SMAI Assignment 2

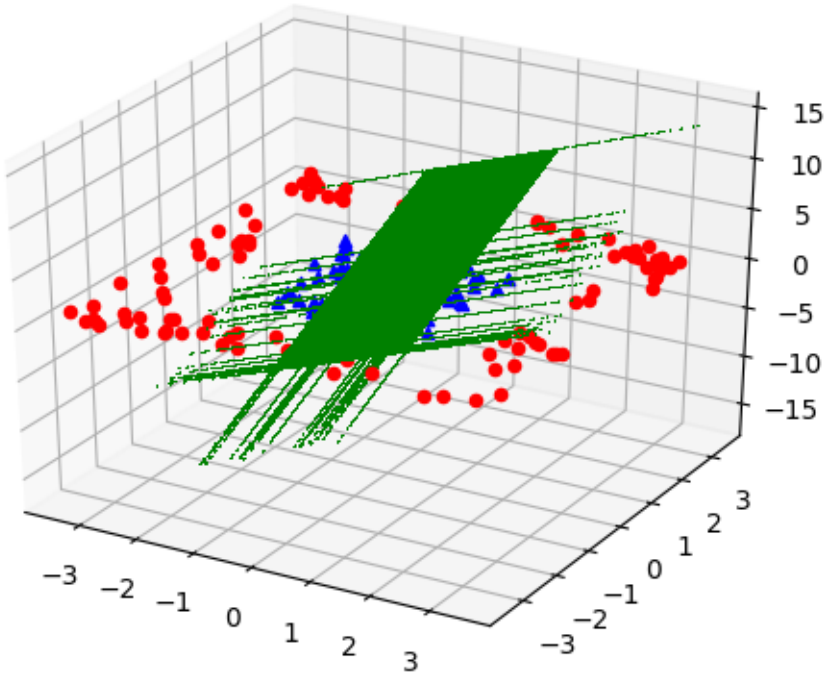
Question 1 Report

Chaitanya Patel

201501071

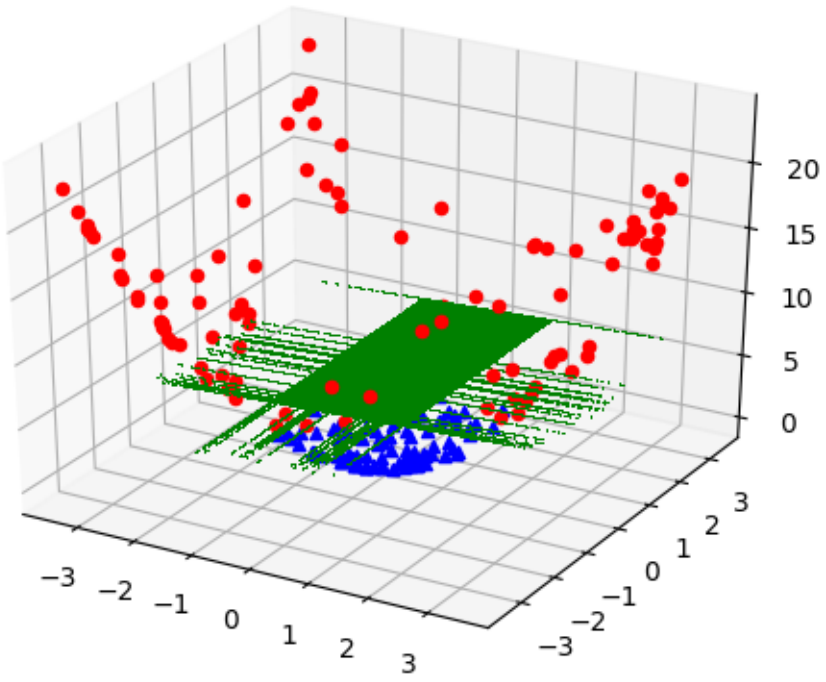
Perceptron Kernel Trick

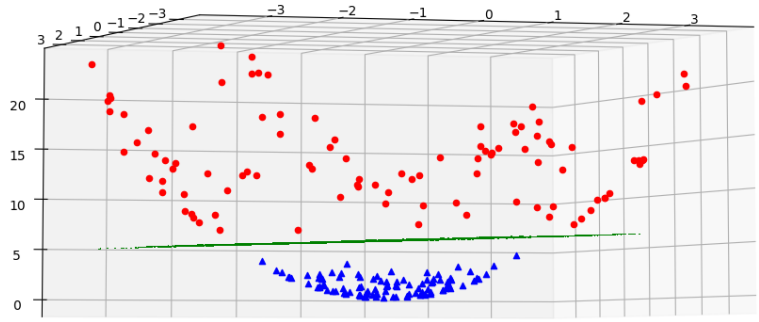
Raw Data in 3D : (accuracy : 0.585)



Polynomial Kernel :

(x, y, x^2 + y^2) : (accuracy : 1.000)


20
15
10
5
0
−3
−2
−1
0
1
2
3

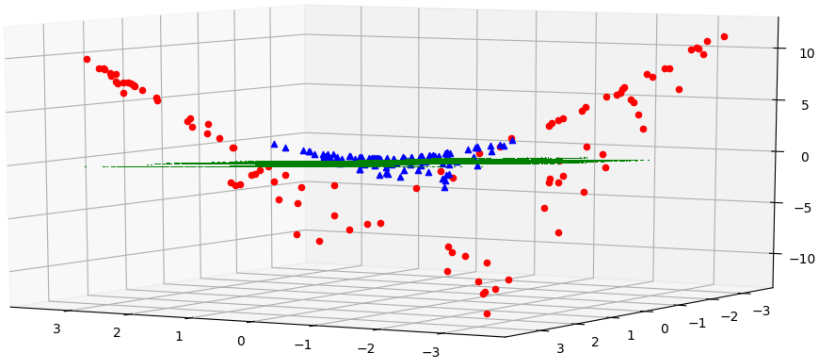

