

## 숙제 #5 (40점)

Classifier learning, Support vector machine

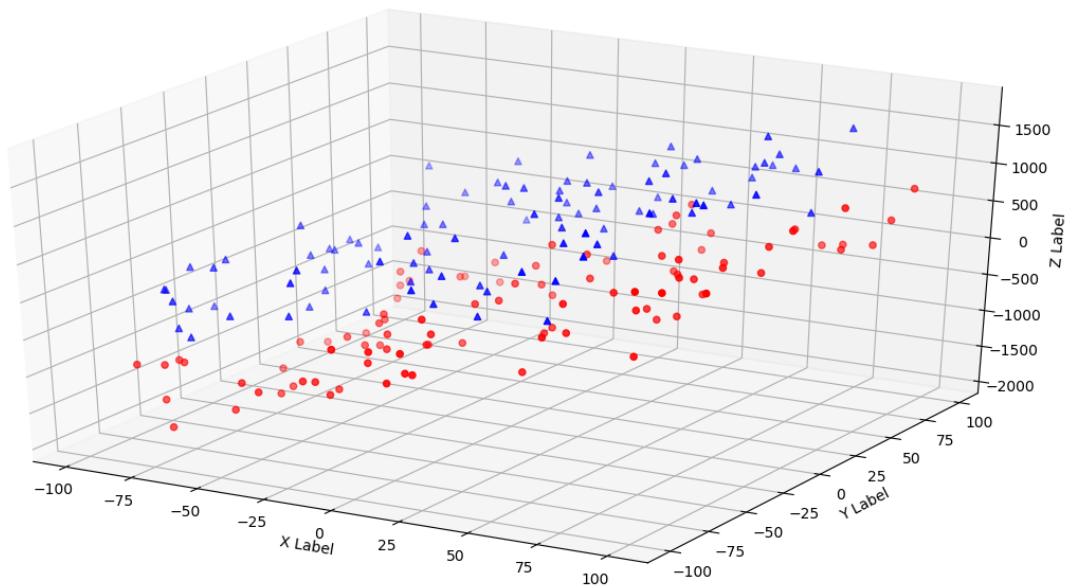
EEE6478 컴퓨터 비전

120180048 김태훈

실험을 위하여 3차원 공간에서 평행한 두 평면을 설정하고 각 평면의 위, 아래에 존재하는 데이터를 2500개씩 생성하여 SVM의 학습데이터로 사용하였다. 또한 학습된 SVM의 분류 성능 실험을 위하여 동일한 조건의 데이터를 추가로 1000개씩 생성하여 다양한 파라미터에 대한 분류 정확도를 분석하였다. 먼저 실험에 필요한 평면은  $z = 10x + y + m$ 에서  $m$ 을 변화시킨 두 평면을 사용하였다. 또한 SVM의 커널로는 Linear, RBF, Polynomial 커널을 사용하였다.

세 커널에 대한 실험에서 공통적으로는 soft margin 값을 변화시켜가며 실험을 진행하였다. 각 커널에 대한 특이사항으로는 RBF 커널의 경우 gamma 값을 변화시켰으며, Polynomial 커널의 경우 degree 값을 변화시키며 실험을 진행하였다. 이후 각 파라미터 변화에 따른 분류 정확도의 변화를 통해 성능 변화를 측정하였다. 실험 결과를 출력하는 과정에서 실험에 사용한 데이터 모두를 플롯에 나타내기엔 너무 뾰뚱하기 때문에 각 class 별로 100개씩을 뽑아 3차원 상에 표시하였다.

