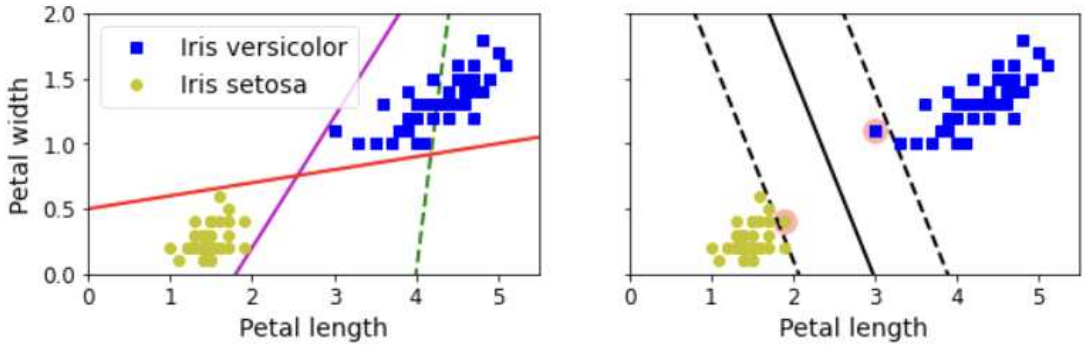
기계학습의 기초 및 전기정보 응용 Assignment02 보고서

전기정보공학부 2017-13758 강정민

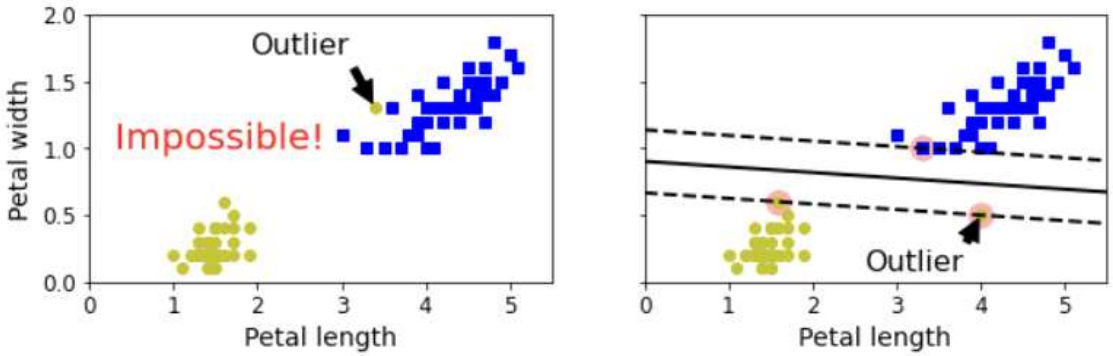
Problem 1

```
# SVM Classifier model
#################### TODO ####################
svm_clf = SVC(kernel='linear')
svm_clf.fit(X, y)
##############################################
```