20
15
10
5
0

(x, y, x^2 + y^2 + 1) : (accuracy : 1.000)

(x, y, 2*x^2 + y^2) : (accuracy : 0.900)

Multiply : (x, y, a*x*y)

(x, y, x*y) : (accuracy : 0.580)



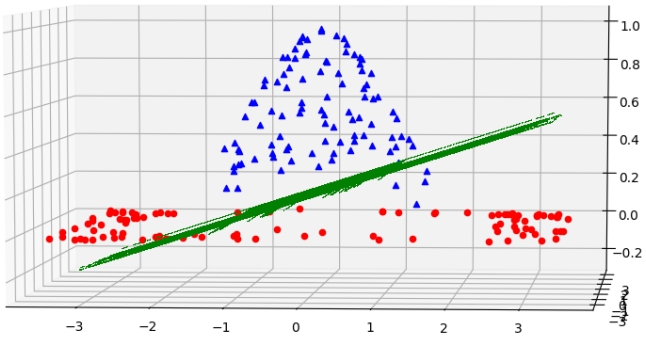
(x, y, 2*x*y) : (accuracy : 0.635)

(x, y, 3*x*y) : (accuracy : 0.530)

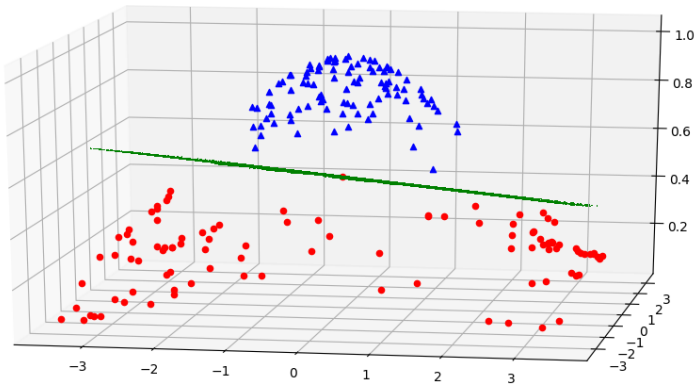
Radial Basis Kernel (Gaussian Kernel)

