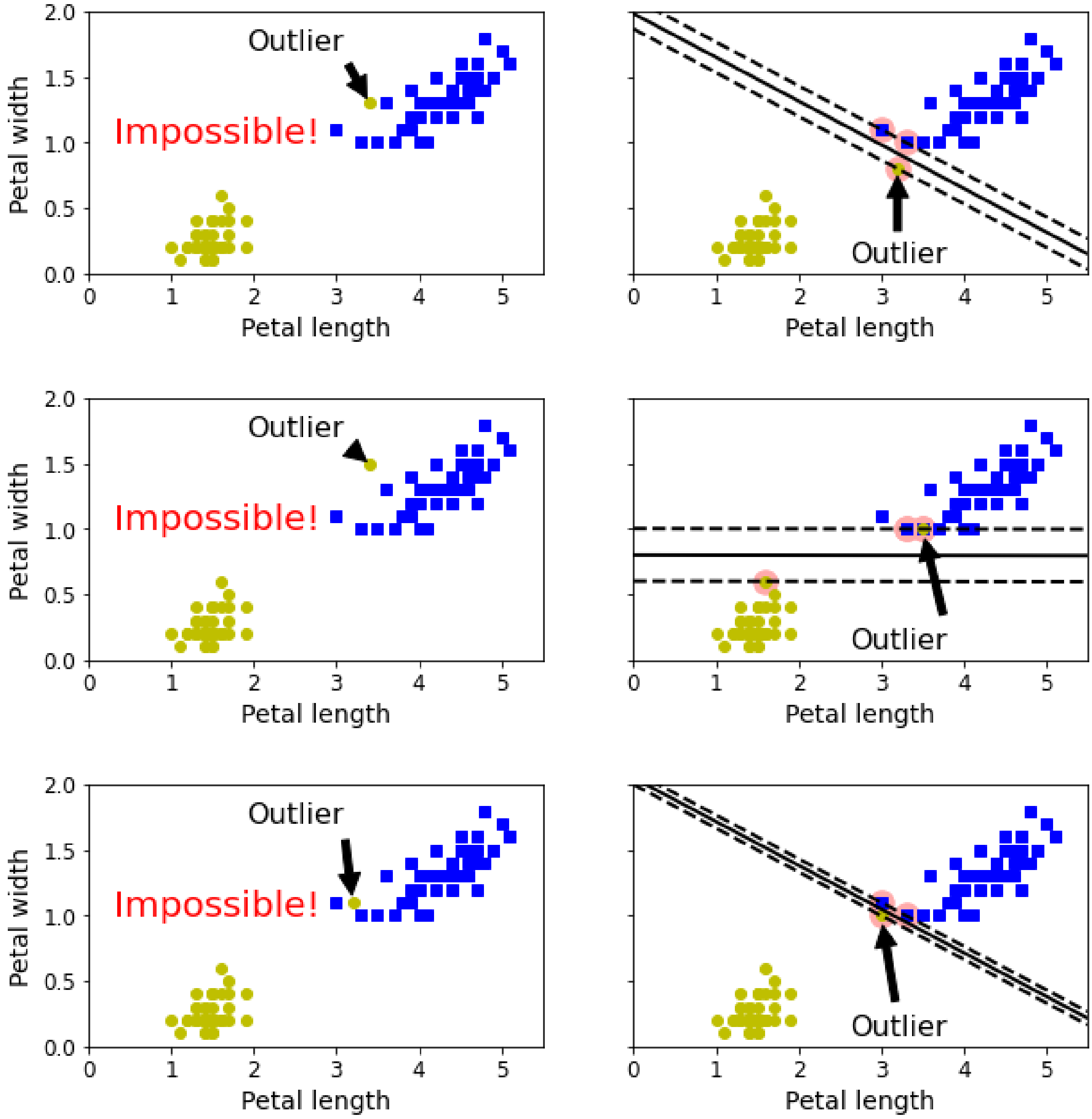
기계학습의 기초 및 전기정보 응용

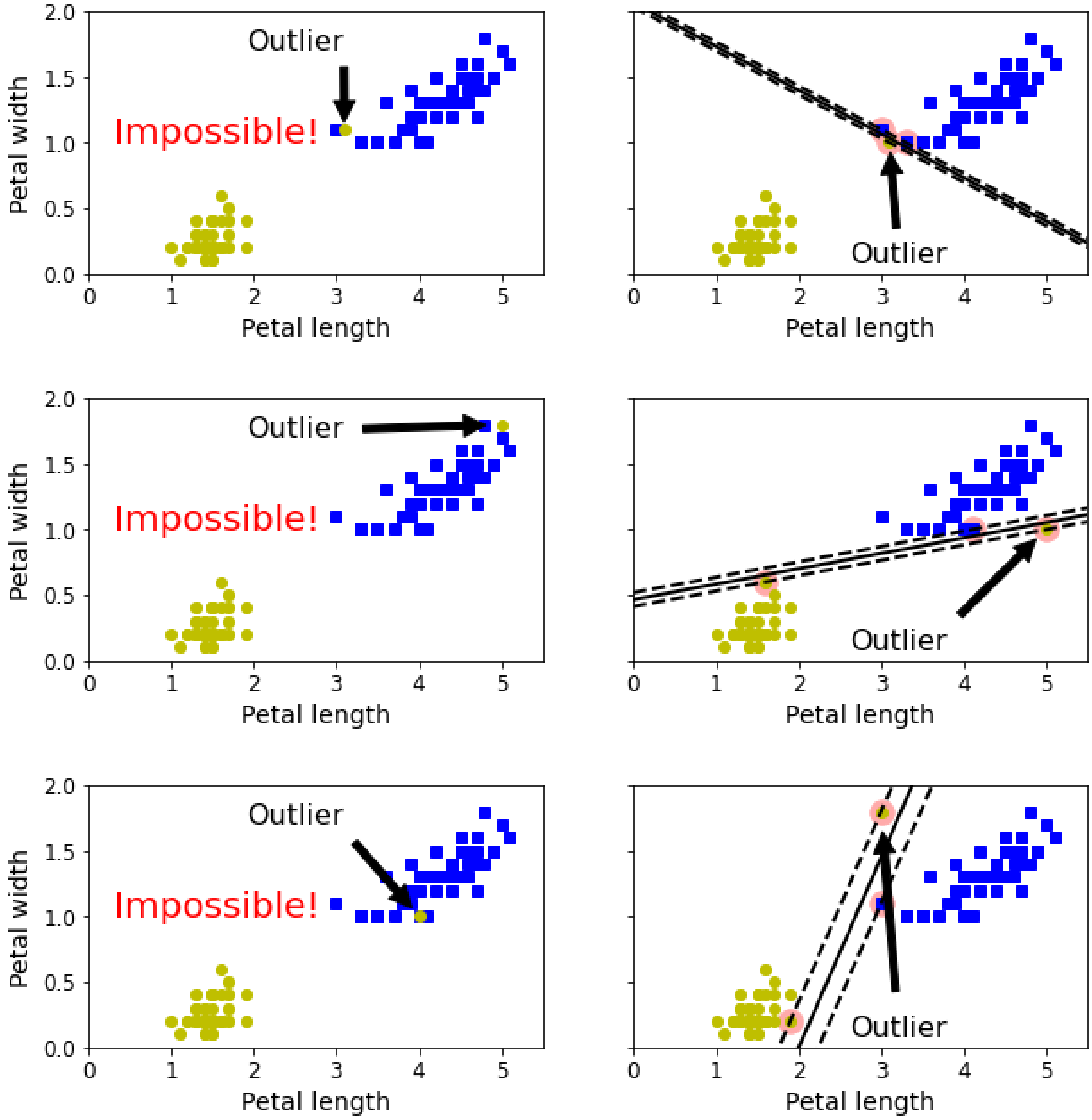
Assignment 2