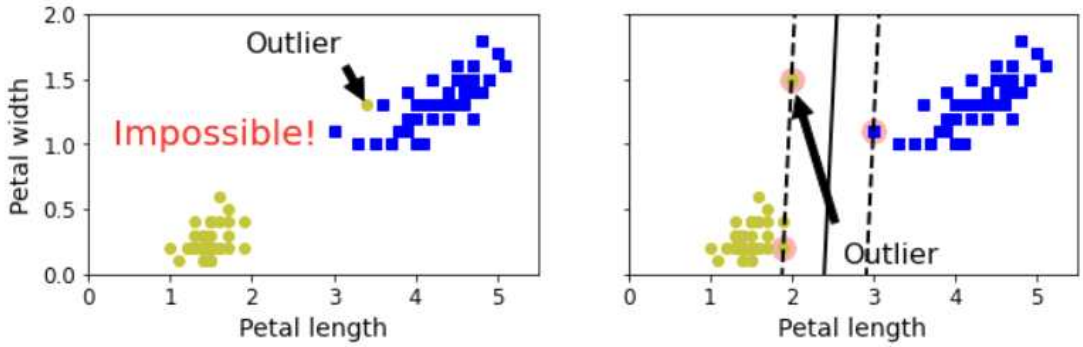


Problem 2

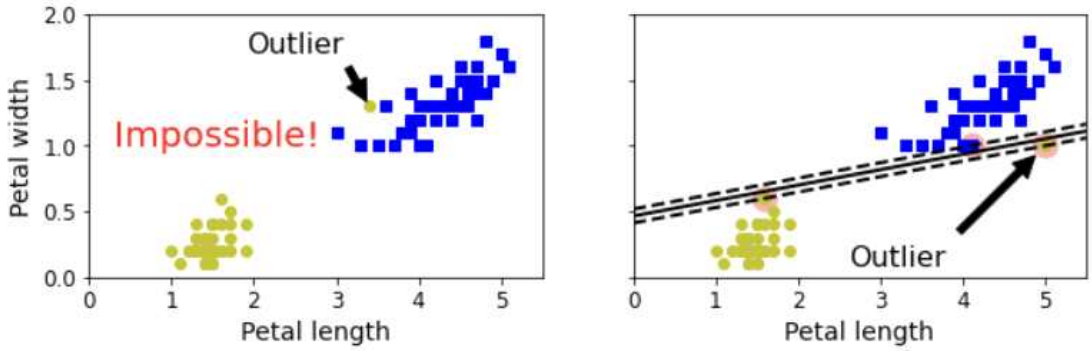
1) X_outliers = np.array([[3.4, 1.3], [4.0, 0.5]])



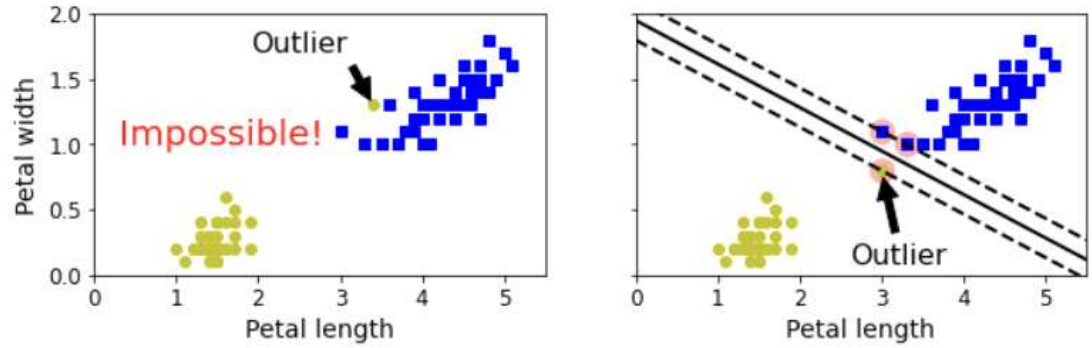
2) X_outliers = np.array([[3.4, 1.3], [2.0, 1.5]])



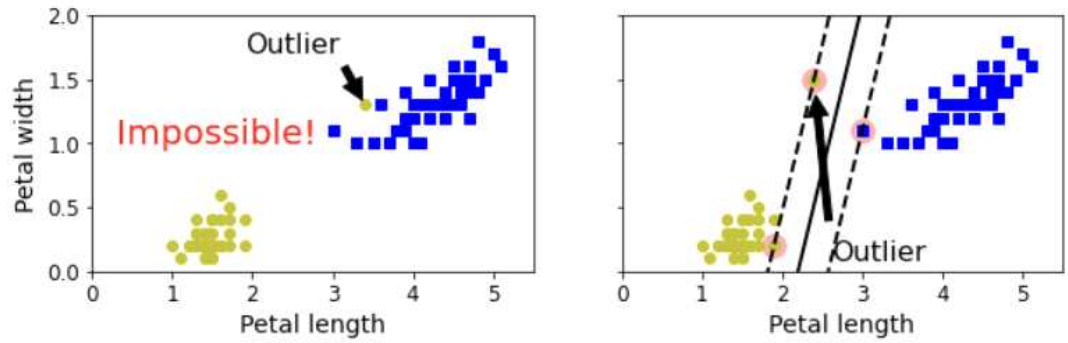
3) X_outliers = np.array([[3.4, 1.3], [5.0, 1.0]])



4) X_outliers = np.array([[3.4, 1.3], [3.0, 0.8]])