(x, y, exp(- (x^2 + y^2) / (2*sigma^2)))

Sigma : 1.0 (accuracy : 0.760)



Sigma : 2.0 (accuracy : 1.000)



Sigma : 4.0 (accuracy : 0.825)

SVM

One to all SVM (LinearSVC in sklearn)

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7101

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7079

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7111

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7079

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7077

Accuracy : 0.6996

'Precision'

array([0.80765302, 0.57056424, 0.61243728, 0.65645098, 0.61887116,
        0.75279945, 0.5459108 , 0.49111111, 0.80121291, 0.8725526 ,
        0.62190982, 0.73544901, 0.81011679, 0.79643996, 0.70954483,
        0.76281709, 0.60690178, 0.59044123, 0.52758842, 0.79380972,
        0.77461405, 0.79649795, 0.8104024 , 0.54318482, 0.7221267 ,
        0.7180371])

'Recall'

array([0.86835443, 0.77468354, 0.76486486, 0.765   , 0.38441558,
        0.60769231, 0.21842105, 0.18356164, 0.81891892, 0.78933333,
        0.51081081, 0.74133333, 0.88148148, 0.74545455, 0.6961039 ,
        0.83544304, 0.76623377, 0.768   , 0.39726027, 0.81481481,
        0.84938272, 0.73589744, 0.90649351, 0.70512821, 0.785   ,
        0.8   ])

'F1'

array([0.83631026, 0.65587603, 0.67960672, 0.70601307, 0.47382724,
0.67224599, 0.31080251, 0.26321985, 0.80910841, 0.82708766,
0.56071342, 0.73795978, 0.84408401, 0.76871114, 0.70253669,
0.79736917, 0.67707028, 0.66647139, 0.4509526 , 0.80397883,
0.80979782, 0.76390188, 0.85539379, 0.61300583, 0.75047051,
0.75518613])

SVC (one to one)

Linear

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7845

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7830

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7851

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7819

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7834

Accuracy : 0.7785

'Precision'

array([0.92460245, 0.5331464 , 0.89169109, 0.71650011, 0.70809173,
0.78929083, 0.68683325, 0.54996943, 0.95428456, 0.95167162,
0.71053718, 0.98983051, 0.82169789, 0.89366995, 0.74455666,
0.93665504, 0.82866009, 0.63746321, 0.51797685, 0.848902 ,
0.90122147, 0.81014633, 0.86020991, 0.78293148, 0.79721201,
0.88641564])

'Recall'

array([0.87341772, 0.88607595, 0.75135135, 0.7975   , 0.72987013,

0.77692308, 0.70789474, 0.49589041, 0.85945946, 0.81066667,
0.70810811, 0.75466667, 0.88641975, 0.85714286, 0.74285714,
0.78734177, 0.76103896, 0.776 , 0.6 , 0.8 ,
0.85925926, 0.87435897, 0.9012987 , 0.81025641, 0.7025 ,
0.68333333])

'F1'

array([0.89801617, 0.66449669, 0.81530842, 0.75391871, 0.71811624,
0.78212636, 0.69668049, 0.51958427, 0.90257695, 0.87321955,
0.7076291 , 0.85602035, 0.85264051, 0.87378118, 0.74284576,
0.85544374, 0.79244762, 0.69950531, 0.55319734, 0.8230468 ,
0.87894155, 0.84086211, 0.87957508, 0.79626084, 0.74583506,
0.7685817])

Sigmoid

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.4991

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.5031

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.4980

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.5029

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.5056

Accuracy : 0.4932

'Precision'

array([0.71948677, 0.16113925, 0.53590338, 0.27109982, 0.33627588,
0.43339372, 0.52666667, 0. , 0.8454951 , 0.90387067,
0.27295726, 1. , 0.65042997, 0.65053904, 0.49561732,
0.74501997, 0.45611011, 0.65 , 0.83333333, 0.52461277,

