Part 1 : Linear SVM

classification result

Linear SVM with different penalty weight	classification rate (testing data = 50)
C = 1.0	0.90
C = 10.0	0.78
C = 100.0	0.92

alpha and optimal model bias

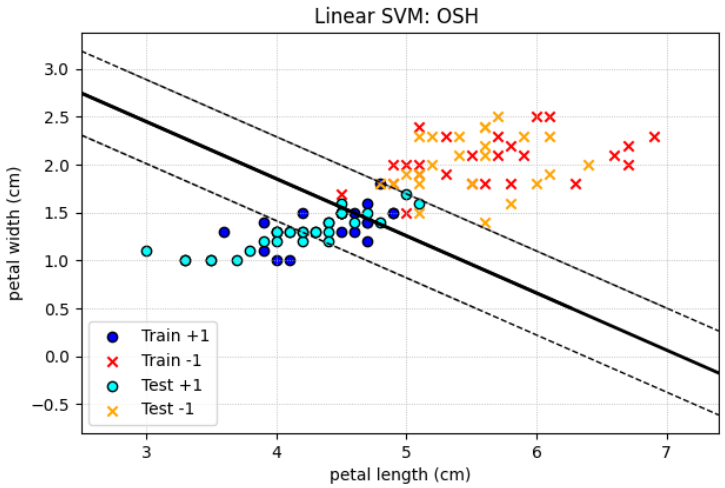
```
α of penalty weight = 1.0 :
 [ 1.0008  0.0572  1.0008  0.0012  1.0009  0.0009  1.0008  0.0018  0.0009
  0.0013  0.0016  0.0011  0.0012  1.0008  0.0015  0.0010  0.0572  0.0011
  0.0572  0.0012  1.0008  0.0012  1.0008  0.0008  0.0010 -0.0005  0.9993
 -0.0005 -0.0006 -0.0005 -0.0004  0.9991 -0.0004 -0.0005 -0.0005  0.9993
  0.2086 -0.0006  0.9993 -0.0007 -0.0006 -0.0006 -0.0004 -0.0003  0.9993
 -0.0005  0.9992 -0.0004  0.9992 -0.0005]
b of sigma = 1.0 :  9.674873563960059
classification rate of linear SVM with penalty weight of 1.0:  0.9
```

```
α of penalty weight = 10.0 :
 [-0.0009 -0.0010  9.0313 -0.0014 -0.0009 -0.0010  9.9991 -0.0025 -0.0010
 -0.0015 -0.0021 -0.0012 -0.0015 -0.0009 -0.0019 -0.0011 -0.0010 -0.0014
 -0.0010 -0.0016  9.9992 -0.0014  9.0313 -0.0009 -0.0012  0.0003  8.0218
  0.0004  0.0005  0.0004  0.0002 10.0010  0.0003  0.0004  0.0003  0.0006
  0.0006  0.0005  0.0007  0.0006  0.0005  0.0005  0.0002  0.0002 10.0007
  0.0004  0.0007  0.0002 10.0007  0.0004]
b of sigma = 10.0 :  16.711647621172343
classification rate of linear SVM with penalty weight of 10.0:  0.78
```