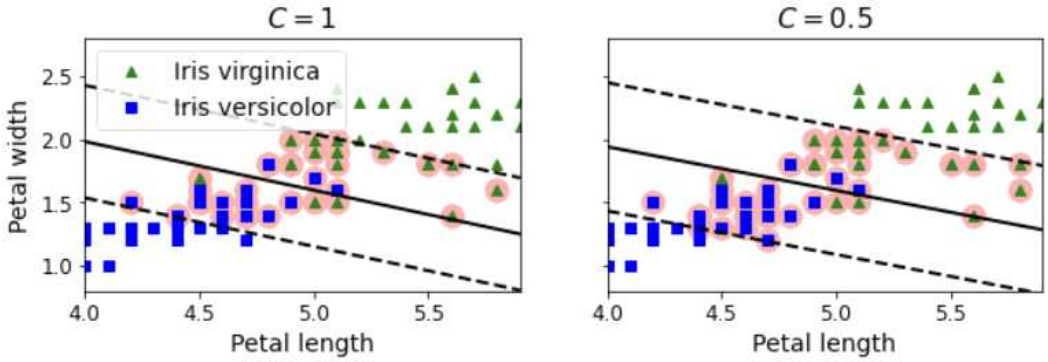


5) X_outliers = np.array([[3.4, 1.3], [2.4, 1.5]])

