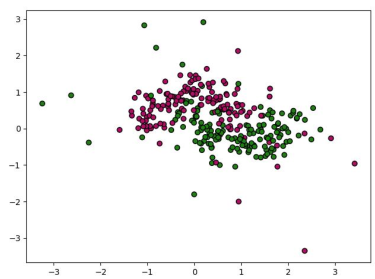
SUPPORT VECTOR MACHINES

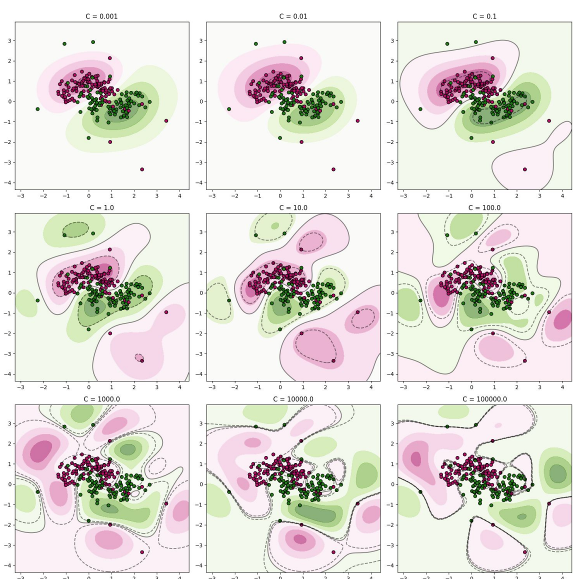
1. Support Vector Machines with Synthetic Data

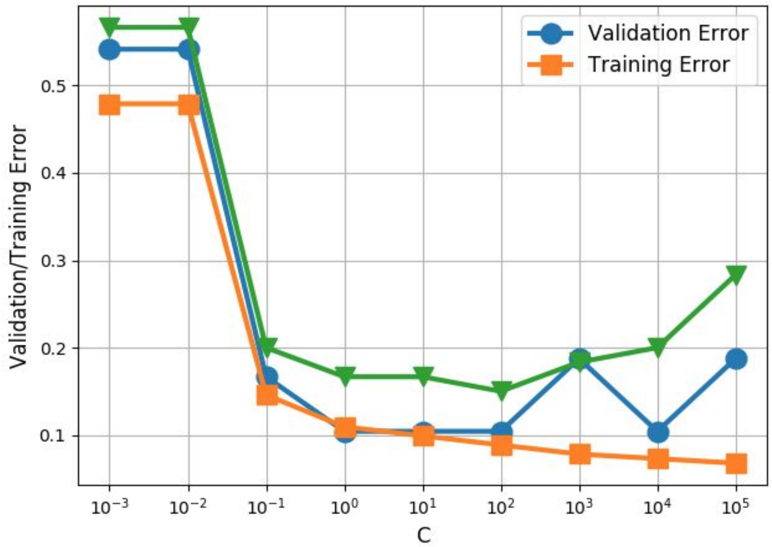
Behavior of SVMs with Radial-Basis Function (RBF) kernels with different values of C and γ



a) Effect of Regularization Parameter, C

Plot





Final Model Selection

C_{best}	1.0
Final Test Set Accuracy for C_{best}	0.8333333333333334

Discussion

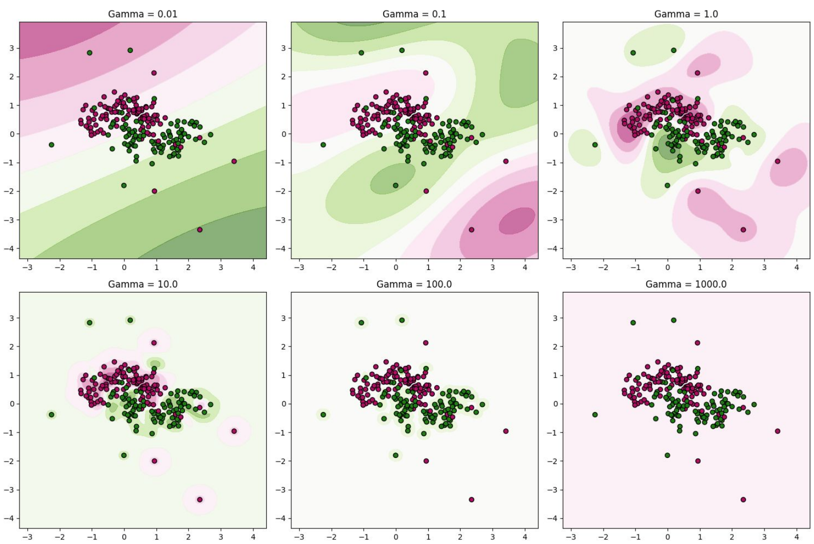
We notice that on increasing the value of C, Validation Error and Training error decreases initially. Lower value of C encourages a larger margin, training accuracy decreases. When the value of C is the lowest the validation and training errors are the highest. Value of error is decreasing till C = 1 and then validation error keeps on decreasing as C increases. By visualizing the model, for lower values of C, like 0.001, 0.01 and 0.1, the model is simple whereas the bias is high, it underfits. For higher values of C, like 1000, 10000 and 100000, the model becomes really complex, variance is high, it overfits. When the value of C =1, we see a trade-off between bias and variance.