전기정보공학부

2017-17088 박찬정

Problem 2



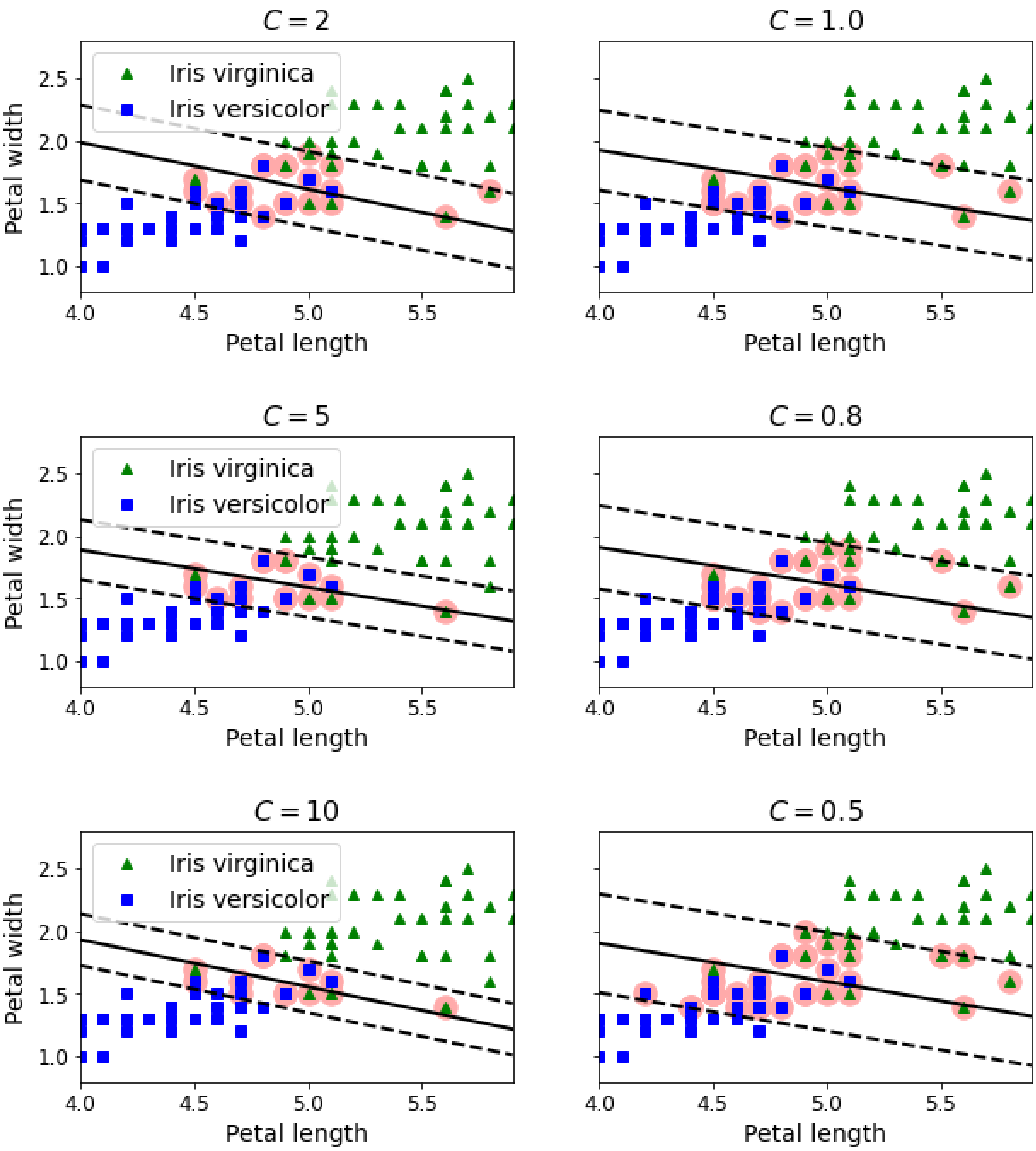


위에서부터 표현된 outlier의 좌표는 각각 아래와 같다.