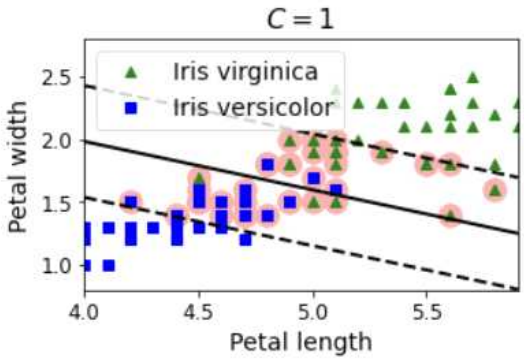


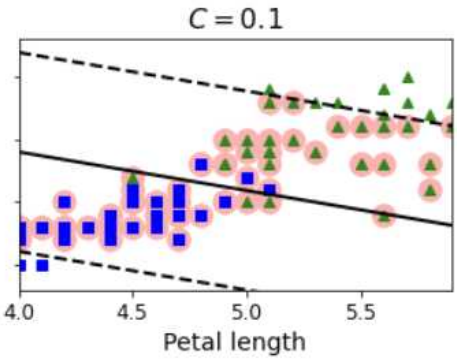
Problem 3

1) C=1, C=0.5

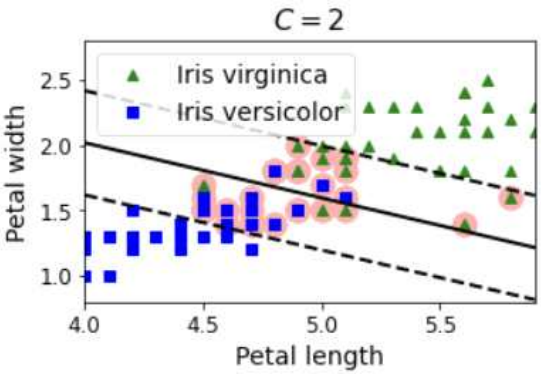


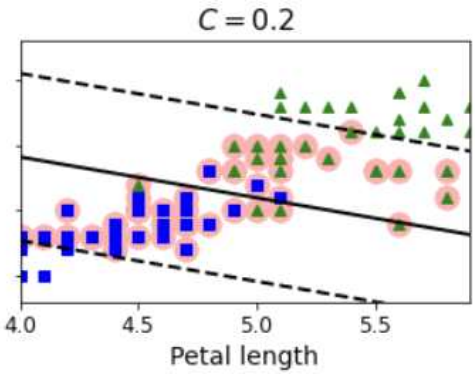
2) C=1, C=0.1



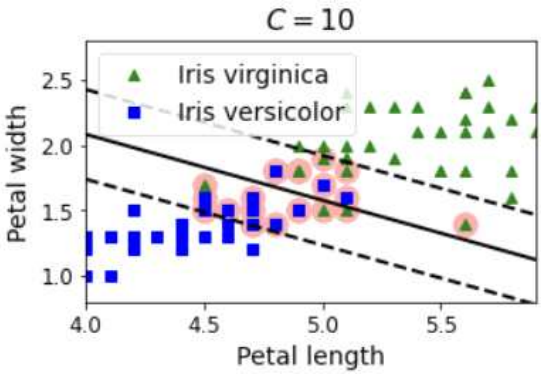


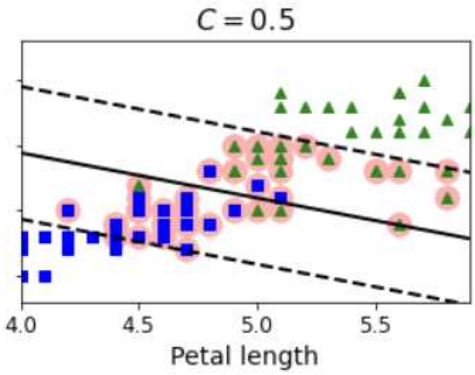
3) C=2, C=0.2



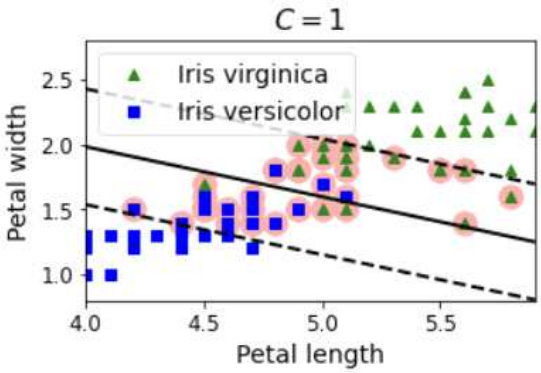


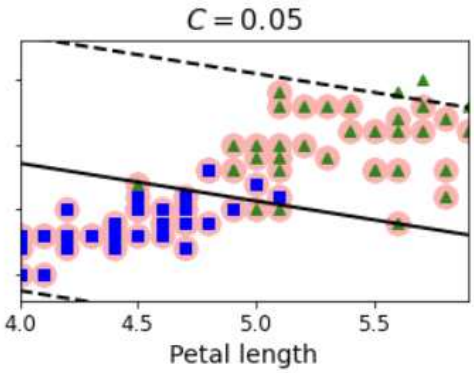