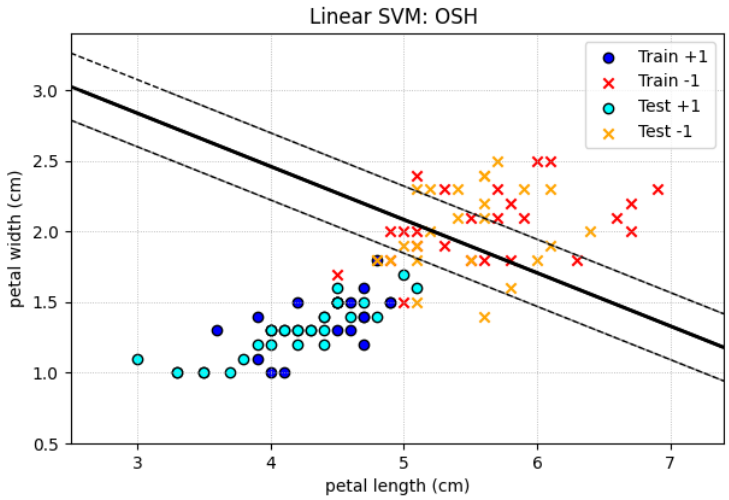
```
α of penalty weight = 100.0 :
 [ -0.0146   0.0005  44.4725   0.0097  -0.0038  -0.0145  99.9991   0.0306
  -0.0182   0.0253   0.0128   0.0153  -0.0188  -0.0146   0.0384  -0.0026
   0.0005  -0.0234   0.0005  -0.0035 100.0079   0.0097  44.4725  -0.0288
  -0.0060  -0.0043  -0.0026   0.0097   0.0167   0.0030   0.0223  99.9855
   0.0295   0.0209  -0.0019  -0.0080   0.0037  -0.0004  -0.0117  -0.0283
  -0.0161   0.0143   0.0201   0.0196 100.0175  -0.0039  -0.0156   0.0272
  88.8445   0.0049]
b of sigma = 100.0 :  12.956177049965795
classification rate of linear SVM with penalty weight of 100.0:  0.92
```

optimal separation hyperplane visualization

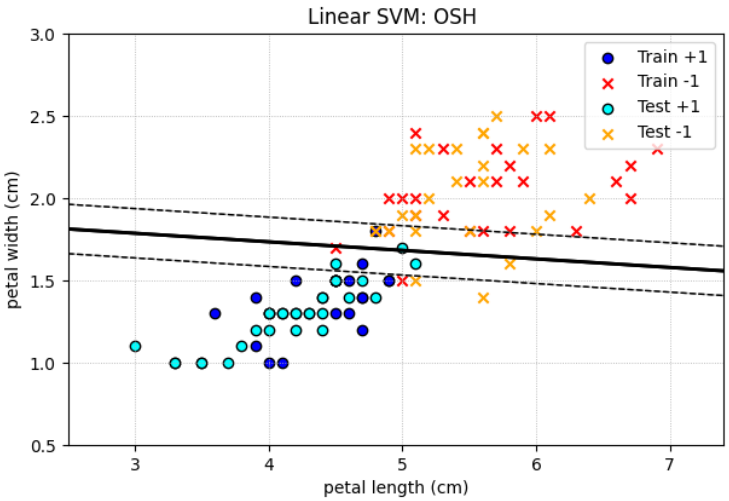
penalty weight=1.0



penalty weight=10.0



penalty weight=100.0



Part 2 : RBF kerneal-based SVM

classification result

RBF kenel-based SVM with different sigma	classification rate (testing data = 50)
σ = 5.0	0.92
σ = 1.0	0.96
σ = 0.5	0.96
σ = 0.1	0.88
σ = 0.05	0.82

alpha and optimal model bias

```
α of sigma = 5 :
 [ 9.9990  8.9773  9.9992 -0.0018  9.9989 -0.0013  9.9991 -0.0026 -0.0012
 -0.0018 -0.0024 -0.0015 -0.0020  9.9990 -0.0022 -0.0013  8.9773 -0.0019
  8.9773 -0.0020  9.9993 -0.0018  9.9992 -0.0012 -0.0015 -0.0009 10.0003
 -0.0006  6.6047 -0.0005 -0.0012 10.0010 -0.0007 -0.0003 -0.0010 10.0002
 10.0001 -0.0002 10.0003  0.3015 -0.0001 10.0000 -0.0013 -0.0015 10.0007
 -0.0005 10.0004 -0.0012 10.0006 -0.0004]
b of sigma = 5 :  0.28033255481659025
classification rate of RBF SVM with sigma of 5:  0.92
```

