Multi-class Classification

Friday 24 May 2024 12:17 AM

■ Multiclass classification : One-vs-One (OvO), One-vs-Rest (OvR) = One-vs-All (OvA)

(y=[+1,-1]) y=[0,1,2...]

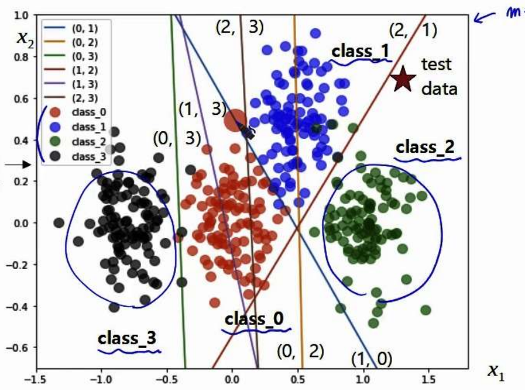
- SVM is a mathematical model for binary classification. Multiclass classification requires performing binary classification multiple times.
- There are two methods for multiclass classification: One-vs-One (OvO), One-vs-Rest (OvR), or One-vs-All (OvA).

▪ One-vs-One (OvO)

- Select two classes to find the boundary.
- A total of ${}_nC_2$ boundaries are created. m = n * (n-1) / 2, (n: the number of classes, m: the number of boundaries)
- Create m boundaries using the training data and use these boundaries to classify the test data. If there are many classes, it takes a long time because many boundaries need to be created, but the results are stable.

$m = {}_4C_2 = 6$

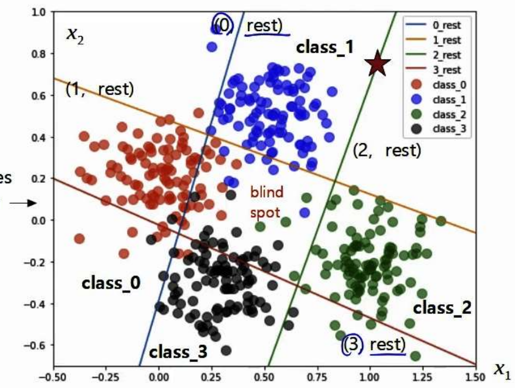
Six boundaries divide the four classes.



▪ One-vs-Rest (OvR) or One-vs-All (OvA)

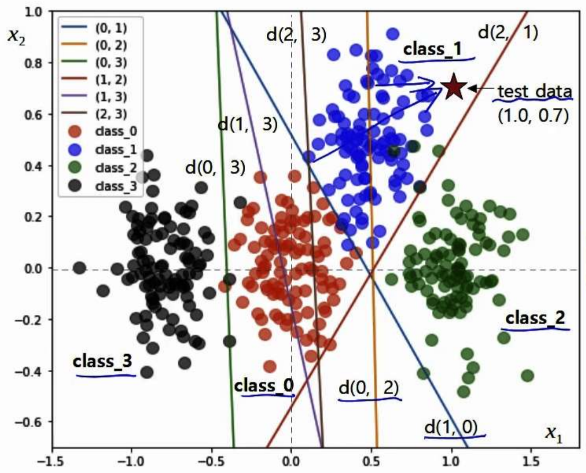
- Find the boundary between one class and the rest.
- A total of n boundaries are created. m = n, (n: the number of classes, m: the number of boundaries)
- Create m boundaries using the training data and use these boundaries to classify the test data. Compared to OvO, it has fewer boundaries and faster learning speed, but may have blind spots and is less stable.

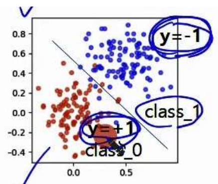
Four boundaries divide the four classes.



■ One-vs-One (OvO)

- Classify the classes of the test data using six boundaries. (majority voting)





The decision functions below are obtained by defining y as shown in this figure.

▪ Decision boundary	▪ Decision function	▪ Test data: (1.0, 0.7)
$d(0,1)$: $-3.6x_1 - 3.0x_2 + 1.74 = 0$	$\hat{y} = -3.6x_1 - 3.0x_2 + 1.74$	$\hat{y} = -3.96$ <0
$d(0,2)$: $-3.72x_1 - 0.41x_2 + 1.83 = 0$	$\hat{y} = -3.72x_1 - 0.41x_2 + 1.83$	$\hat{y} = -2.18$
$d(0,3)$: $4.62x_1 + 0.73x_2 + 1.8 = 0$	$\hat{y} = 4.62x_1 + 0.73x_2 + 1.8$	$\hat{y} = 6.93$
$d(1,2)$: $-3.63x_1 + 3.56x_2 + 1.82 = 0$	$\hat{y} = -3.63x_1 + 3.56x_2 + 1.82$	$\hat{y} = 0.68$
$d(1,3)$: $2.46x_1 + 0.89x_2 + 0.19 = 0$	$\hat{y} = 2.46x_1 + 0.89x_2 + 0.19$	$\hat{y} = 3.27$
$d(2,3)$: $1.84x_1 + 0.21x_2 - 0.2 = 0$	$\hat{y} = 1.84x_1 + 0.21x_2 - 0.2$	$\hat{y} = 1.79$

For d(0,1), the test data is to the right of the boundary, so it is classified as class_1. If y_hat of d(0,1) > 0, it is classified as class_0, otherwise it is classified as class_1.