4) C=10, C=0.5





5) C=1, C=0.05

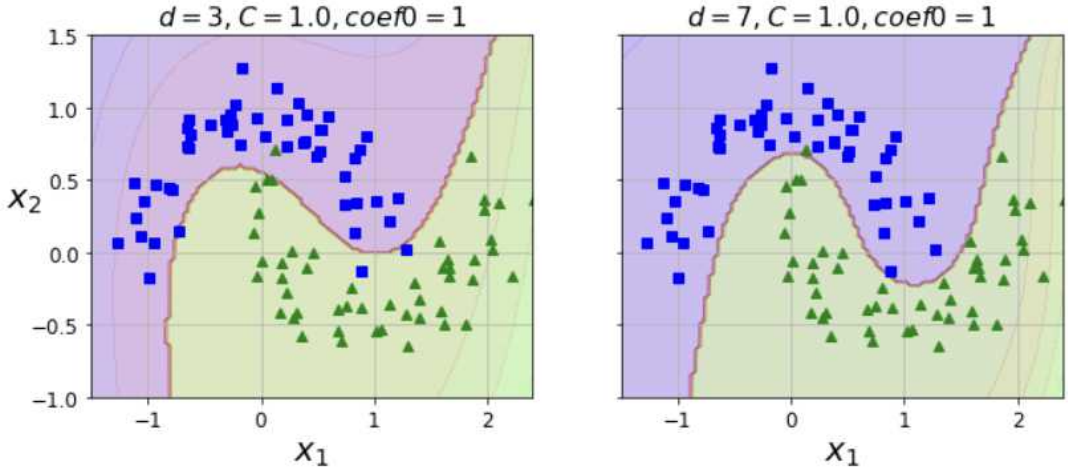




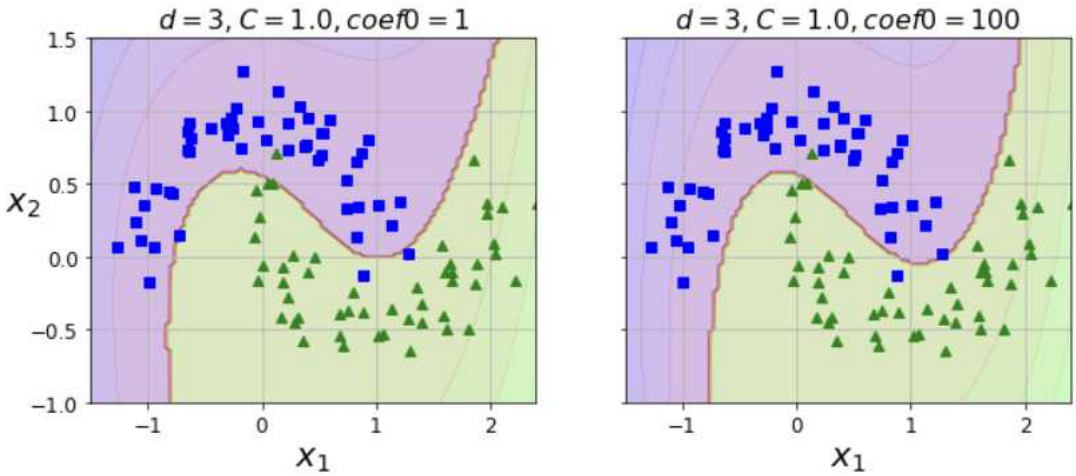
-> SVM 알고리즘은 maximize margin과 최대한 많은 점을 제대로 분류하는 것을 목표로 한다. 이때 C가 작을수록 전자의 목표에 집중하고, C가 커질수록 후자의 목표에 집중하게 됨을 확인할 수 있다.

Problem 4-1

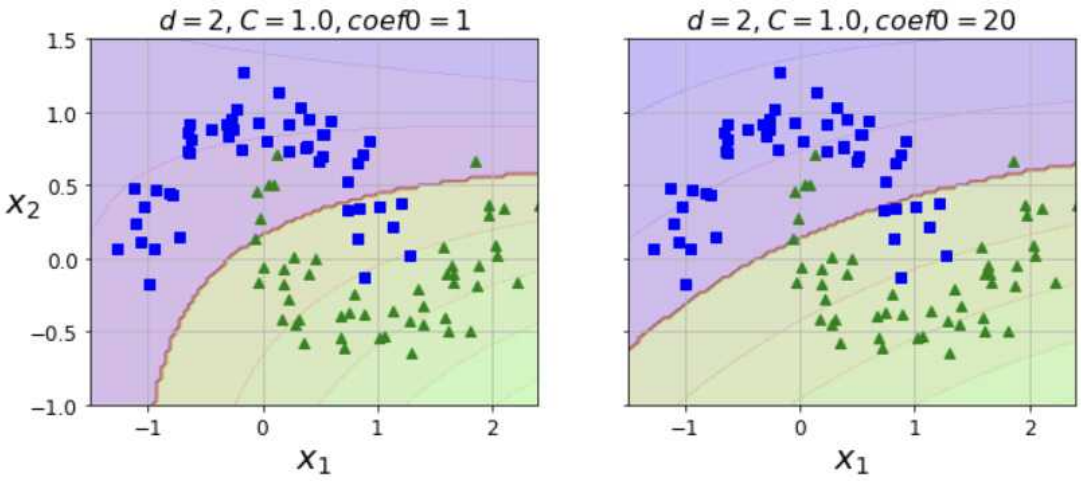
1) (d=3, coef0=1), (d=7, coef0=1)



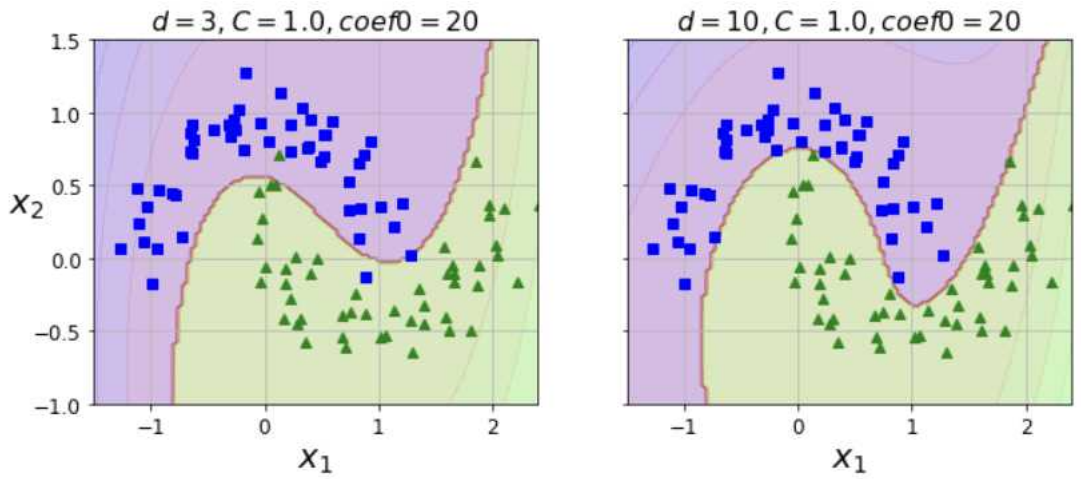
2) (d=3, coef0=1), (d=3, coef0=100)



