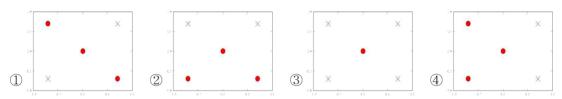
2021학년도 2학기 기계학습 기말고사 (50점 만점)

아래의 선다형 문제는 보기에서 옳은 것을 모두 고르는 문제입니다.

(선다형 1번) 다음 각 보기는 5개의 훈련 데이터 샘플 $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ 이 속하는 두 부류 (class)를 o와 \mathbf{x} 로 나타낸 것이다. $\phi(\mathbf{x}) = \mathbf{x}$ 로 정의되는 입력 특성에 퍼셉트론 알고리즘을 적용하여 유한 번 만에 결정 경계를 구할 수 있는 보기를 모두 고르시오. (선다형 3점)



(선다형 2번) 다음 500개의 입력 특성을 이용하는 분류모델을 훈련 데이터로 학습시켜 90%의 정확도(accuracy)를 얻었지만, 이 모델을 이용하여 구한 테스트 데이터셋에 대한 정확도는 70%에 그쳤다. 다음 보기 중 이 문제점을 해결하기 위한 방안으로 옳은 것을 모두고르시오. (선다형 3점)

- ① 고려하는 입력 특성의 개수를 줄여서 학습시킨다.
- ② 입력 특성을 새롭게 추가한다.
- ③ 훈련 데이터를 이용하여 학습시킬 때 교차검증을 이용한다.
- ④ 다른 모델을 선택한다.

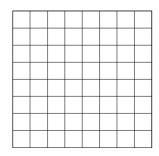
아래의 단답형 문제는 간단한 계산 근거와 답을 쓰면 되고, 서술형 문제는 상세한 풀이과정을 모두 써야 합니다.

(단답형 3번) 아래 표와 같이 7개의 훈련 데이터 샘플이 주어질 때, 다음 물음에 답하시오.

$\boldsymbol{x} = (x_1, x_2)$	(1,1)	(2,2)	(3,1)	(4,5)	(5,7)	(6,5)	(7,7)
y	1	1	1	-1	-1	-1	-1

(3-1) 이 훈련 데이터셋을 이용하여 하드마진 서포트벡터머신 분류기를 학습시킬 때, 서포 트벡터를 모두 구하시오. (단답형 3점)

(3-2) 레이블이 -1인 훈련 샘플 (5,1)을 추가하여 (1-1)에서 구한 분류기를 다시 학습시킨 때, 서포트벡터를 모두 구하시오. **(단답형 4점)**



(이 모눈종이를 활용할 수 있습니다.)