0.9303891 , 0.69231703, 0.70473879, 0.23833983, 0.66946913,
0.98888889])

'Recall'

array([0.88101266, 0.5721519 , 0.46756757, 0.79 , 0.2961039 ,
0.64615385, 0.02894737, 0. , 0.7972973 , 0.70666667,
0.08648649, 0.656 , 0.84444444, 0.54545455, 0.2987013 ,
0.74177215, 0.68311688, 0.016 , 0.02465753, 0.4691358 ,
0.76296296, 0.65384615, 0.78441558, 0.57179487, 0.1275 ,
0.23888889])

'F1'

array([0.79193426, 0.25139438, 0.49913958, 0.40342545, 0.31422189,
0.51811178, 0.05452007, 0. , 0.82010392, 0.792821 ,
0.1310104 , 0.79084313, 0.73420858, 0.59335849, 0.36979231,
0.74262441, 0.54637508, 0.03091099, 0.04785965, 0.49381989,
0.83797692, 0.67103828, 0.74156683, 0.33623739, 0.210679 ,
0.38226779])

Polynomial Degree 2

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.0484

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.0469

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.0504

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.0492

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.0486

Accuracy : 0.05

'Precision'

array([0. , 0. , 0. , 0. , 0. ,
        0. , 0. , 0. , 0. , 0. ,
        0. , 0. , 0.04101688, 0. , 0. ,
        0. , 0. , 0. , 0. , 0.75424383,
        0. , 0. , 0. , 0. , 0. , 0.])

'Recall'

array([0. , 0. , 0. , 0. , 0. ,
        0. , 0. , 0. , 0. , 0. ,
        0. , 0. , 1. , 0. , 0. ,
        0. , 0. , 0. , 0. , 0.2345679,
        0. , 0. , 0. , 0. , 0. , 0.])

'F1'

array([0. , 0. , 0. , 0. , 0. ,
        0. , 0. , 0. , 0. , 0. ,
        0. , 0. , 0.07880156, 0. , 0. ,
        0. , 0. , 0. , 0. , 0.35717622,
        0. , 0. , 0. , 0. , 0. , 0.])

Polynomial Degree 4, Bias 1

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6996

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6991

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.7024

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6994

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6994

Accuracy : 0.6931

'Precision'

array([0.85765626, 0.4278185 , 0.79522943, 0.61578474, 0.67844571,
0.67966639, 0.51465441, 0.35546991, 0.95858112, 0.9451742 ,
0.61265265, 0.98965312, 0.75124585, 0.75843044, 0.64412208,
0.9080199 , 0.717961 , 0.64532893, 0.40046629, 0.7829618 ,
0.80051919, 0.75979097, 0.82637981, 0.44907494, 0.74904071,
0.8712548])

'Recall'

array([0.86835443, 0.8835443 , 0.73513514, 0.755   , 0.37142857,
0.75384615, 0.57368421, 0.15616438, 0.7972973 , 0.79466667,
0.51891892, 0.73333333, 0.87654321, 0.83896104, 0.73246753,
0.7164557 , 0.6987013 , 0.74133333, 0.39726027, 0.72098765,
0.83703704, 0.85128205, 0.87012987, 0.58717949, 0.4875   ,
0.65555556])

'F1'

array([0.86282503, 0.57609319, 0.76349827, 0.67717603, 0.4791792 ,
0.71452518, 0.54213434, 0.21587886, 0.86961665, 0.86304871,
0.56154874, 0.8413498 , 0.80873852, 0.79596629, 0.68512474,
0.79989592, 0.70808287, 0.68912692, 0.39751952, 0.74999055,
0.81723156, 0.80260987, 0.8468115 , 0.50838035, 0.58901989,
0.74739207])