### a) 데이터 분포 예시

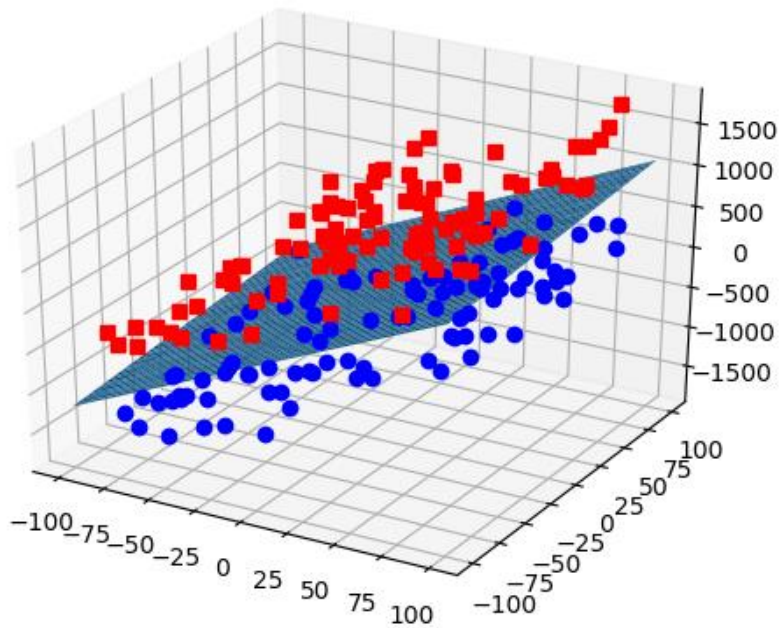


위 예시는  $m$ 을 각각 -30, -10으로 둔 두 평면을 경계로 하는 데이터를 생성하여 그 중 100개씩을 3차원 공간에 나타낸 예시이다. 실제 실험에서는 두  $m$  값의 차이를 변화시켜가며 해당 차이에 따라 분류 정확도가 얼마나 떨어지는 지를 비교해보았다.

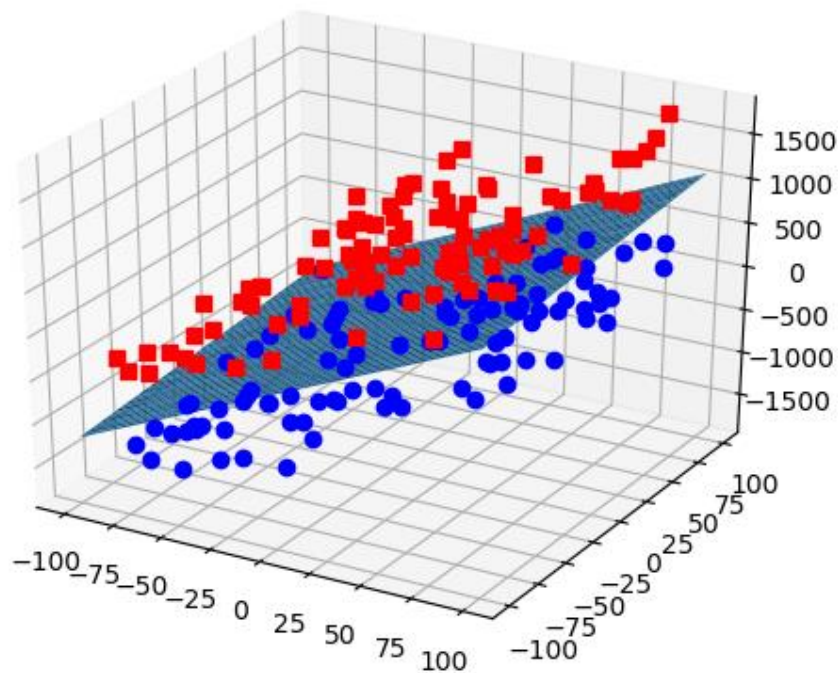
b) 커널 별 Decision Boundary 예시

a)와 동일한 조건에서 각 커널을 이용하여 SVM의 decision boundary를 구한 예시는 다음과 같다. 예시에서 soft margin은 1.0, RBF의 gamma는 0.7, Polynomial의 degree는 3으로 설정하였다.

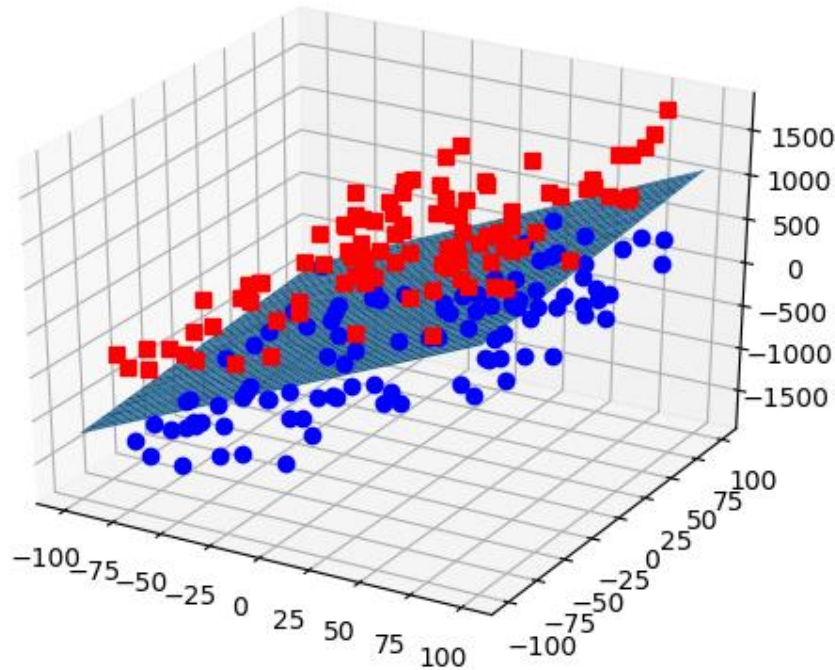
**[Linear]**



**[RBF]**



## [Polynomial]



### c) 분류 정확도 변화

각 실험을 위해서 class 별 2500개, 총 5000개의 데이터를 이용하여 SVM을 학습하고 class 별 1000개의 데이터를 추가로 생성하여 이를 얼마나 정확하게 분류하는지를 확인하였다. 이후 그 정도를 분류 정확도를 통해 비교하면서 파라미터에 따른 SVM의 성능 변화를 확인하였다.

