

Assignment #11

머신러닝 이론과 실전

Due: November 29, 2021

- SVM 방법을 이용하여 분류분석을 수행하고자 합니다.
- SVM 은 python 의 'from sklearn import svm' 패키지를 사용하시기 바랍니다.
- 데이터는 다음과 같은 방식으로 training 1000 개, test 1000 개의 관찰값을 랜덤하게 생성합니다.

where $p_i = \text{sigmoid}(X_{1i} \cdot \sqrt{|X_{4i}| + 1} + \sin(X_{3i}) - \frac{|X_{2i}|}{\sqrt{|X_{5i}| + 1}} + 1.5|X_{6i}| + X_{7i})$, $i = 1, \dots, 1000$
where $X_1, X_2, X_3 \sim \text{Unif}(-10, 10)$ and $X_4, \dots, X_7 \sim N(0, 10)$.

1. Linear SVM 을 training 데이터를 이용하여 학습하시오. 이때 'C' 값은 0.1, 0.5, 1, 5, 10, 30, 50, 70, 90, 150 의 값을 모두 시도합니다. 그외의 파라미터는 디폴트를 사용합니다.
2. 위 1 번에서 사용된 'C' 값별로 test 데이터의 오분류율을 그림으로 나타내시오.
3. 위 2 번에서 가장 최적의 'C' 값은 무엇인가?
4. Non-linear SVM 을 수행하시오. 이때 커널함수는 rbf (radial basis function) 함수를 사용하시오. 이때 'C' 값은 0.1, 0.5, 1, 5, 10, 30, 50, 70, 90, 150 의 값을 모두 시도합니다. 그외의 파라미터는 디폴트를 사용합니다.
5. 위 4 번에서 사용된 'C' 값별로 test 데이터의 오분류율을 그림으로 나타내시오.
6. 위 5 번에서 가장 최적의 'C' 값은 무엇인가?
7. 위 3 번과 6 번의 'C'값으로 아래의 결과물을 output 파일에 추가하시오.

결과물 예시)

Linear Kernel SVM (C=0.1)

```
Confusion Matrix (SVM-linear)
-----
                Predicted Class
                1      2
Actual   1      239   14
Class    2       12  153

Model Summary (SVM-linear)
-----
Overall accuracy = .793
```

RBF Kernel SVM (C=50)

```
Confusion Matrix (SVM-rbf)
-----
                Predicted Class
                1      2
Actual   1      239   14
Class    2       12  153

Model Summary (SVM-rbf)
-----
Overall accuracy = .793
```