**[첫 번째 과제 - p. 66 문제 2.1, 문제 2.2]**

**IT응용시스템공학과**

**1494053**

**김희택**

**2.1 아래의 문제를 해결하기 위해 데이터마이닝 기법을 이용하려고 한다. 지도학습과 비지도학습 중 어느 기법을 사용하는 것이 보다 적절한지 구분하시오.**

a. 과거 소비자들의 인구통계학적, 금융 데이터를 바탕으로 새로운 소비자들의 대출여부를 결정

* 지도 학습 (입력변수와의 관계를 통해 모델링하여 새로운 입력데이터의 목표변수 값을 예측하므로)

b. 온라인 서점에서 기존 구매 고객들의 구매 패턴을 파악하여 책을 추천

* 비지도 학습 (과거 데이터를 분석하여 구매패턴을 파악해 협력적 필터링을 통한 상품 추천)

c. 기존에 알고 있는 네트워크 데이터 패킷 상태정보를 이용하여 새로운 네트워크 데이터패킷의 상태분류 (예: 바이러스, 해커 공격)

* 지도 학습 (입력변수와의 관계를 통해 모델링하여 새로운 입력데이터의 목표변수 값을 분류하므로)

d. 비슷한 성향을 가지고 있는 고객군 파악

* 비지도 학습 (비슷한 성향을 가진 고객군을 통해 다른 특정 성향의 고객군을 찾는 연관 규칙 방법)

e. 파산한 회사와 그렇지 않은 회사들의 재정상태 데이터를 기반으로 특정회사의 파산여부를 예측

* 지도 학습 (입력변수와의 관계를 통해 모델링하여 새로운 입력데이터의 목표변수 값을 예측하므로)

f. 수리이력 정보를 바탕으로 항공이 수리 시간을 추정

* 지도 학습 (입력변수와의 관계를 통해 모델링하여 새로운 입력데이터의 목표변수 값을 예측하므로)

g. 우편번호 스캔 결과 정보를 바탕으로 우편물 자동분류

* 지도 학습 (입력변수와의 관계를 통해 모델링하여 새로운 입력데이터의 목표변수 값을 분류하므로)
* 비지도 학습도 가능 (비슷한 입력변수끼리 묶어준다면)

h. 본인의 구매물품과 다른 고객들의 과거 구매이력정보를 바탕으로 할인 쿠폰의 발행 여부 판단

* 비지도 학습 (본인의 구매물품과 과거 고객들의 구매결과들의 연관 규칙을 통해 어떤 물품에 대해 할인 쿠폰을 발행할 지 판단하므로)
* 지도 학습도 가능

**2.2 검증데이터와 평가데이터의 차이를 서술하시오.**

평가데이터는 모델의’**최종** **성능**’을 평가하기 위해서 쓰인다. 즉 트레이닝의 과정에 관여하지 않는 것이다. 하지만 검증데이터는 여러 모델 중에서 최종 모델을 선정하기 위한 성능 평가에 관여한다. 즉 트레이닝 과정에 관여하는 것이고 트레이닝이 된 여러 가지 모델 중 가장 좋은 하나의 모델을 고르기 위한 셋이다.