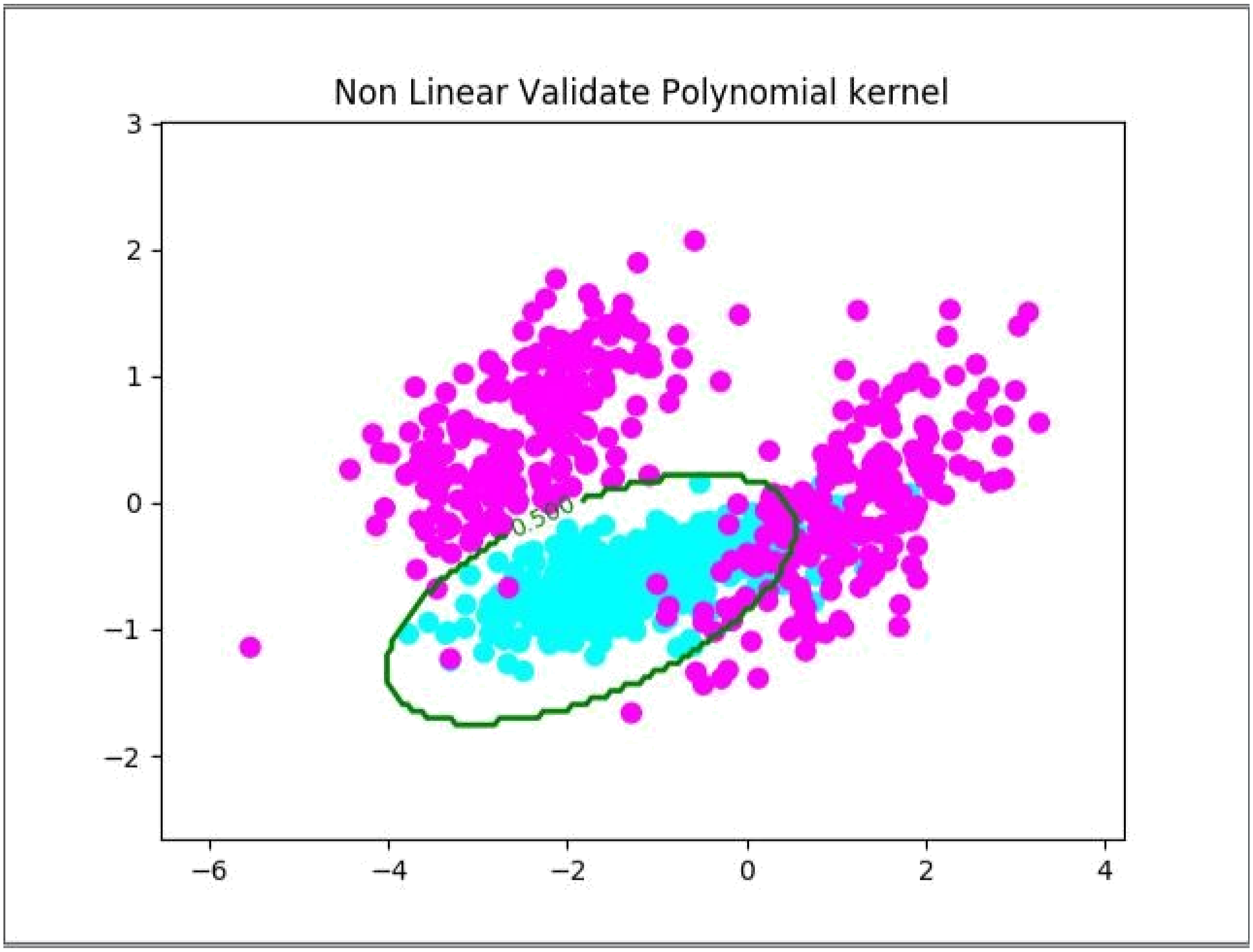
No. of classification errors: **67**



No of classification errors: **156**



No of classification errors: **67**



No of classification errors: 127