신경망, SVM 과제

* 이 과제에서는 결정나무, ANN, SVM의 성능을 비교한다.

1. 나무모델에서 사용하였던 데이터를 사용함. 또는, coupon data 또는 HCV data – UCI 사용. 이 데이터로 인공 신경망을 학습시켜라.   
   단, hidden layer는 1 개로 가정한다. 또한, 명목형 변수를 2개 이상 포함하도록 한다.   
   아래 사항을 참고 바랍니다.
   1. 결측치 처리, 변수 타입변환, 데이터 변환 등 적절한 전처리 과정이 필요합니다.
   2. 각 파라미터에 대한 최적화가 필요합니다. (CV 필요)   
      최적 모델을 제시할 것.
   3. 적절한 plot 또는 학습곡선을 제시하여 모델이 최적화되었음을 보여야 합니다.
   4. 인공 신경망의 경우 epoch 수 vs 성능의 학습곡선이 필요함.   
      - 모델이 충분한 학습이 되었는지를 확인
2. 위 문항의 내용에 준하여 SVM 모델을 학습시켜라.
3. 결정나무, 인공신경망, SVM 모델의 성능을 비교하라.
4. 어떠한 데이터들이 분포기반과 경계기반 모델에 각각 적합한지 토론해 보라.   
   이상치/극단치에 더 강건한-robust 한 모델은 무엇인가?  
   분포기반 – 분포는 소수의 이상치가 있더라도 크게 변하지 않는다.   
   경계기반 – 소수의 이상치가 경계 부근에 있다면 경계를 왜곡시킬 수 있다.

- 끝 -