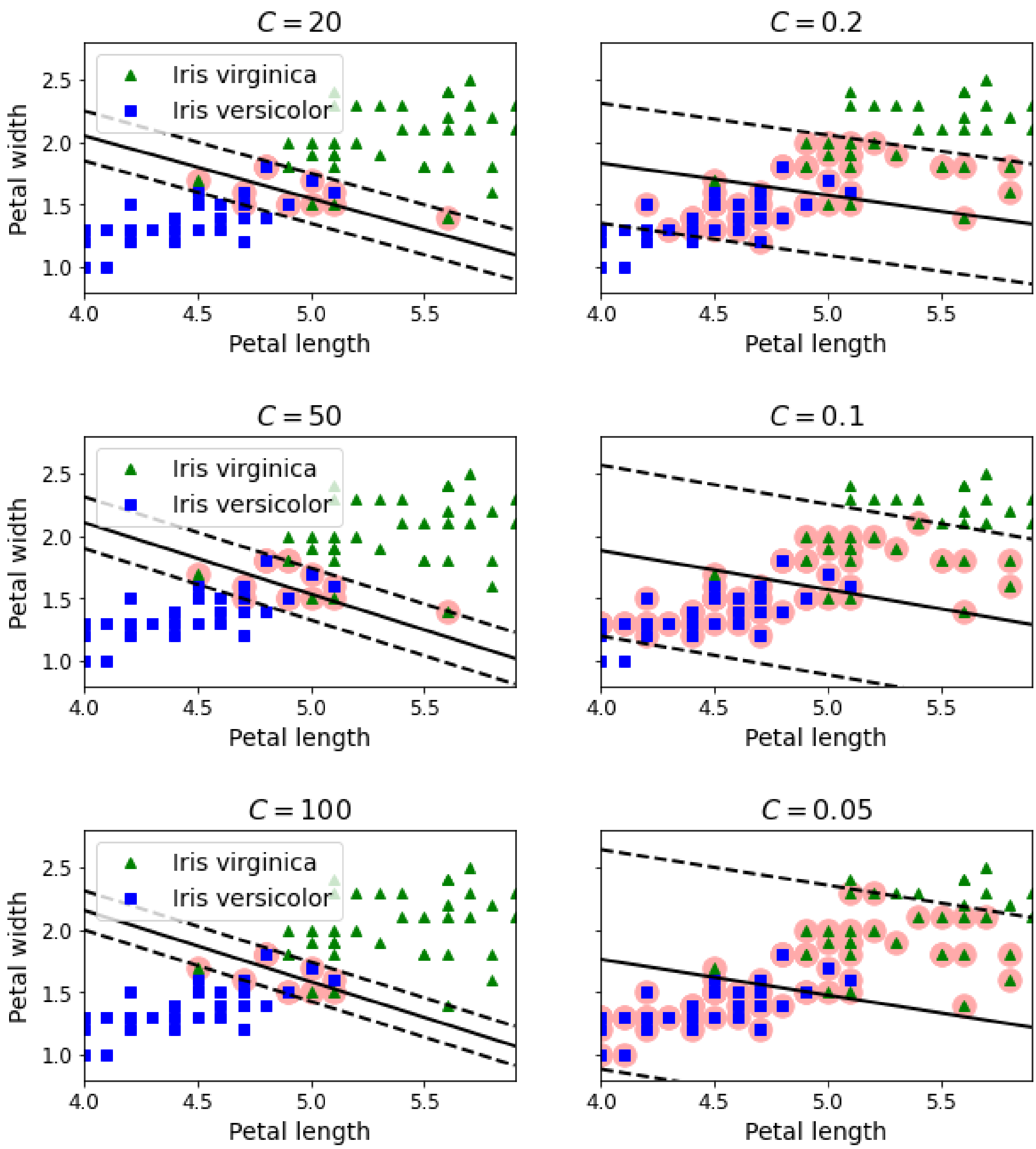
```
X_outliers = np.array([[3.4, 1.3], [3.2, 0.8]])
X_outliers = np.array([[3.4, 1.5], [3.5, 1.0]])
X_outliers = np.array([[3.2, 1.1], [3.0, 1.0]])
X_outliers = np.array([[3.1, 1.1], [3.1, 1.0]])
X_outliers = np.array([[5.0, 1.8], [5.0, 1.0]])
X_outliers = np.array([[4.0, 1.0], [3.0, 1.8]])
```

하나의 outlier 만으로도 경계선이 민감하게 변화하고, 불가능한 경우도 많이 생기는 것을 확인할 수 있었다.