*Decision rule: if the y-hat of d(A, B) > 0, then class=A, else class = B

```
             w                b
d(0, 1) : [-3.6 , -3.03]   [1.74]
d(0, 2) : [-3.72, -0.41]   [1.83]
d(0, 3) : [ 4.62,  0.73]   [1.8 ]
d(1, 2) : [-3.63,  3.56]   [1.82]
d(1, 3) : [ 2.46,  0.89]   [0.19]
d(2, 3) : [ 1.84,  0.21]   [-0.2]
```

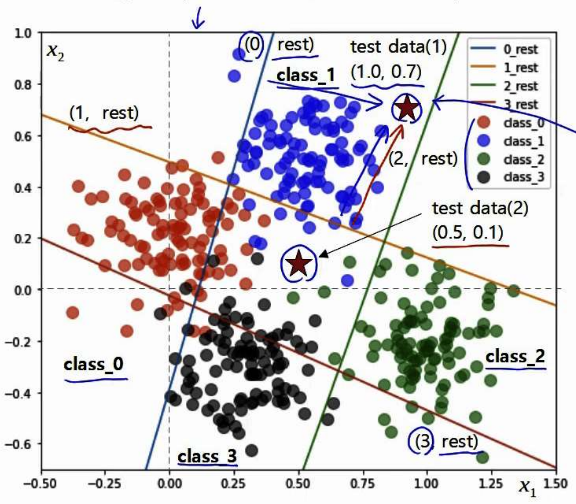
$\hat{y}$

- d(0,1) = -3.96 < 0 → class = 1
- d(0,2) = -2.18 < 0 → class = 2
- d(0, 3) = 6.93 > 0 → class = 0
- d(1, 2) = 0.68 > 0 → class = 1
- d(1, 3) = 3.27 > 0 → class = 1
- d(2, 3) = 1.79 > 0 → class = 2

Among the classes found with six boundaries, "1" is the most common, so it is classified as class = 1 by majority vote.

■ One-vs-Rest (OvR)

- Among multiple boundaries, classification is performed using the boundary with the largest decision function value.



▪ Decision boundary	▪ Decision function	▪ Test data (1) : (1.0, 0.7)
$d(0, rest)$: $-5.87x_1 + 1.71x_2 + 0.67 = 0$	$\hat{y} = -5.87x_1 + 1.71x_2 + 0.67$	$\hat{y} = -4.00$
$d(1, rest)$: $2.13x_1 + 5.73x_2 - 2.83 = 0$	$\hat{y} = 2.13x_1 + 5.73x_2 - 2.83$	$\hat{y} = 3.31$
$d(2, rest)$: $4.88x_1 - 1.74x_2 - 3.76 = 0$	$\hat{y} = 4.88x_1 - 1.74x_2 - 3.76$	$\hat{y} = -0.10$
$d(3, rest)$: $-2.59x_1 - 5.82x_2 - 0.14 = 0$	$\hat{y} = -2.59x_1 - 5.82x_2 - 0.14$	$\hat{y} = -6.80$

*OvO Decision rule: if the y-hat of d(A, B) > 0, then class=A, else class = B

test data (1) : x = (1.0, 0.7)

- d(0, rest) = -4.00 → rest → "It is not class 0"
- d(1, rest) = 3.31 → 1 → "It is class 1"
- d(2, rest) = -0.10 → rest → "It is not class 2"
- d(3, rest) = -6.80 → rest → "It is not class 3"

This test data point can be classified as class 1. The d(1, rest) value is positive and the largest. It can have multiple positive values, and even so, the largest d value can be used to classify.

test data (2) : x = (0.5, 0.1)

- d(0, rest) = -2.09 → "It is not class 0"
- d(1, rest) = -1.19 → "It is not class 1"
- d(2, rest) = -1.49 → "It is not class 2"
- d(3, rest) = -2.02 → "It is not class 3"

The test data is in a blind spot. The d(1, rest) value is the largest. This means that it is not-class-1, but because the test data is closest to this boundary, it means that it is least likely to be not-class-1. Therefore, it is reasonable to classify it as class 1.

```
          w               b
0-rest: [-5.87,  1.71]   [ 0.67]
1-rest: [ 2.13,  5.73]   [-2.83]
2-rest: [ 4.88, -1.74]   [-3.76]
3-rest: [-2.59, -5.82]   [-0.14]
```

*OvR Decision rule: Classify the test data using the boundary with the highest decision function value.