RBF gamma = 1 / #features

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6191

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6161

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6234

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6171

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6213

Accuracy : 0.6204

'Precision'

array([0.79066122, 0.31856732, 0.62121997, 0.50361445, 0.48178571,
0.55322504, 0.43704269, 0. , 0.95740839, 0.9192482 ,
0.35569997, 0.9929198 , 0.70033952, 0.67962227, 0.57586487,
0.83281565, 0.64834908, 0.76156905, 0.5150564 , 0.67455266,
0.78399577, 0.73056988, 0.77381723, 0.32580149, 0.76923915,
0.96405041])

'Recall'

array([0.87088608, 0.90632911, 0.73243243, 0.75 , 0.12467532,
0.75384615, 0.56315789, 0. , 0.77567568, 0.76 ,
0.18918919, 0.71733333, 0.8691358 , 0.75584416, 0.76363636,
0.7164557 , 0.60779221, 0.49866667, 0.22739726, 0.5654321 ,
0.80987654, 0.81794872, 0.86233766, 0.55897436, 0.255 ,
0.58888889])

'F1'

array([0.82882603, 0.47132272, 0.67152964, 0.60152904, 0.19485813,
0.63793352, 0.49180649, 0. , 0.85633142, 0.83196867,
0.24596272, 0.8319574 , 0.77484746, 0.71490121, 0.65622599,
0.76920735, 0.62691169, 0.59938714, 0.31477742, 0.61438981,
0.79606913, 0.77144871, 0.81486108, 0.41090523, 0.37951269,
0.73053847])

RBF gamma = 1

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.8256

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.8243

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.8287

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.8247

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.8256

Accuracy : 0.8174

'Precision'

array([0.931645 , 0.60966565, 0.91583102, 0.74755048, 0.75930761,
0.80470656, 0.74445979, 0.6364611 , 0.95699392, 0.94956288,
0.74659847, 0.98134463, 0.87038864, 0.94370213, 0.75929702,
0.94425314, 0.88626361, 0.66714385, 0.6220084 , 0.88904304,
0.91458058, 0.85956638, 0.84562417, 0.81881204, 0.84206198,
0.89182875])

'Recall'

array([0.88607595, 0.87594937, 0.76486486, 0.8325 , 0.77402597,
0.8 , 0.77105263, 0.54520548, 0.87027027, 0.824 ,
0.78378378, 0.808 , 0.89382716, 0.86753247, 0.82337662,
0.81012658, 0.81558442, 0.784 , 0.69041096, 0.83950617,
0.87901235, 0.89487179, 0.92727273, 0.87948718, 0.7775 ,
0.79722222])

'F1'

array([0.90743939, 0.71819672, 0.83347246, 0.78746489, 0.76562999,
0.80203919, 0.75674587, 0.58571227, 0.91024792, 0.88074593,
0.76396016, 0.88582576, 0.88188414, 0.90300287, 0.78871207,
0.87188296, 0.84914141, 0.72030394, 0.65144067, 0.86301191,
0.89598443, 0.8764549 , 0.88378686, 0.84722493, 0.80779365,
0.84014392])

RBF gamma = 0.1

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6641

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6652

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6674

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6644

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.6664

Accuracy : 0.6635

'Precision'

array([0.82892291, 0.39379901, 0.75984089, 0.56731207, 0.66798663,
        0.62165683, 0.46722928, 0.21469697, 0.97345665, 0.93548603,
        0.49643195, 0.98953418, 0.72946079, 0.71366612, 0.62899148,
        0.8635423 , 0.696491 , 0.64616916, 0.38756287, 0.73282495,
        0.78206512, 0.75350545, 0.80142358, 0.39420707, 0.72605172,
        0.91188173])

'Recall'

array([0.87088608, 0.9164557 , 0.72972973, 0.7475   , 0.31168831,
        0.77435897, 0.56315789, 0.02739726, 0.78648649, 0.78133333,
        0.36756757, 0.72533333, 0.8691358 , 0.82857143, 0.75844156,
        0.7164557 , 0.65974026, 0.632   , 0.33424658, 0.6345679 ,
        0.82716049, 0.84615385, 0.87012987, 0.57692308, 0.3725   ,
        0.64444444])