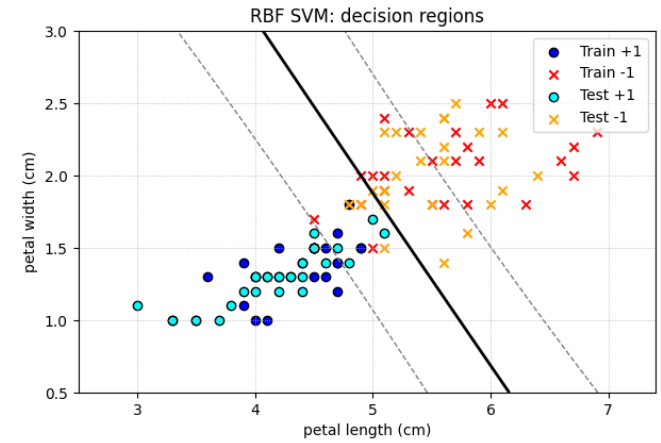
```
α of sigma = 1 :
 [-0.0002 -0.0008  8.9188 -0.0007 -0.0007 -0.0003  9.9993  0.3298 -0.0002
 -0.0008 -0.0001 -0.0010 -0.0001 -0.0002 -0.0005 -0.0007 -0.0008 -0.0000
 -0.0008 -0.0003  9.9991 -0.0007  8.9188  0.0003 -0.0006 -0.0047  8.2239
 -0.0043 -0.0027 -0.0037 -0.0070 10.0012 -0.0061 -0.0038 -0.0052  0.0001
 -0.0009 -0.0019  0.0005  0.0001 -0.0008 -0.0021 -0.0071 -0.0070  9.9997
 -0.0031  0.0008 -0.0071 10.0006 -0.0031]
b of sigma = 1 :  -0.19719189887460142
classification rate of RBF SVM with sigma of 1:  0.96
```

```
α of sigma = 0.5 :
 [-0.0003  0.0000  5.9800 -0.0007 -0.0000 -0.0004  5.4099  1.0062 -0.0004
 -0.0006 -0.0005  3.0096 -0.0013 -0.0003 -0.0006 -0.0002  0.0000 -0.0013
  0.0000 -0.0012 10.0000 -0.0007  5.9800 -0.0007 -0.0005  0.2981 -0.0009
 -0.0018 -0.0030 -0.0016  0.0002  9.9998  0.2996 -0.0027  0.1055 -0.0008
 -0.0020 -0.0022 -0.0003  0.0002 -0.0007 -0.0029  0.0001  0.5834  9.9998
 -0.0011  0.1115  0.0006  9.9998 -0.0023]
b of sigma = 0.5 :  -0.4785196195627407
classification rate of RBF SVM with sigma of 0.5:  0.96
α of sigma = 0.1 :
```

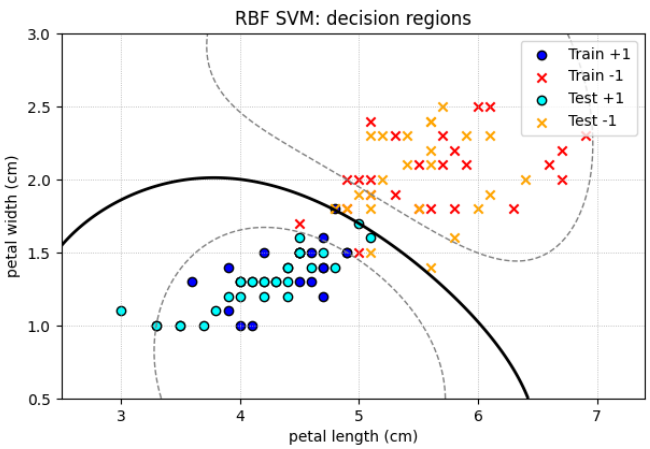
```
α of sigma = 0.05 :
 [0.5377 0.3213 0.6269 0.5453 0.9401 0.9632 1.0929 1.1103 0.9401 1.0907
 1.1103 1.1105 0.9579 0.5377 1.1106 1.0552 0.3213 0.9809 0.3213 1.0930
 1.2539 0.5453 0.6269 1.0930 1.0909 0.7834 0.7830 0.8737 0.7831 0.8414
 0.8574 0.8898 0.8894 0.8891 0.7834 0.6920 0.8890 0.8891 0.6734 0.8893
 0.8893 0.7834 0.8733 0.8893 1.0591 0.8737 0.7977 0.8734 1.0587 0.8731]
b of sigma = 0.05 :  -0.11062735461203056
classification rate of RBF SVM with sigma of 0.05:  0.82
```