Problem 3

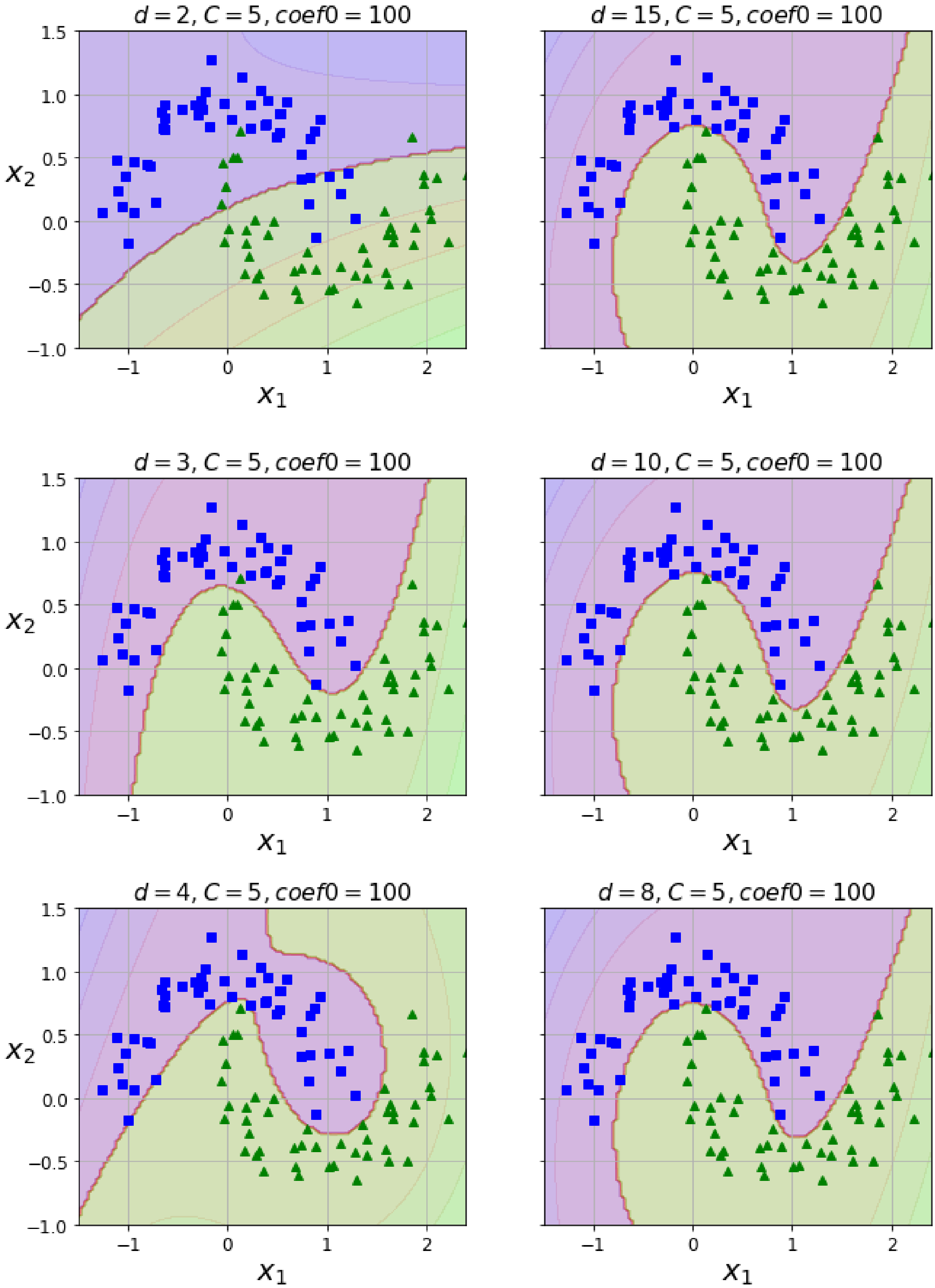


$C = 2$
$C = 1.0$
$C = 5$
$C = 0.8$
$C = 10$
$C = 0.5$
Iris virginica
Iris versicolor
Petal width
Petal length
2.5
2.0
1.5
1.0
4.0
4.5
5.0
5.5

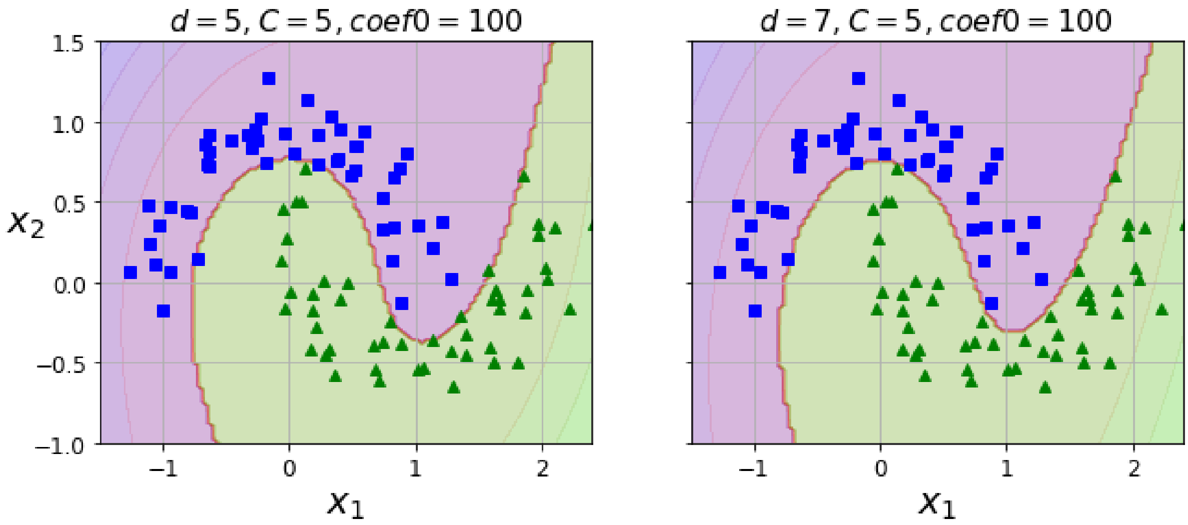


C 값이 작아질수록 margin이 넓어지고, C 값이 클수록 margin이 좁아진다. 그에 따라서 경계선의 형태도 조금씩 달라진다.

