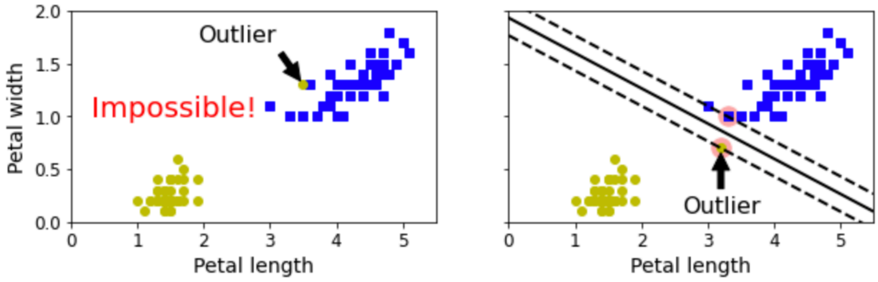
Assingment 2 결과 보고서

2017-11621 전기정보공학부 배병욱

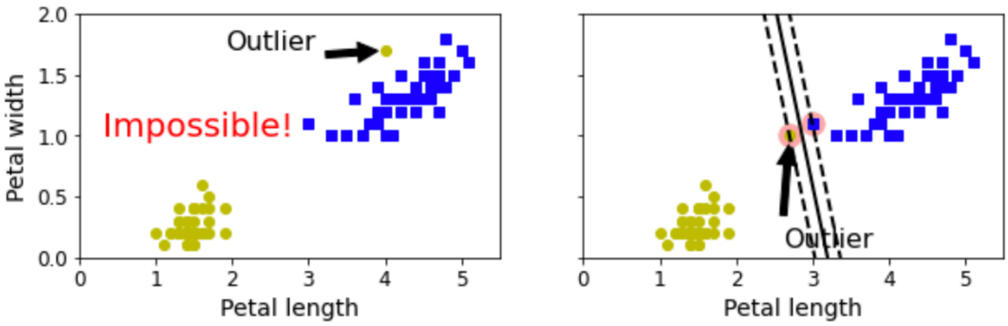
Problem 2

Problem 2에서는 outlier가 hard margin SVM에 주는 영향을 주기 위한 영향을 확인하기 위해 임의의 outlier를 추가한 후 그 결과를 plot하였다. 그 결과는 다음과 같다.

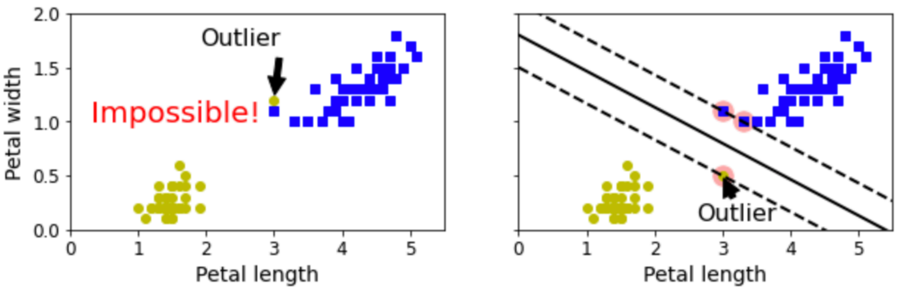
Case 1. outlier ([3.5, 1.3]) 좌측, outlier ([3.2, 0.7]) 우측



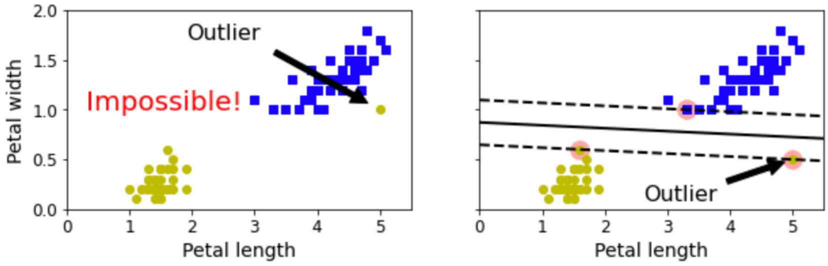
Case 2. outlier ([4.0, 1.7]) 좌측, outlier ([2.7, 1.0]) 우측