optimal separation hyperplane visualization

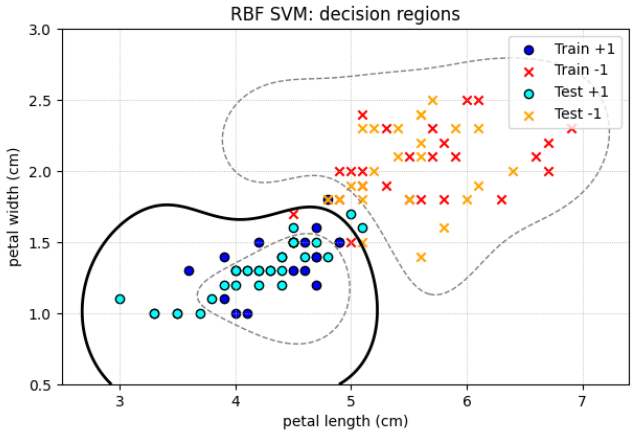
sigma=5



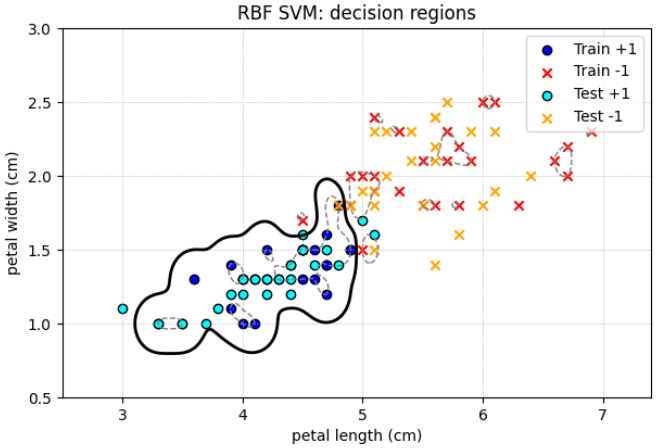
sigma=1



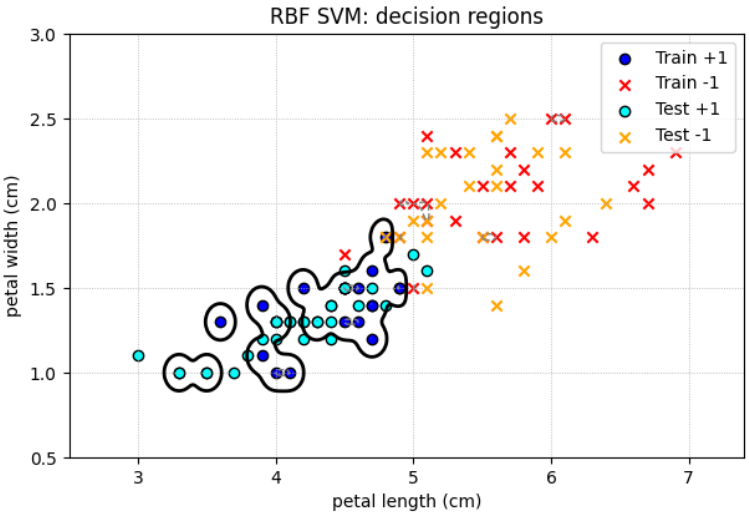
sigma=0.5



sigma=0.1



sigma=0.05



Part 3 : Polynomial kerneal-based SVM

classification result

Polynomial kenel-based SVM with different power	classification rate (testing data = 50)
p = 1	0.90
p = 2	0.94
p = 3	0.94
p = 4	0.82
p = 5	0.62