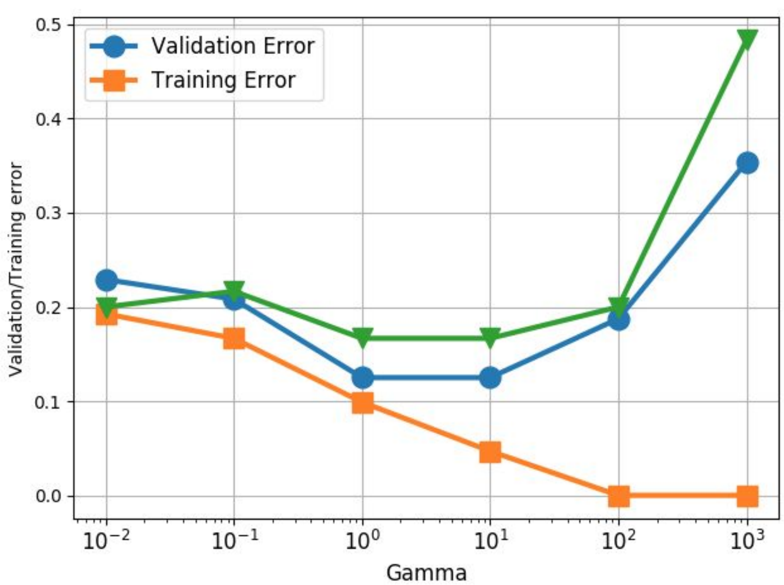
When C = 1, accuracy is more.

Objective function of SVM is given by $1/2(\mathbf{w}'\mathbf{w})+C\,\Sigma\ell(\mathbf{w}|\mathbf{x}i,yi)$; i ranges from 1 to n; $\mathbf{x}i$ -> training examples; yi -> the training labels; $\mathbf{w}'\mathbf{w}$ is the regularization term. Regularization is a technique to avoid overfitting. $C\,\Sigma\ell(\mathbf{w}|\mathbf{x}i,yi)$ is the hinge loss term. In soft margin SVM, it penalizes misclassifications. In hard margin SVM there are no misclassifications (by definition).

b) Effect of RBF Kernel Parameter, γ

Plot





Final Model Selection

γ_{best}	1.0
Final Test Set Accuracy for γ_{best}	0.8333333333333334

Discussion

We notice that on increasing the value of gamma, Validation Error and Training error decreases initially. Value of error is decreasing till gamma = 1.By visualizing the model, for lower values of γ, like 0.001,0.01, the model is very simple, the bias is high and it underfits. For higher values of γ like 1000, 10000, 100000, the model becomes highly complex, it tries to fit all the data points and it also fits noise and variance is high. When γ value is 1, we see a trade-off between bias and variance.

When γ=1, accuracy is more.

RBF kernel: $\kappa(\mathbf{x},\mathbf{z})$=exp($-\gamma\cdot\ \|\mathbf{x}-\mathbf{z}\|$ 2) $\|\mathbf{x}-\mathbf{z}\|$ 2 is the squared Euclidean distance between the two feature vectors. γ is the regularization factor.

2. Breast Cancer Diagnosis with Support Vector Machines

Final Model Selection

C_{best}, γ_{best}	(100.0 , 0.01)
Final Test Set Accuracy for C_{best}, γ_{best}	0.9652173913043478

Tables: Training Error , Validation Error

C , γ **, Training Error**
{ (0.01, 0.001): 0.37168141592920356, (0.01, 0.01): 0.37168141592920356, (0.01, 0.1): 0.37168141592920356, (0.01, 1.0): 0.37168141592920356, (0.01, 10.0): 0.37168141592920356, (0.01, 100.0): 0.37168141592920356, (0.1, 0.001): 0.30678466076696165, (0.1, 0.01): 0.05014749262536877, (0.1, 0.1): 0.035398230088849563, (0.1, 1.0): 0.37168141592920356, (0.1, 10.0): 0.37168141592920356, (0.1, 100.0): 0.37168141592920356, (1.0, 0.001): 0.0471976401179941, (1.0, 0.01): 0.029498525073746285, (1.0, 0.1): 0.011799410029498858, (1.0, 1.0): 0.0,

```
 (1.0, 10.0): 0.0,
 (1.0, 100.0): 0.0,
 (10.0, 0.001): 0.02654867256637l723,
 (10.0, 0.01): 0.011799410029498S8,
 (10.0, 0.1): 0.0,
 (10.0, 1.0): 0.0,
 (10.0, 10.0): 0.0,
 (10.0, 100.0): 0.0,
 (100.0, 0.001): 0.014749262536873142,
 (100.0, 0.01): 0.002949852507374673,
 (100.0, 0.1): 0.0,
 (100.0, 1.0): 0.0,
 (100.0, 10.0): 0.0,
 (100.0, 100.0): 0.0,
 (1000.0, 0.001): 0.005899705014749235,
 (1000.0, 0.01): 0.0,
 (1000.0, 0.1): 0.0,
 (1000.0, 1.0): 0.0,
 (1000.0, 10.0): 0.0,
 (1000.0, 100.0): 0.0,
 (10000.0, 0.001): 0.0,
 (10000.0, 0.01): 0.0,
 (10000.0, 0.1): 0.0,
 (10000.0, 1.0): 0.0,
 (10000.0, 10.0): 0.0,
 (10000.0, 100.0): 0.0 }
```