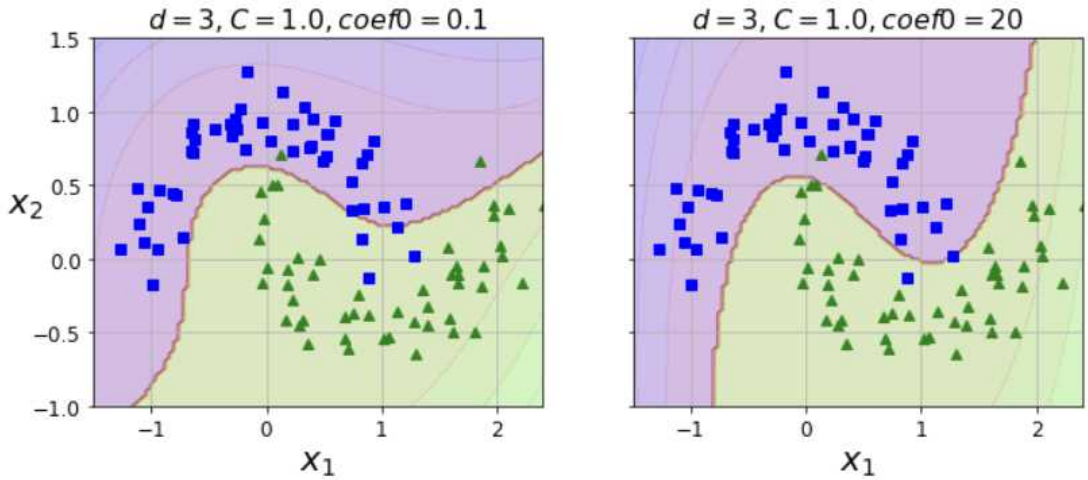
3) (d=2, coef0=1), (d=2, coef0=20)



4) (d=3, coef0=20), (d=10, coef0=20)



5) (d=3, coef0=0.1), (d=3, coef0=20)

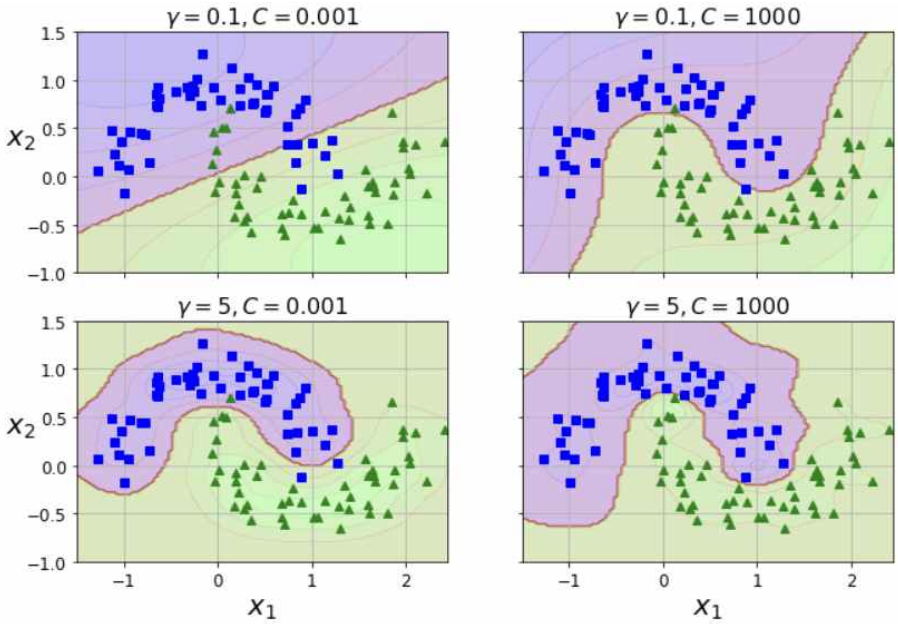


-> Polynomial Kernel을 이용해 분류하기 위해서는 특정한 dimension 이상의 값이 필요하고(과제에서는 3) dimension이 크다고 해서 성능에 급격한 향상을 가져오지도 않는다. 그리고 coef0값은 1 이상의 값만 주어진다면 좋은 성능을 보여줌을 알 수 있