Problem 4-1

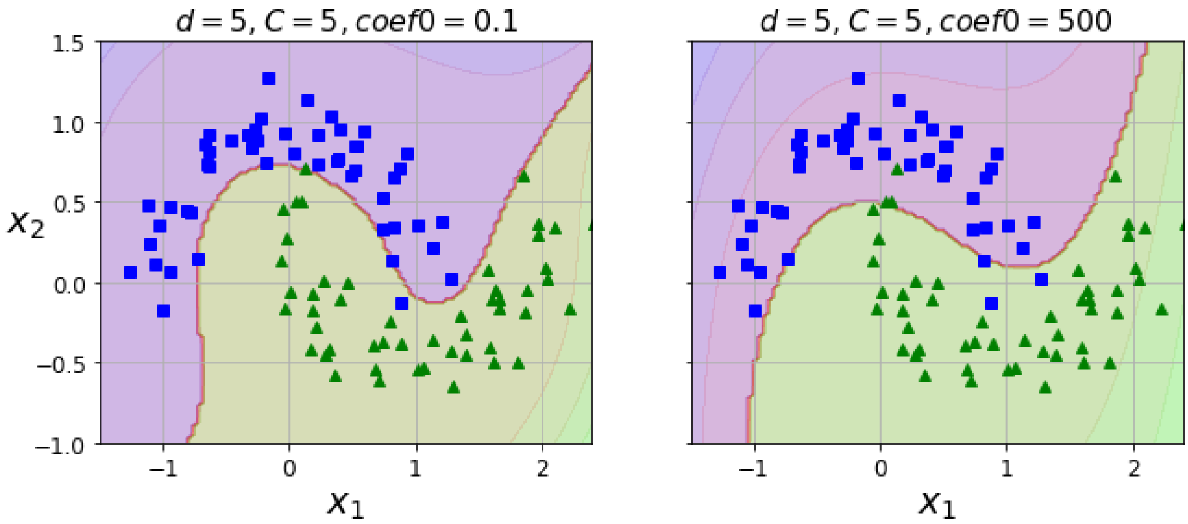


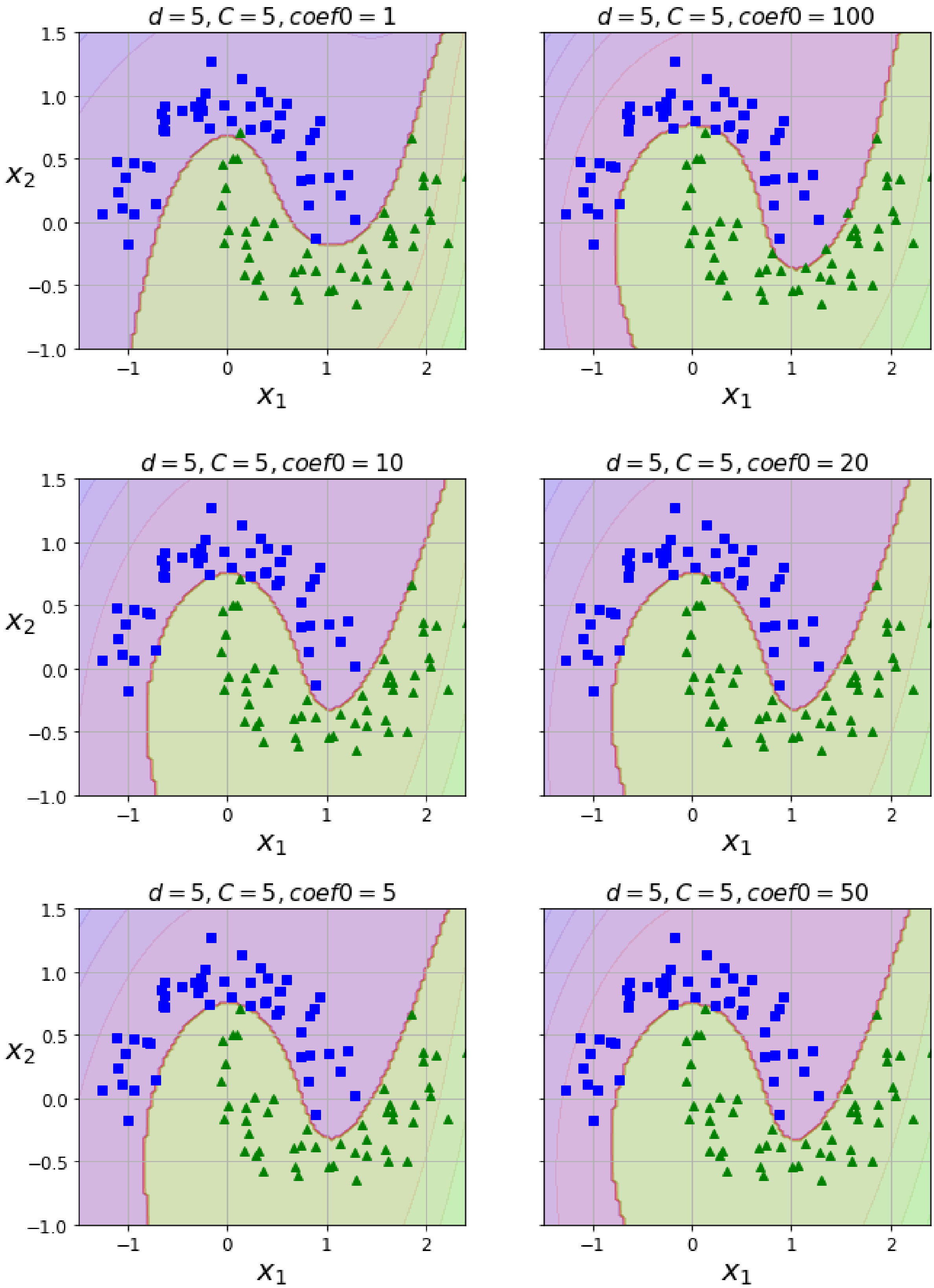
$d = 2, C = 5, coef0 = 100$
$d = 15, C = 5, coef0 = 100$
$d = 3, C = 5, coef0 = 100$
$d = 10, C = 5, coef0 = 100$
$d = 4, C = 5, coef0 = 100$
$d = 8, C = 5, coef0 = 100$
x_1
x_2
1.5
1.0
0.5
0.0
−0.5
−1.0
−1
0
1
2



coef0를 100으로 고정하고 degree를 변화시켜보았다.

