

## <3주차 – 머신러닝 개요>

### 1. 머신러닝 개요

1. 머신러닝의 학습법(교사법) 중 하나인 **강화학습**에 대하여 서술하시오.  
(정의 및 사례 1가지 포함.)

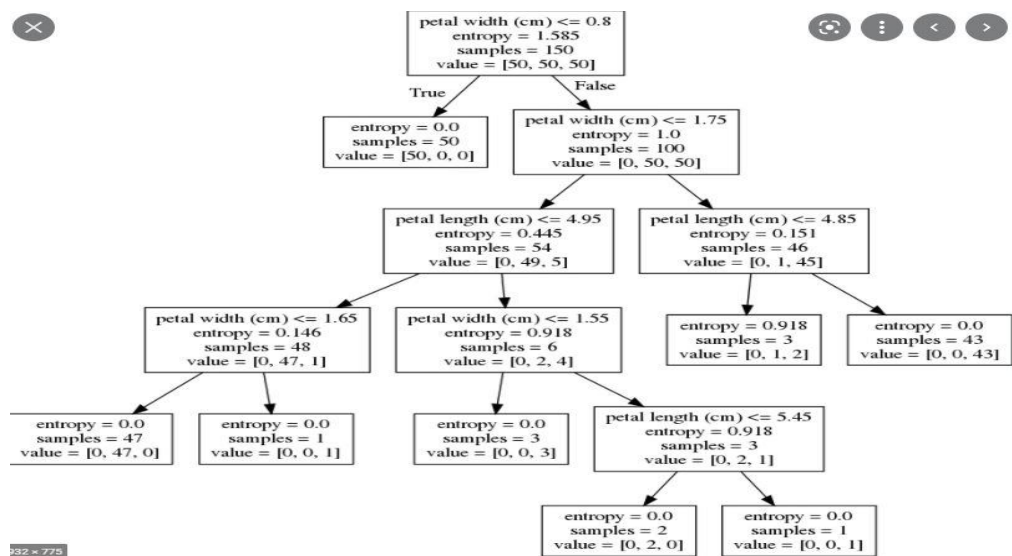
**강화학습**이란 현재의 상태에서 어떤 행동을 취하는 것이 최적인지를 학습하는 것입니다. 행동을 취하는 경우 외부환경에서 보상이 주어지고 이러한 보상을 최대화하는 방향으로 학습이 진행됩니다. 대표적인 예로는 AI에게 주입한 인슐린 양에 따른 혈당이 어떻게 변하는지 나타낸 데이터를 학습시켜 당뇨병 환자의 혈당 데이터를 입력하는 경우 최적의 인슐린양을 찾도록 하는 강화학습을 이용한 AI알고리즘이 있습니다.

## 2. 지도학습 비지도학습

2-(1) 지도학습과 비지도학습을 **차이점** 기준으로 설명해주세요.

지도 학습이란 정답이 있는 데이터를 활용하여 데이터를 학습시키는 것입니다. 대표적인 예로는 분류, 회귀가 있습니다. 비지도 학습이란 정답 라벨이 없는 데이터를 비슷한 특징끼리 군집화 하여 새로운 데이터에 대한 결과를 예측하는 방법입니다. 대표적인 예로는 군집(clustering), k-means가 있습니다.

2-(2) 다음은 지도학습 중, ( 결정트리 )에 해당하고, ( graphviz )기법에 해당한다.



2-(3) 2-(1)에서 나온 이 기법은 하부구조가 너무 많으면, ( 과대적합 ) 문제가 발생한다. 이를 피하기 위해서 ( max\_depth )를 통해 트리가 너무 복잡하지 않도록 단순화시킨다.

2-(4) 비지도학습의 대표적인 예로 '군집'(Clustering)이 있는데 아래의 절차로 진행하는 기법의 이름은?

1. 일단  $K$  개의 임의의 중심점(centroid)을 배치하고
2. 각 데이터들을 가장 가까운 중심점으로 할당한다. (일종의 군집을 형성한다.)
3. 군집으로 지정된 데이터들을 기반으로 해당 군집의 중심점을 업데이트한다.
4. 2 번, 3 번 단계를 그래서 수렴이 될 때까지, 즉 더이상 중심점이 업데이트 되지 않을 때까지 반복한다.

**K 평균 군집화입니다.**

### 3. 분류/회귀

3-(1) 분류와 회귀의 **공통점** 및 **차이점**을 서술하시오.

공통점으로는 둘 다 지도학습을 사용합니다.

차이점으로는 가지고 있는 데이터에 독립, 종속변수가 있고 예측하고자 하는 값이 숫자인 경우 회귀를 이용합니다.

가지고 있는 데이터에 독립, 종속변수가 있고 예측하고자 하는 값이 이름이나 문자(범주형 데이터)인 경우 분류를 사용합니다.

3-(2) 회귀 방법론에는 **회귀분석만**이 존재한다. (O/X)

※만약, O라면 어떤 방법들이 있는지 기법 이름과 함께 분석 방법을 2줄 이내로 서술하고, X라면 어떤 다른 방법들이 있는지 기법 이름과 함께 분석 방법을 2줄 이내로 서술하시오.

X -> 로지스틱 회귀의 경우는 분류 모델의 역할을 합니다. 이진분류의 경우는 시그모이드 함수 다중분류의 경우는 소프트맥스 함수를 사용합니다.

3-(3) 로지스틱 회귀분석은 '**회귀**'에 해당하는가, '**분류**'에 해당하는가? 답을 쓰고 이유를 2줄 이내로 서술하시오.

분류 -> 이진분류의 경우 시그모이드 함수 출력 값을 통해 0.5이상이면 양성클래스, 0.5이하이면 음성클래스로 분류합니다.

3-(4) 멀티 클래스 분류와 멀티 레이블 분류의 사례를 하나만 서술하시오.  
(책 교재 및 PPT 사례 제외)

멀티 클래스 분류의 경우 손글씨 숫자를 인식하는 인공 신경망 사례가 있습니다. 멀티 레이블의 경우 실시간 전략 게임(클래시로얄)의 경기 데이터와 멀티 레이블 분류를 사용해 상대방의 행동을 예측하는 사례가 있습니다.

## 4. 과적합

4-(1) 과소적합(Underfitting)이 발생하는 이유와 해결방법을 서술하시오.

모델이 단순해서 데이터의 내재된 구조를 학습하지 못하는 경우 발생합니다. 해결방법으로는 모델을 바꾸거나 규제하는 하이퍼 파라미터 값이 있는 경우 줄이는 방법이 있습니다.

4-(2) 과대적합(Overfitting)이 발생하는 이유와 해결방법을 서술하시오.

훈련데이터에 지나치게 학습이 되어서 결국 테스트 데이터에서의 score점수나 accuracy점수가 훈련데이터보다 많이 낮게 됩니다. 사이킷런의 경우 릿지, 라쏘로 특성 규제를 하였고 딥러닝의 경우 `keras.callbacks.EarlyStopping`을 하여 조기종료를 하는 방법이 있습니다.