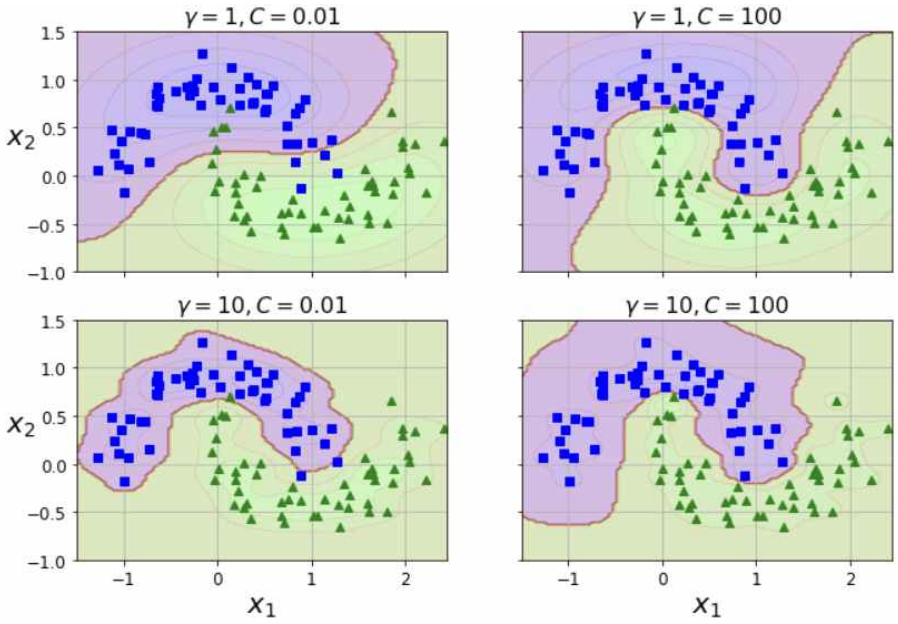
다.

Problem 4-2

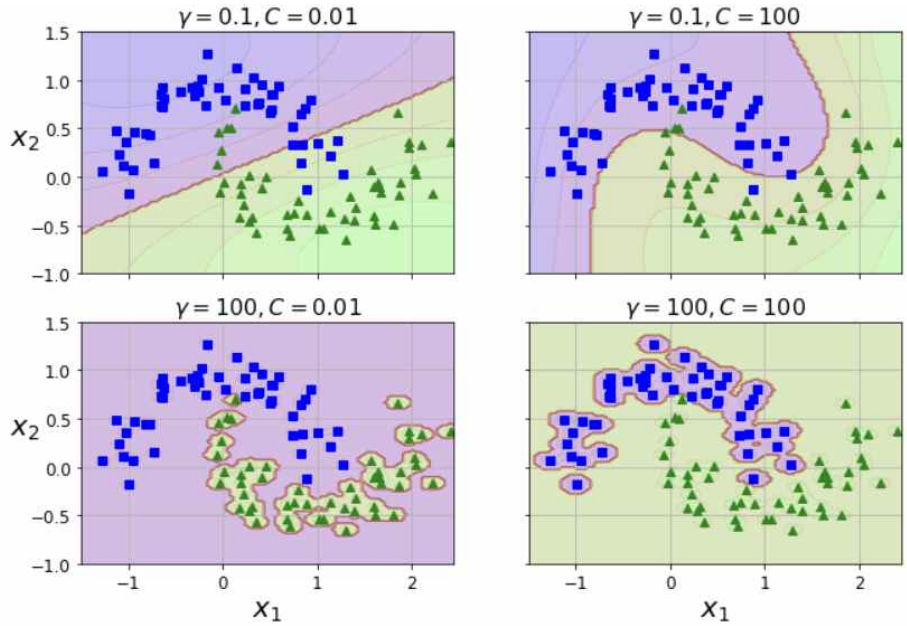
1) gamma=(0.1, 5), C=(0.001, 1000)