Degree 4 까지는 경계선의 모양이 크게 바뀌는 모습을 보여주었으나, 5부터는 전체적인 모습이 크게 변하지 않았다.



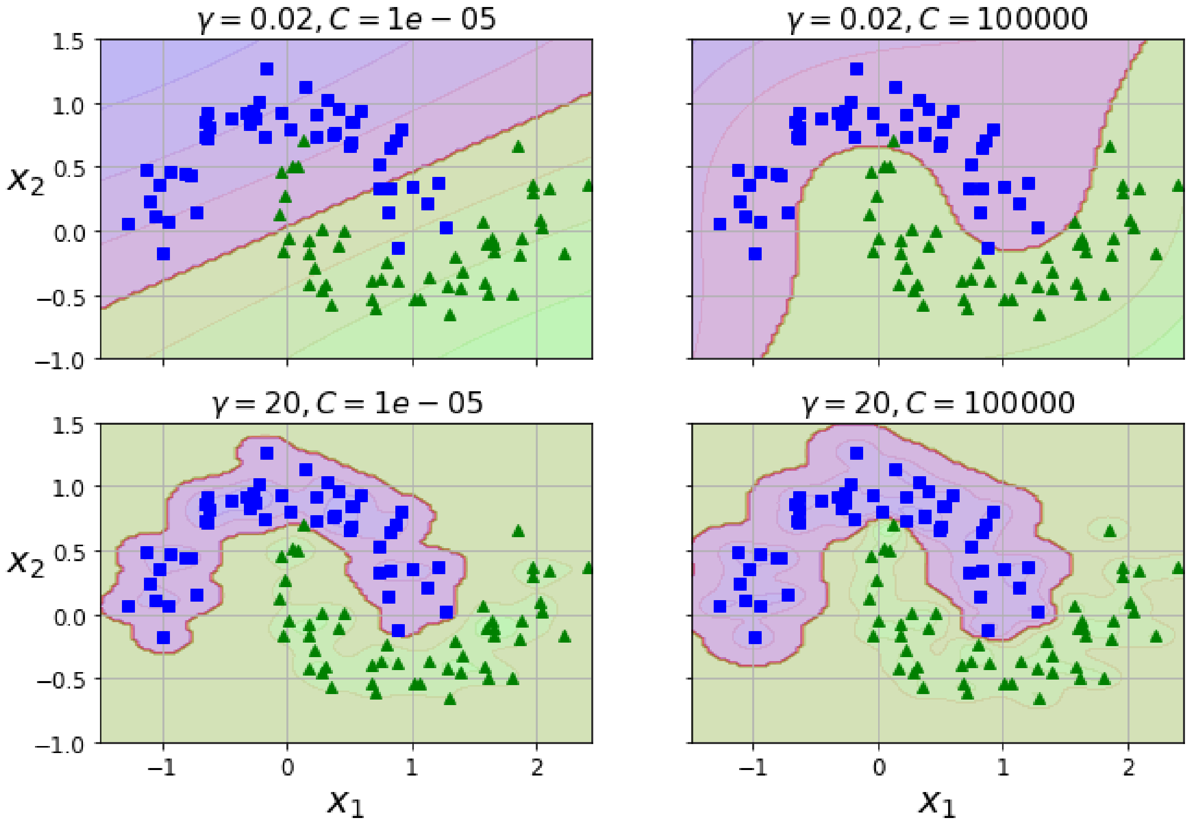


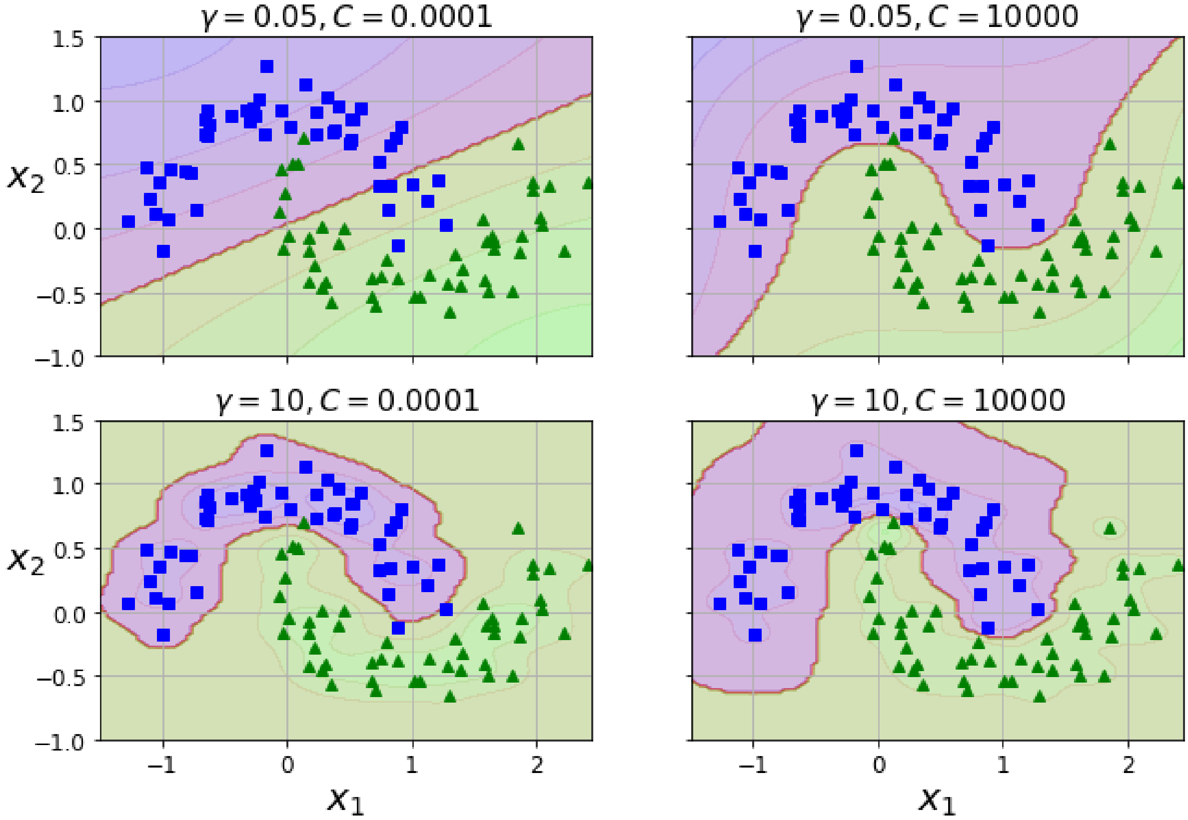
다음으로 degree를 5로 고정하고 coef0를 변화시켜 보았다.