alpha and optimal model bias

```
α of power = 1 :
 [ 1.0000  0.0669  1.0000 -0.0000  1.0000 -0.0000  1.0000 -0.0000 -0.0000
 -0.0000 -0.0000 -0.0000 -0.0000  1.0000  0.0000 -0.0000  0.0668 -0.0001
  0.0668 -0.0000  1.0000 -0.0000  1.0000 -0.0000 -0.0000 -0.0000  1.0000
  0.0000  0.0000 -0.0000  0.0000  1.0000  0.0000  0.0000 -0.0000  1.0000
  0.2000 -0.0000  1.0000 -0.0000 -0.0000  0.0000  0.0000  0.0000  1.0000
 -0.0000  1.0000  0.0000  1.0000  0.0000]
b of sigma = 1 :  10.997577531808343
classification rate of polynomial SVM with power of 1:  0.9
```

```
α of power = 2 :
 [-0.0000 -0.0000  0.5004 -0.0001 -0.0000 -0.0000  1.0000 -0.0002 -0.0000
 -0.0001 -0.0002 -0.0000 -0.0001 -0.0000 -0.0001 -0.0000 -0.0000 -0.0001
 -0.0000 -0.0001  1.0000 -0.0001  0.5004 -0.0000 -0.0000 -0.0000 -0.0000
 -0.0000 -0.0000 -0.0000 -0.0000  1.0000 -0.0000 -0.0000 -0.0000 -0.0000
 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000  1.0000
 -0.0000 -0.0000 -0.0000  1.0000 -0.0000]
b of sigma = 2 :  10.862540140560572
classification rate of polynomial SVM with power of 2:  0.94
```

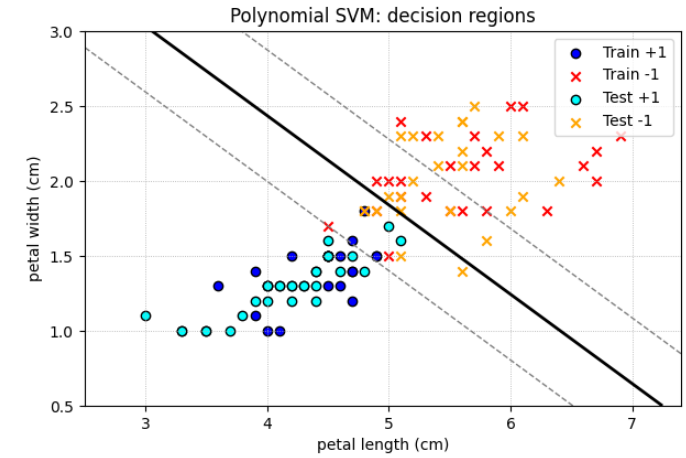
```
α of power = 3 :
 [-0.0000 -0.0001  0.3970 -0.0003 -0.0000 -0.0001  0.9146 -0.0014 -0.0001
 -0.0003 -0.0009 -0.0002 -0.0003 -0.0000 -0.0006 -0.0001 -0.0001 -0.0003
 -0.0001 -0.0004  1.0000 -0.0003  0.3970 -0.0000 -0.0001 -0.0000 -0.0000
 -0.0000 -0.0000 -0.0000 -0.0000  1.0000 -0.0000 -0.0000 -0.0000 -0.0000
 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -0.0000  1.0000
 -0.0000 -0.0000 -0.0000  0.7037 -0.0000]
b of sigma = 3 :  6.9812564459815265
classification rate of polynomial SVM with power of 3:  0.94
```

```
α of power = 4 :
 [ 0.0000 -0.0000  0.5620 -0.0000 -0.0000 -0.0000 -0.0000 -0.0002 -0.0000
 -0.0000 -0.0003  0.1723 -0.0002  0.0000  0.0228 -0.0000 -0.0000 -0.0002
 -0.0000 -0.0001  1.0000 -0.0000  0.5620  0.0313 -0.0000  0.0000  0.0000
  0.0000 -0.0000  0.0000 -0.0000  0.9094 -0.0000 -0.0000  0.0000  0.0000
  0.0000  0.0000  0.0000  0.0001  0.0001 -0.0000 -0.0000  0.0000  1.0000
  0.0000  0.0000 -0.0000  0.4394  0.0000]
b of sigma = 4 :  3.6068303178103784
classification rate of polynomial SVM with power of 4:  0.82
```