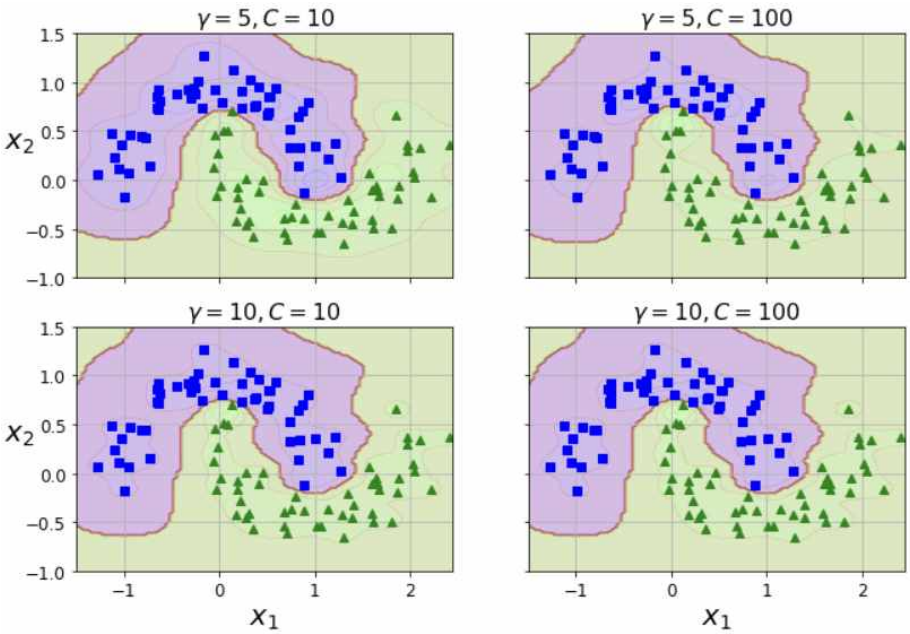
2) gamma=(1, 10), C=(0.01, 100)



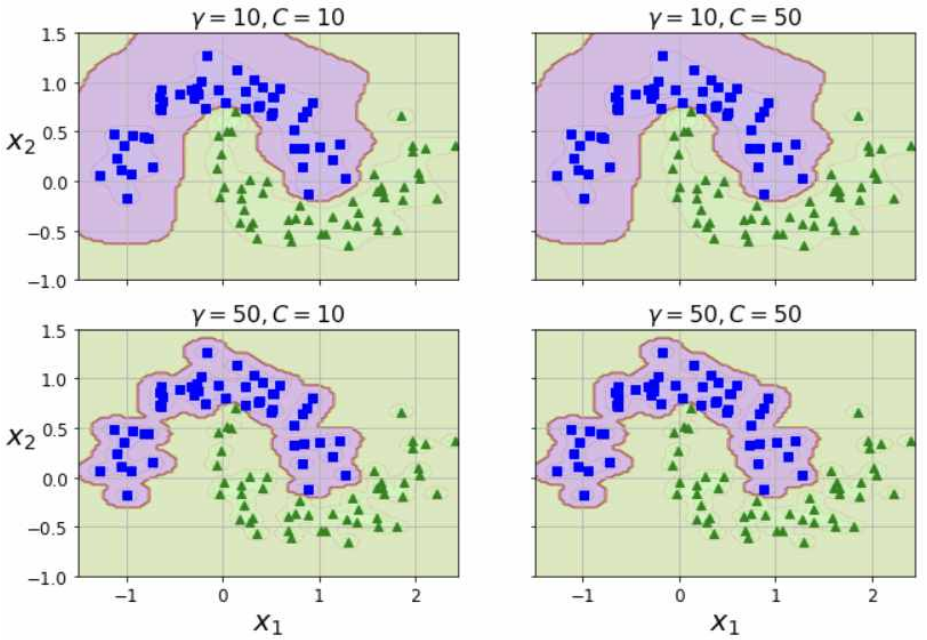
3) gamma=(0.1, 100), C=(0.01, 100)



4) gamma=(5, 10), C=(10, 100)



5) gamma=(10, 50), C=(10, 50)



-> RBF의 Gamma 값은 커질수록 경계면의 곡률이 높아지는 경향을 보이고 Gamma 값이 작아질수록 경계면의 곡률이 낮아지는 경향을 보인다. 그리고 C와 Gamma값이 반비례 관계를 가질 때 이상적인 경계면이 만들어짐을 알 수 있다. (예를 들어, Gamma = 0.1, C = 1000, Gamma = 1, C = 100)