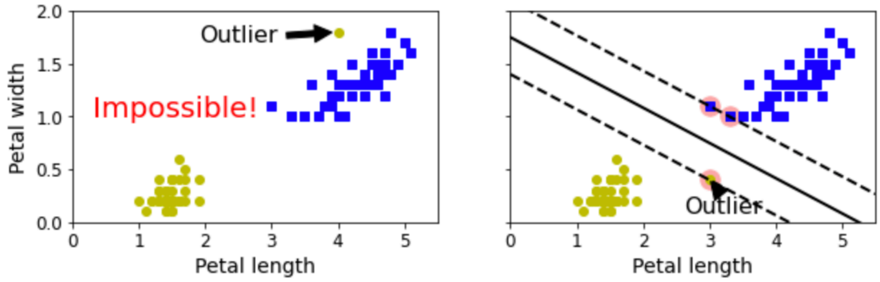
Case 3. outlier ([3.0, 1.2]) 좌측, outlier ([3.0, 0.5]) 우측



Case 4. outlier ([5.0, 1.0]) 좌측, outlier ([5.0, 0.5]) 우측



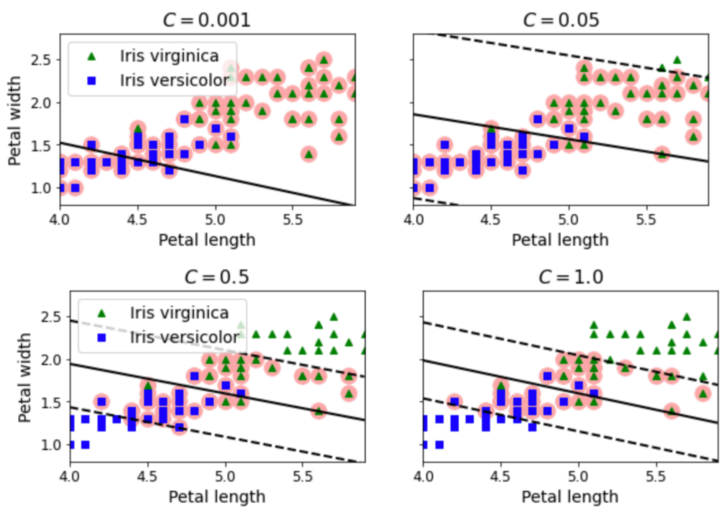
Case 5. outlier ([4.0, 1.8]) 좌측, outlier ([3.0, 0.4]) 우측

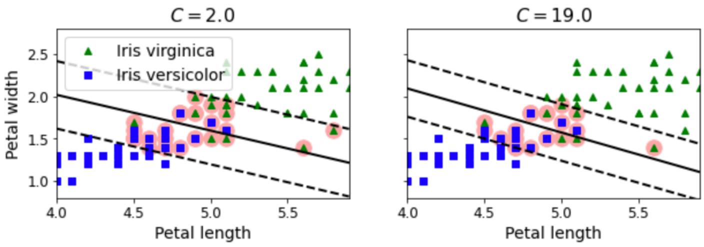


위의 Case 1~5를 확인해보면 좌측의 outlier들은 hard margin의 SVM이 불가능한 point이다. Catregory의 관점에서는 C3 category, slack variable의 관점에서는 ξ_t 가 1보다 큰 점, 즉 misclassified된 outlier들이다. 우측의 outlier들은 본래의 hard margin의 SVM에서 misclassified되지는 않았지만 margin 안의 점으로 다시 새로 SVM을 했을 때에 margin을 크게 감소시키는 점들이다. Catregory의 관점에서는 C2 category, slack variable의 관점에서는 ξ_t가 0과 1 사이의 점, 즉 fall inside band된 outlier들이다.

Problem 3

Problem 3에서는 hyperparameter의 C값의 변화에 따른 SVM의 변화를 알아보도록 한다.