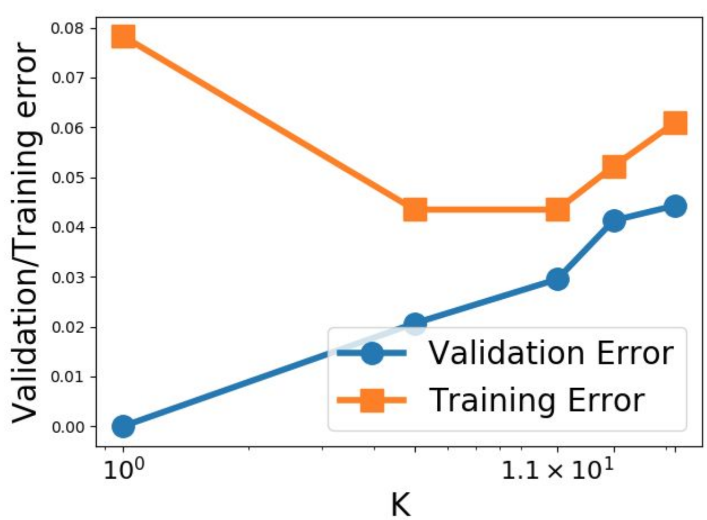
C , γ , Validation Error

```
{ (0.01, 0.001): 0.3739130434782608,
 (0.01, 0.01): 0.3739130434782608,
 (0.01, 0.1): 0.3739130434782608,
 (0.01, 1.0): 0.3739130434782608,
 (0.01, 10.0): 0.3739130434782608,
 (0.01, 100.0): 0.3739130434782608,
 (0.1, 0.001): 0.30434782608695654,
 (0.1, 0.01): 0.06956521739130439,
 (0.1, 0.1): 0.0782608695652l738,
 (0.1, 1.0): 0.3739130434782608,
 (0.1, 10.0): 0.3739130434782608,
 (0.1, 100.0): 0.3739130434782608,
 (1.0, 0.001): 0.060869565217391286,
 (1.0, 0.01): 0.060869565217391286,
 (1.0, 0.1): 0.04347826086956519,
 (1.0, 1.0): 0.3739130434782608,
 (1.0, 10.0): 0.3739130434782608,
```

```
(1.0, 100.0): 0.3739130434782608,
(10.0, 0.001): 0.034782608695652195,
(10.0, 0.01): 0.04347826086956519,
(10.0, 0.1): 0.034782608695652195,
(10.0, 1.0): 0.3739130434782608,
(10.0, 10.0): 0.3739130434782608,
(10.0, 100.0): 0.3739130434782608,
(100.0, 0.001): 0.034782608695652195,
(100.0, 0.01): 0.02608695652173909,
(100.0, 0.1): 0.034782608695652195,
(100.0, 1.0): 0.3739130434782608,
(100.0, 10.0): 0.3739130434782608,
(100.0, 100.0): 0.3739130434782608,
(1000.0, 0.001): 0.034782608695652195,
(1000.0, 0.01): 0.02608695652173909,
(1000.0, 0.1): 0.034782608695652195,
(1000.0, 1.0): 0.3739130434782608,
(1000.0, 10.0): 0.3739130434782608,
(1000.0, 100.0): 0.3739130434782608,
(10000.0, 0.001): 0.02608695652173909,
(10000.0, 0.01): 0.02608695652173909,
(10000.0, 0.1): 0.034782608695652195,
(10000.0, 1.0): 0.3739130434782608,
(10000.0, 10.0): 0.3739130434782608,
(10000.0, 100.0): 0.3739130434782608 }
```

3. Breast Cancer Diagnosis with K-Nearest Neighbors

Plot



K, Training Error, Validation Error
1   0.0   0.07826086956521738
5   0.020648967551622377   0.04347826086956519
11   0.029498525073746285   0.04347826086956519
15   0.041297935103244865   0.05217391304347829
21   0.04424778761061943   0.060869565217391286

Final Model Selection

K_{best}	1
Final Test Set Accuracy for K_{best}	0.9478260869565217

Discussion: SVM vs KNN

SVM	0.9652173913043478
KNN	0.9478260869565217

We can observe that whenever training data size increases, SVM performs better than kNN and it also has got more accuracy comparatively. kNN seems to be a good classifier at the same time its performance depends on the value of k. It does not give good results when the value of k is small and we would get better results as the k value grows higher.

SVM is better suited for this classification task