coef0가 낮을수록 outlier에 둔감한 모습을 보여준다. coef0가 10 이상일 때 부터는 misclassify된

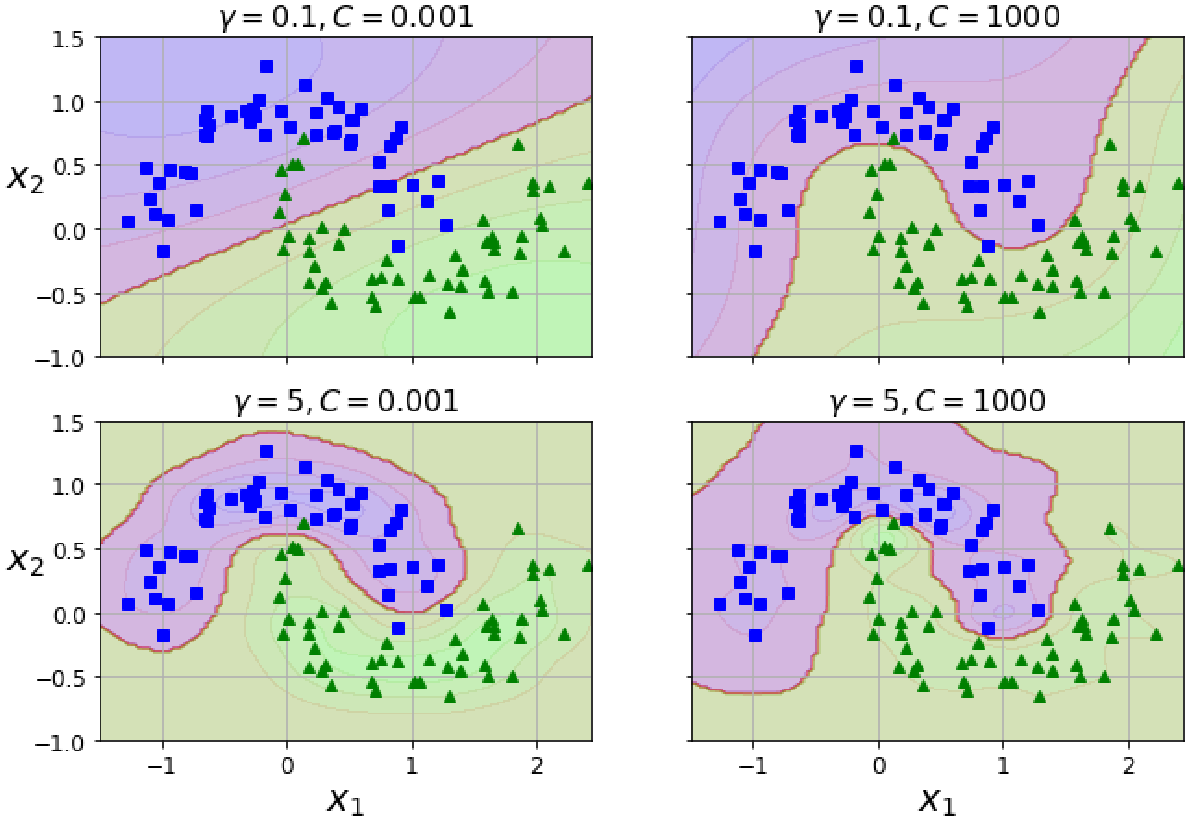
point는 없었고, 전체적인 형태도 변화하지 않았다. coef0가 커지면서 경계선이 초록색 point들에 더 가까워지는 모습을 보였다.

