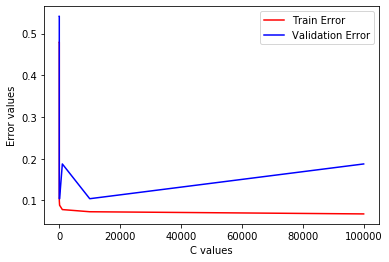
MACHIINE LEARNING

Assignment 4

**SVM with synthetic data:**

**The effect of the regularization parameter,**𝐶



c\_best : 1.0

test error : 0.16666666666666663

* Consistent with the standard theory train error reduces with increase in c\_value
* There is a small fluctuation in the pattern of the validation error (it spiked for c=1000 and came back to the same value again), but in the overall picture, it reduces along with the training error till a point and then increases at a slow rate.
* Least value of validation error is obtained for c = [1,10,100,10000]. As we have multiple minimum values first one will be picked using min(list) as that is less computationally intensive.
* From the visualization, it can be inferred that the model is overfitting from C = 1000

### 

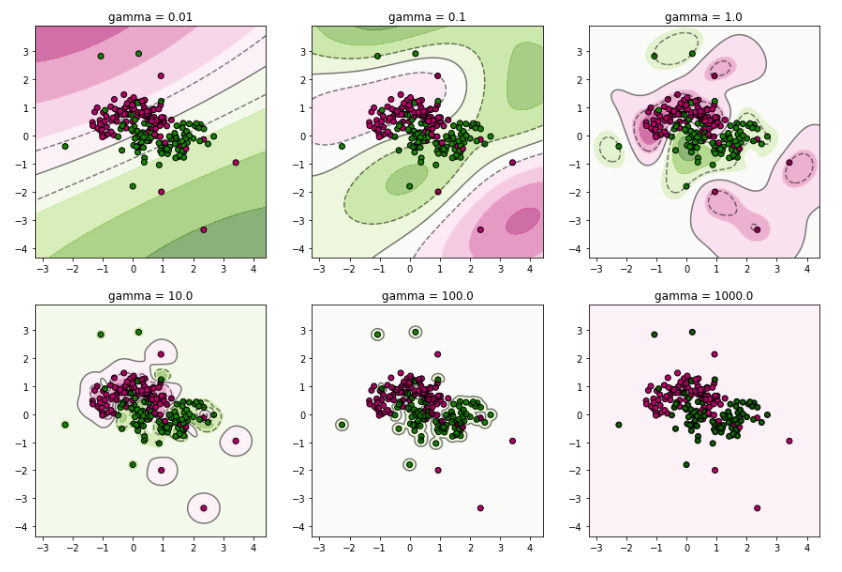
### The effect of the RBF kernel parameter, 𝛾

### 

gamma\_best : 1.0

test error : 0.16666666666666663

* Similar to the previous case, Train error decreases with an increase in gamma value.
* Validation error decreases until a point and then increases drastically compared to the slope (growth rate) seen in changing c\_value
* Least validation error is obtained at gamma = [1, 10]
* From the visualization, it can be inferred that the model is starting to overfitting from gamma = 10



**SVM with breast cancer data:**

validation Error :

[[0.36842105 0.36842105 0.07017544 0.06140351 0.04385965 0.02631579]

[0.36842105 0.36842105 0.07894737 0.04385965 0.03508772 0.03508772]

[0.36842105 0.36842105 0.36842105 0.36842105 0.36842105 0.36842105]

[0.36842105 0.36842105 0.36842105 0.36842105 0.36842105 0.36842105]

[0.36842105 0.36842105 0.36842105 0.36842105 0.36842105 0.36842105]

[0.36842105 0.36842105 0.36842105 0.36842105 0.36842105 0.36842105]

[0.36842105 0.36842105 0.36842105 0.36842105 0.36842105 0.36842105]]

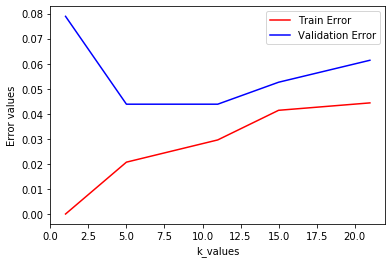
c\_best : 0.001

gamma\_best : 1000

test error : 0.368421052631579

* Each row contains data for a particular ‘c\_value’. C varies from 10^-2 to 10^4
* Each column contains data for a particular ‘gamma value’. Gamma varies from 10-3 to 103
* Least value in the table was found, and the corresponding c and gamma value are considered to be c\_best and gamma\_best to test the test set. C\_best = 0.001 and gamma\_best = 1000

**KNN with breast cancer data:**

****

k\_best : 5

test error : 0.03508771929824561

* Train error increases with an increase in k value but Validation error decreases first, and it is consistent and then increases
* Least validation error is obtained when it is consistent (where k = 5 and k = 11)
* test error for SVM is approximately 0.3684, and test error for KNN is approximately 0.0350.
  + KNN gives 10% less error than SVM
  + Best model in KNN uses k = 5 but Best model in SVM uses gamma = 1000 & c = 0.001 (so, I think KNN is computationally less intense than SVM)
* So, I prefer KNN than SVM for this case