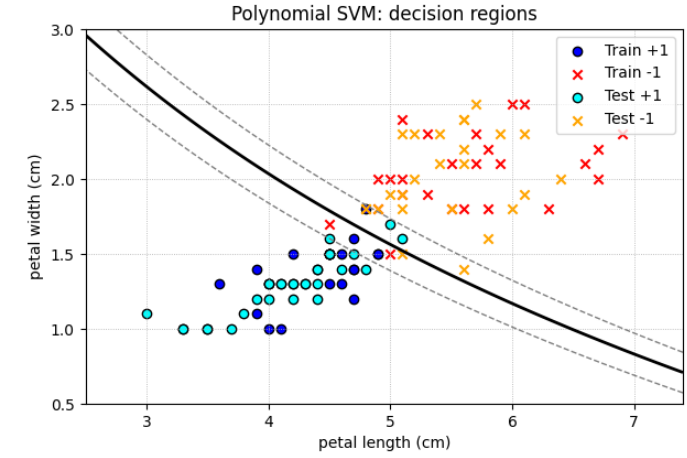
```
α of power = 5 :
 [ 0.0000  0.0000  0.0785  0.0000  0.0000  0.0000 -0.0000  0.0001  0.0000
 -0.0000  0.0001 -0.0000  0.0001  0.0000  0.0859  0.0000  0.0000  0.0001
  0.0000  0.0000  1.0000  0.0000  0.0785  0.0000  0.0000  0.0100  0.0000
  0.0000  0.0000  0.0000  0.0000  0.5619  0.0000  0.0000  0.0000 -0.0000
  0.0000  0.0000 -0.0000 -0.0000 -0.0000  0.0000  0.0000  0.0000  0.1080
 -0.0000 -0.0000  0.0000  0.5633  0.0000]
b of sigma = 5 :  28.90161155850051
classification rate of polynomial SVM with power of 5:  0.62
```

