DD2431 Machine Learning - Lab 2: SVM

Jun Yamada and Philipson Samuel

6 Running and Reporting

1

Moving the label 1 cluster to the same side, make it easier for classifier to find a descend boundary, since both clusters are completely separated. Moreover, we have to take variance into consideration, otherwise it would not be separated completely.

```
classA = [(random.normalvariate(1.5, 0.5),
           random.normalvariate(0.5, 0.5),
           1)
           for i in range(5)] + \
         [(random.normalvariate(1.5, 0.5),
           random.normalvariate(0.5, 0,5),
           1)
           for i in range(5)]

classB = [(random.normalvariate(0.0, 0.5),
            random.normalvariate(-0.5, 0.5),
            -1)
            for i in range(10)]
```

The optimal solution can be found in above case, since we change variable of classA and make the data separate completely.

On the other hand, if we set the different sign of mean in the normal distribution in class A, the data of classB would be located between the data of ClassA. Therefore, This make the classification harder.

Polynomial function

P=2, P=3

```
    classA = [(random.normalvariate(-1.5, 1.5),
               random.normalvariate(0.5, 1.5),
               1)
               for i in range(intA)] + \
             [(random.normalvariate(1.5, 1.5),
               random.normalvariate(0.5, 1.5),
               1)
               for i in range(intA)]
```

```
    classB = [(random.normalvariate(0, 1),
                random.normalvariate(-0.5, 1),
                -1)
                for i in range(intB)]
    data = classA + classB
    random.shuffle(data)
    return classA, classB, data
```

2

put images plotted hyperplane with non-linear kernel function.

3 (rewrite)

If we set the parameter bigger, the smaller margin of hyperplane is chosen. Maximizing the margin of the hyperplane means to maximize the variance. The bigger parameter makes the variance lower and it cause making the margin smaller.

7 Slack Implementation

1

As for polynomial kernel, if p=2,

2

3

4

the role of the parameter is weight for the error. if C is very small value, it is more strict for the error.

5

If it is solvable in our current dimensional space, we would rather use slack. Applying the complexity model makes variance larger.