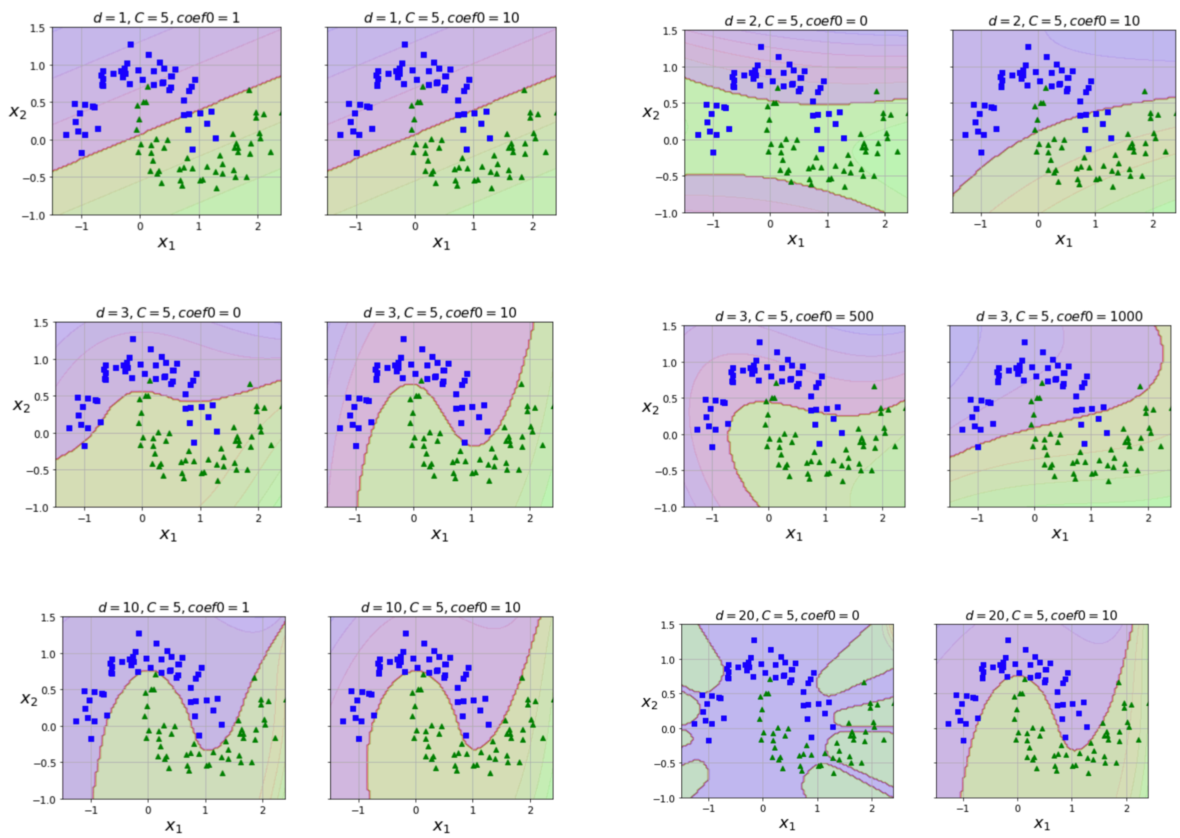




Problem 3 에서 hyperparameter C는 cost를 의미하며 model이 얼마나 많은 outlier를 허용하는지에 대한 수치이다. 즉 C가 클수록 noise에 대한 tolerance가 커지고 이에 따라 overfitting의 가능성이 커지게 된다. 위의 결과를 보면 C가 커질수록 margin이 점점 작아지는 것을 확인할 수 있다. 반대로 C가 아주 작은 0.0001과 같은 case에서는 모든 data들이 margin안에 들어가게 된다. C값이 너무 작다면 margin이 커져 overfitting할 확률은 작아지지만 underfitting 되며 반대의 경우 overfitting은 커지며 underfitting은 작아지게 된다. 즉, hyperparameter C값에 따라 tradeoff가 일어난다는 것을 확인 가능하다.