optimal separation hyperplane visualization

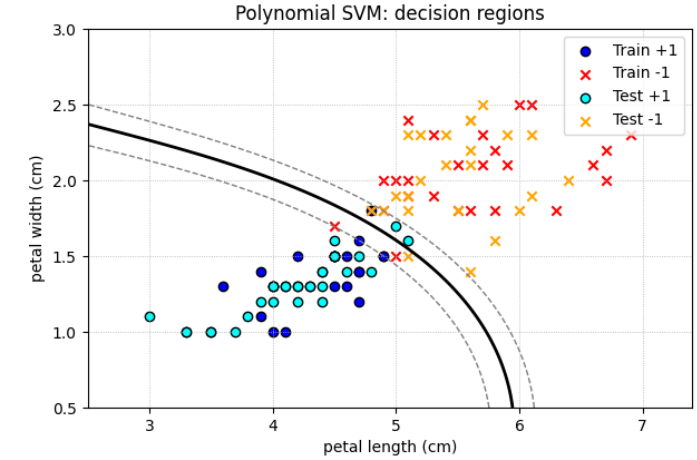
power=1



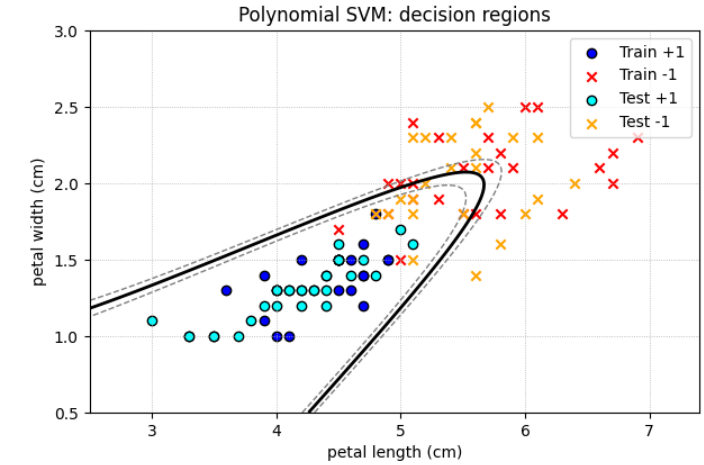
power=2



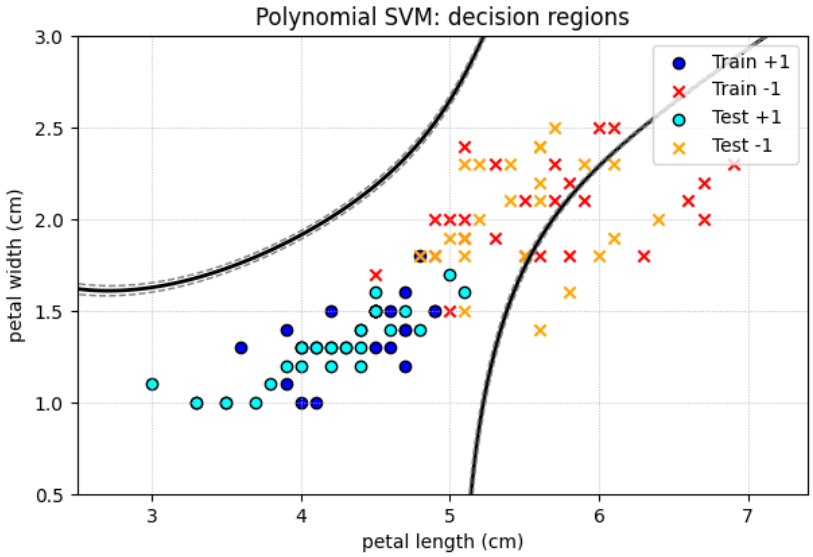
power=3



power=4



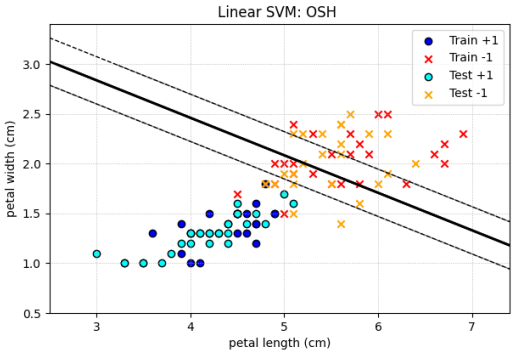
power=5



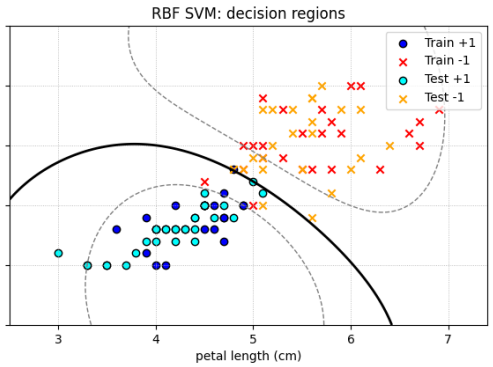
Part 4 Discussion and result presenting

1. The difference of hyperplane between linear SVM and nolinear SVM (kernel-based) is visually significant in shape, where linear SVM is a straight line and others are curve (except polynomial when power = 1, this is identical to linear SVM)

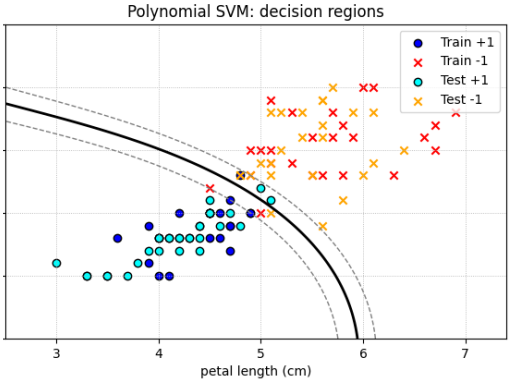
Linear SVM



RBF SVM

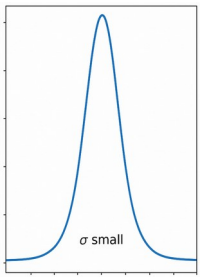


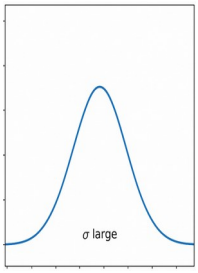
Polynomial SVM



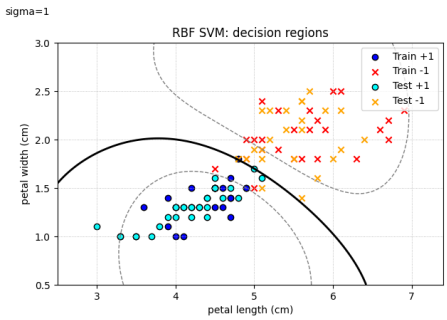
2. As kernel parameter varying, the hyperplane of both kernel-based SVM is changing as well.

because RBF kernel is actually followed by Gaussian distribution, hence when when the sigma value is increasing, the distribution curve gets wider (as shown below) means the coverage of hyperplane is getting larger as well.