'F1'

array([0.84928872, 0.55035282, 0.74363393, 0.64412164, 0.42383249,
        0.68951969, 0.51006595, 0.04844103, 0.86933761, 0.85117611,
        0.42186501, 0.83615084, 0.7928224 , 0.76640301, 0.68712793,

0.78231553, 0.67735816, 0.63796998, 0.35689625, 0.67950171,
0.80295992, 0.79686813, 0.83350147, 0.46762625, 0.49075377,
0.75439407])

RBF gamma = 10

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9554

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9543

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9558

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9542

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9547

Accuracy : 0.9394

'Precision'

array([0.98737261, 0.83292406, 0.98276225, 0.88295609, 0.90273998,
0.90473785, 0.88255336, 0.90043165, 0.97526115, 0.97303113,
0.93411331, 0.9939819 , 0.97004985, 0.96585955, 0.9222206 ,
0.98355753, 0.93767182, 0.82120188, 0.9471329 , 0.97275713,
0.97051635, 0.97617255, 0.96455287, 0.93036117, 0.98012039,
0.98626126])

'Recall'

array([0.98481013, 0.94177215, 0.91891892, 0.9375 , 0.91428571,
0.94358974, 0.94210526, 0.83287671, 0.94864865, 0.952 ,
0.9 , 0.88 , 0.95555556, 0.94545455, 0.96623377,
0.90632911, 0.96623377, 0.87733333, 0.91506849, 0.95555556,
0.97283951, 0.94102564, 0.98441558, 0.95897436, 0.9825 ,
0.98611111])

'F1'

array([0.98606758, 0.88381332, 0.94960218, 0.90928419, 0.90837483,
0.92351901, 0.91078567, 0.86497561, 0.96123989, 0.96224748,
0.91601944, 0.93329801, 0.96263558, 0.95532027, 0.94334145,
0.94328715, 0.95155081, 0.84815376, 0.93045119, 0.96392198,
0.97162596, 0.9582185 , 0.97434169, 0.94441334, 0.98125699,
0.98611004])

RBF gamma = 100

Using ./letter_classification_train.data as the dataset file

Fold1 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9981

Fold2 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9986

Fold3 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9987

Fold4 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9984

Fold5 -> Number of training samples: 14000 | Number of testing samples: 2000

Training Accuracy: 0.9989

Accuracy : 0.9643

'Precision'

array([0.98496835, 0.94550535, 0.98102894, 0.92628371, 0.95401957,
0.946015 , 0.9453765 , 0.90331429, 0.97582824, 0.97349615,
0.93436675, 0.96875739, 0.98301499, 0.97111267, 0.94508228,
0.96411235, 0.96419497, 0.93460456, 0.99722222, 0.98274466,
0.99015128, 0.97941181, 0.97964281, 0.97252693, 0.98496676,
0.99459459])

'Recall'

array([0.98987342, 0.96202532, 0.95135135, 0.97 , 0.95844156,

0.95897436, 0.95526316, 0.89863014, 0.96486486, 0.968 ,
0.93243243, 0.97866667, 0.99012346, 0.94805195, 0.96883117,
0.93417722, 0.97922078, 0.91466667, 0.97260274, 0.98271605,
0.98024691, 0.96410256, 0.98701299, 0.98974359, 0.9825 ,
0.98055556])

'F1'

array([0.98740546, 0.95359596, 0.96569841, 0.94757944, 0.95603762,
0.95173327, 0.95012861, 0.90065167, 0.97003949, 0.97055287,
0.93267797, 0.97346677, 0.98651775, 0.95919214, 0.95654365,
0.94851771, 0.97161194, 0.92436352, 0.9847314 , 0.98270034,
0.98509202, 0.97157495, 0.98324063, 0.9810198 , 0.98369447,
0.98741097])