#### 1) 데이터 경계의 m값 차이에 따른 변화

데이터 생성을 위하여 사용한 두 경계 사이의 m값 차이에 따른 분류 정확도 변화에 대한 실험을 진행하였다. 정확한 비교를 위하여 SVM의 커널은 Linear 커널로 고정하고, soft margin 값으로는 1.0을 사용하였다.

$m_1 - m_2$	0	10	20	30	40	50	60	70	80
정확도	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00

두 평면 사이의 m 값 차이가 적어질수록 분류 정확도가 현저하게 떨어질 것으로 예상하였으나 실제 실험에서는 큰 차이를 보이지 않았다. 이는 두 데이터 class의 경계가 Linear한 평면이기 때문에 SVM이 대부분의 경우에서 모두 좋은 성능을 보이기 때문으로 보인다. 이후 실험에서는 두 평면 사이의 m 값 차이를 30으로 고정하고 나머지 과정을 진행하였다.

## 2) Soft margin 변화에 따른 변화

RBF의 gamma는 0.7, polynomial의 degree는 3으로 고정한 채로 실험을 진행하였다.

soft margin	Linear	RBF	Polynomial
0.2	1.00	0.99	0.93
0.4	0.98	0.96	0.93
0.6	1.00	0.97	0.93
0.8	1.00	0.98	1.0
1.0	1.00	0.98	0.00
1.2	1.00	0.98	1.00
1.4	1.00	0.99	0.01
1.6	1.00	0.95	0.01
1.8	0.99	0.99	0.01
2.0	0.99	0.98	0.01

이 실험결과를 봤을 때 Linear 커널은 soft margin의 변화에 큰 영향을 받지 않고 모는 경우 높은 정확도를 기록하는 것을 볼 수 있다. 그러나 Polynomial 커널의 경우에는 soft margin값의 영향을 많이 받아 몇몇 soft margin 값에 대해서는 0에 가까운 점수를 기록하는 것을 확인하였다. RBF 커널의 경우에는 일부 soft margin 값에 대해 상대적으로 낮은 정확도를 기록하기는 했으나 전반적으로 정확도 점수를 기록하는 것을 볼 수 있다.

## 3) gamma 변화에 따른 변화

다음으로는 RBF 커널의 gamma 값 변화에 따른 분류 정확도를 비교하였다. 실험을 위한 soft margin 값은 1.0으로 고정하였다. 몇몇 gamma 값의 경우 상대적으로 낮은 정확도를 보인 경우는 있었으나 Polynomial 커널과 같이 0에 가까운 정확도를 보인 결과는 없었다.

gamma	degree
0.2	0.96
0.4	0.99
0.6	0.96
0.8	0.99
1.0	0.99
1.2	0.98
1.4	0.98
1.6	0.98
1.8	0.98
2.0	0.98

#### 4) degree 변화에 따른 변화

degree	soft margin=0.8	soft margin=1.0
1	0.97	0.98
3	0.98	0.94
5	0.96	0.00
7	0.99	0.98
9	0.97	0.01
11	0.98	0.96
13	계산불가	계산 불가

1)의 실험에서 0에 가까운 낮은 점수와 높은 점수를 기록한 두 soft margin에 대하여 degree를 변화시켜가며 실험을 진행하였다. 그 결과 0에 가까운 정확도를 기록한 soft margin=1.0에 대한 실험의 경우에도 특정 degree에 대해서는 높은 정확도를 보임을 확인할 수 있었다. 이 실험을 통해 Polynomial 커널은 soft margin과 degree 등의 파라미터에 매우 민감함을 알 수 있었다. 또한 degree 값이 너무 높을 경우에는 계산 시간이 너무 오래 걸려 실험 결과를 확인할 수 없었다.

실험의 결과를 전반적으로 종합해 본 결과는 다음과 같다. 전체적인 실험에 사용한 데이터의 class 간 경계를 평면으로 잡았기 때문에 Linear 커널이 가장 좋은 성능을 보인 것으로 판단되나, 데이터가 달라지면 RBF나 Polynomial 커널이 더 좋은 성능을 보일 수도 있을 것으로 판단된다. 또한 다른 두 커널에 비해 Polynomial 커널은 사용한 파라미터에 따른 성능 차이가 심함을 확인하였으며, Polynomial 커널을 사용할 때는 이에 주의해야 할 것으로 보인다.