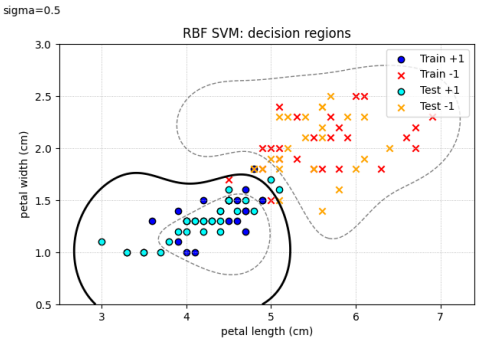




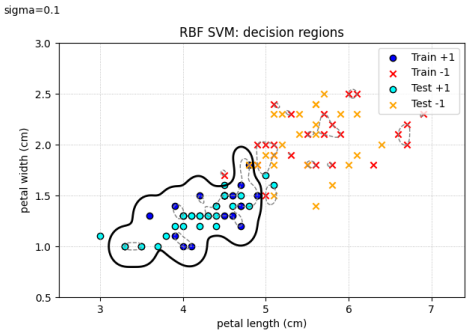
RBF SVM sigma = 1



RBF SVM sigma = 0.5

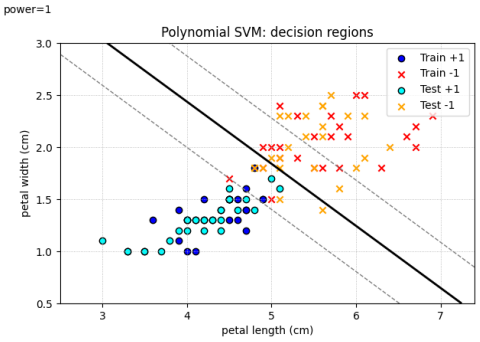


RBF SVM sigma = 0.1

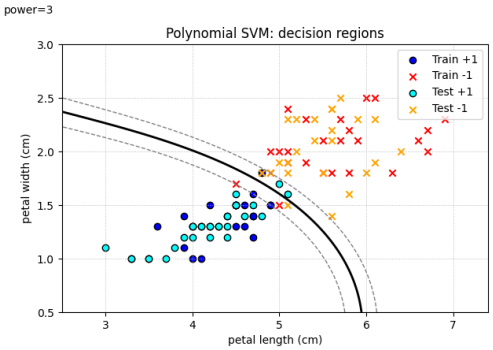


As for the polynomial SVM, when the power of kernel mapped function increase, the complexity of hyperplane is increase as follow.

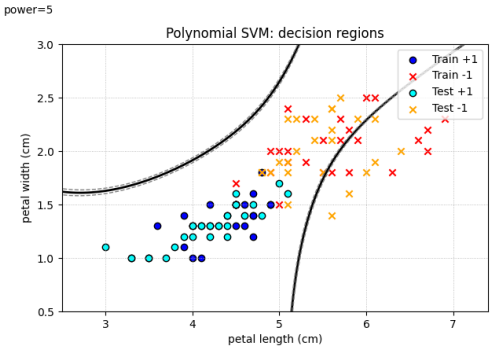
poly SVM power = 1



poly SVM power = 3



poly SVM power = 5



3. To avoid model overfitting, there are two common ways. One is select the appropriate feature combination as possible (the scatter plot of this homework's feature combination is shown below). The other method is to choose a well-perform model and its parameters, in this homework, SVM is chosen, hence the parameter tuning is crucial due to data is not linearly separable.

To adjust kernel parameter and avoid overfitting, we can perform following method.
1. Conduct cross validation (see how well a model can generalize to different data)
2. Feature normalization (tuning value of feature to specific range, e.g. 0~1)

data under feature combination of petal length and petal width is not linear separable, hence kernl-based method is more likely to perform well at first glance. Eventually, the result confirm the hypothesis.