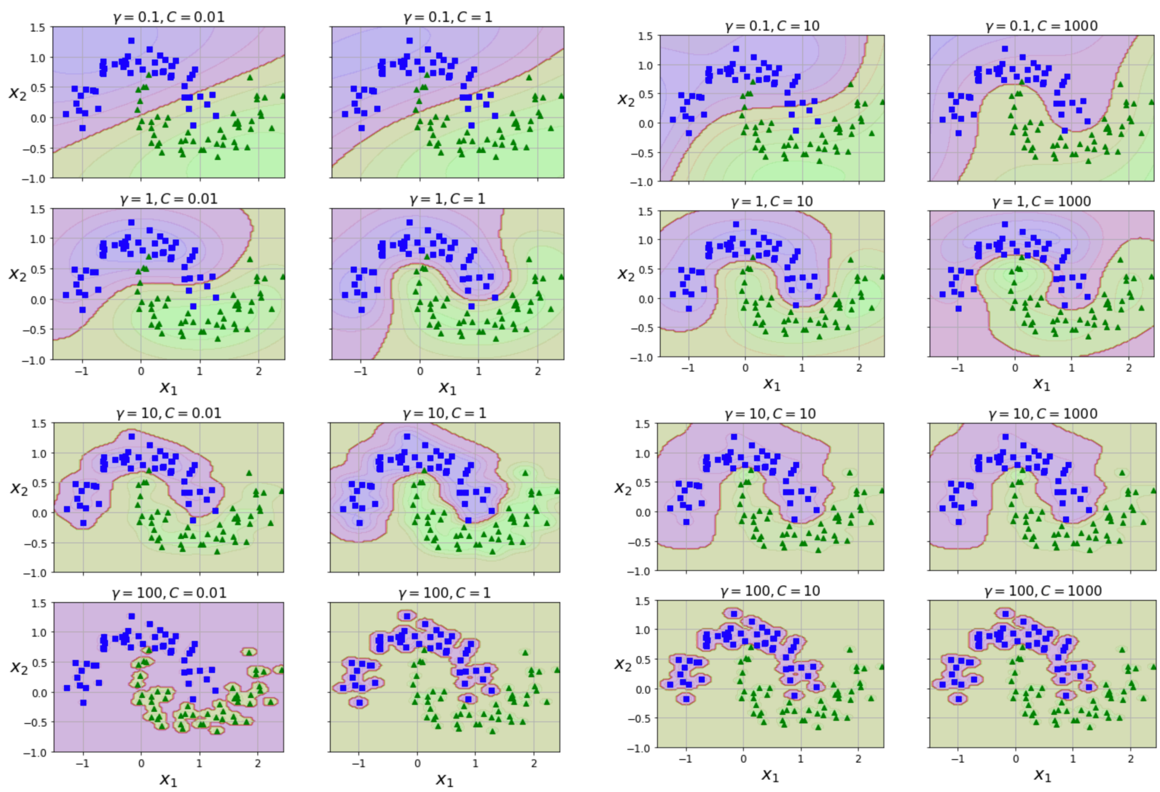
Problem 4

Problem 4에서는 nonlinear classification을 구현한다. 이 때의 nonlinear classification은 linearly separable하지 않은 input space를 nonlinear transform한 뒤에 linear model을 적용하는 방식을 취한다. 우선 Problem 4-1에서는 기존의 SVC의 Kernel을 polynomial로 사용한 뒤 degree와 coef0의 변화에 따른 차이를 확인한다. 이 때 polynomial에서 kernel function은 $(\gamma\langle x, x'\rangle + r)^d$이며, degree는 d를, coef0은 r을 결정한다.



위의 결과를 살펴보면, degree가 1일 경우에는 결국 linear separation과 똑같은 역할을 하고 있으며 degree가 커질수록 decision boundary가 더 flexible해지는 것을 확인할 수 있다. 이번 problem에서 주어진 data의 경우에는 degree가 3이상의 경우에 data를 충분히 separation할 수 있다는 것을 확인할 수 있다. Coef0의 경우에 degree가 커질수록 큰 영향을 주게 되는데 만약 $\langle x, x' \rangle$가 1보다 작은 수의 경우에 d가 커질수록 0에 가까워지게 된다. 이 경우에 위의 d=20, coef0=0의 경우처럼 data가 주는 영향이 거의 없어지게 될 수 있다. 그렇기에 적절한 coef0을 설정하는 것 또한 중요해지는데 이번 problem에서는 coef0이 10이라면 충분히 separable한다는 것 또한 확인 가능하다.

Problem 4-2에서는 kernel function으로 RBF(Radial bias function)를 사용하게 된다. Kernel function은 $\exp(-\gamma\|x - x'\|^2)$으로 hyper parameter인 gamma와 C를 변화시켜가며 그 차이를 확인한다.



Hyperparameter인 gamma는 각각의 training data가 얼마나 멀리까지 영향을 끼치는지를 결정한다. 위의 결과를 보면 gamma값이 작을 경우에 각각의 data가 굉장히 큰 영향을 끼치는 것을 확인 가능하며 클 경우에는 그 점 주위에서 영역이 형성되는 것을 확인 가능하다. C는 Problem 3에서 알 수 있듯 정확한 classification과 margin의 trade-off를 결정하는 수치이다.

gamma가 작은 값을 갖게 된다면 맨 위의 좌측 결과에서 볼 수 있듯 model이 linear classification과 유사한 형태를 띄게 된다. 아주 큰 C를 주지 않는 이상 data가 underfitting된다는 것을 알 수 있다. 반대로 gamma가 큰 값을 갖게 된다면 각각의 data는 거의 자기 자신과 아주 근접한 data만을 포함하게 된다. 이 경우에 C값이 변화해도 overfitting된다는 것을 알 수 있다.

위의 결과를 통해 gamma가 10일 경우에 가장 잘 classification됨을 알 수 있는데 이 경우에 C가 1에서 1000까지의 큰 변화를 주어도 모델에 큰 차이가 없음을 확인 가능하다. 즉, 적절한 gamma를 통해 이미 잘 regularized된 모델에는 C가 영향이 적다는 것을 알 수 있다.