Problem 4-2

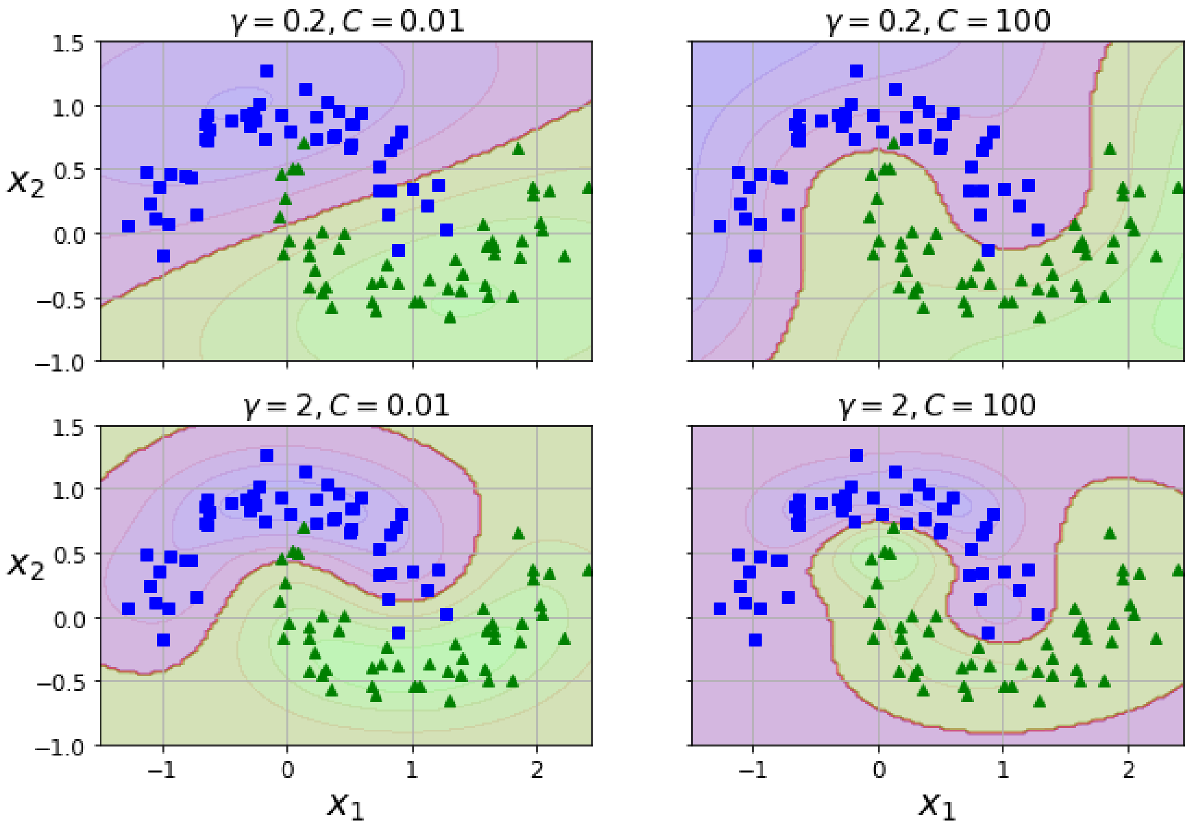


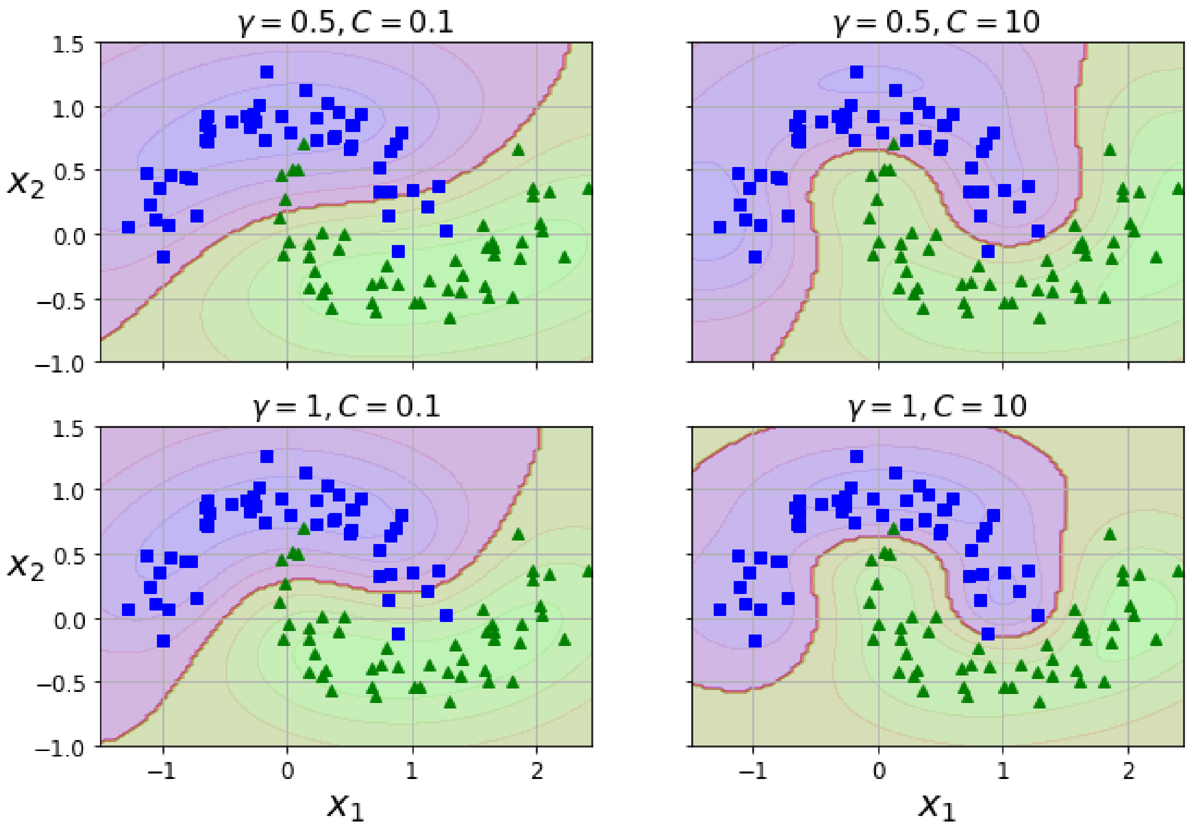


$\gamma = 0.05, C = 0.0001$
$\gamma = 0.05, C = 10000$
$\gamma = 10, C = 0.0001$
$\gamma = 10, C = 10000$
x_2
x_1



$\gamma = 0.1, C = 0.001$
$\gamma = 0.1, C = 1000$
$\gamma = 5, C = 0.001$
$\gamma = 5, C = 1000$
x_2
x_1





감마 값이 클수록 경계선이 point들에 가깝이 붙는 모습을 보였고, 감마 값이 작을수록 point의

구체적인 위치와 무관한 형태를 보였다. 감마 값이 0.1 이하인 경우 거의 선형으로 보일 정도이다.

C값이 작을수록 outlier에 둔감하고 misclassify되는 point의 수가 늘어난다. 반대로 C값이 클수록 outlier에 민감하다.

적절한 감마, C 값을 선택하지 않으면 overfitting이 쉽게 일어나는 것을 확인할 수 있었다.