

<3주차 – 머신러닝 개요>

1. 머신러닝 개요

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

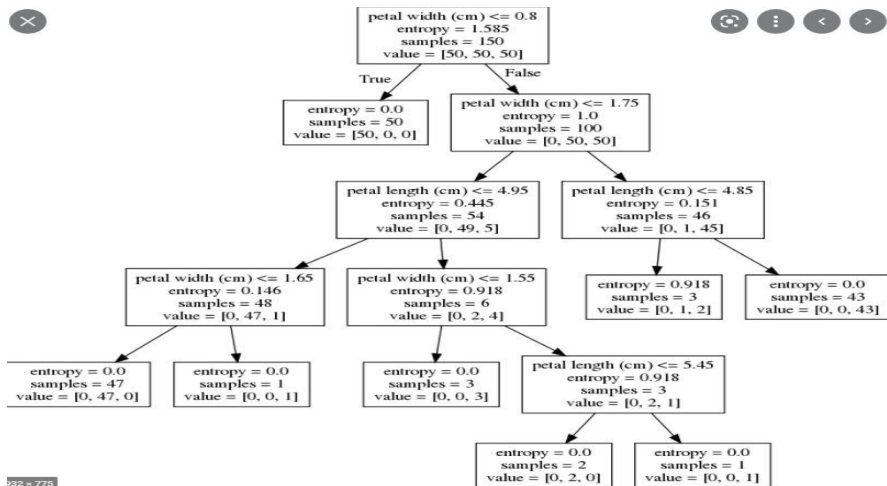
답: 시행착오를 거쳐 보상을 극대화 하는 행동을 찾는 것이다. 대표적인 예로, 자율주행 자동차 주차가 있다. 자동차가 장애물들에 충돌하면 벌점을 주고, 주차 위치와 가까워질 경우 보상을 주면서 수많은 시도 끝에 정확한 위치에 주차할 수 있도록 학습하는 것이다.

2. 지도학습 비지도학습

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

답: 지도학습은 독립변수에 따른 종속변수, 즉 정답이 있는 데이터를 활용하는 것이고, 비지도학습은 독립변수에 따른 종속변수가 없이, 정답을 알려주지 않고 분석 모델을 학습시킨다.

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



답: 의사결정나무 / 분류분석+회귀분석

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

답: 과대적합 / 가지치기

2-(4) 비지도학습의 대표적인 예로 '군집'(Clustering)이 있는데 아래의 절차로 진행되는 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.

답: K – means 군집화

3. 분류/회귀

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

답:

1. 공통점: 독립변수에 따른 종속변수가 존재하는, 정답이 있는 데이터를 활용해 분석 모델을 학습시키는 지도학습이다.
2. 차이점: 분류 분석은 이산 값을 포함하는 유한 집합에서 레이블을 예측하는 것이고, 회귀 분석은 연속형 값을 포함하는 유한 집합에서 연속된 값을 예측하는 것이다.

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.

답: (X),

회귀 방법론에는 분류에서 주로 쓰이는 의사결정나무, ANN, SVM도 쓰인다.

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

답: 분류 분석에 해당한다.

로지스틱 회귀분석은 분류하려는 범주가 성공/실패, 예/아니오, 남자/여자 등 2가지 범주로 나뉜 경우에 적용한다.

3-(4) 멀티 클래스 분류와 멀티 레이블 분류의 사례를 하나만 서술하시오.

(책 교재 및 PPT 사례 제외)

답: 멀티 클래스 분류의 예시로 대상이 되는 객체(사과)가 있으면, 그 객체는 과일 카테고리(사과, 배, 포도 등)에 속하는 경우를 들 수 있다. (+ 학생들의 성적을 A+, A0, ..., D+, D0, F 중 하나로 분류한다.)

멀티 레이블 분류의 예시로는 사과, 배를 입력 받으면 과일 카테고리(여기서는 사과, 배)로 분류되는 것을 들 수 있다.

4. 과적합

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

답: 데이터에서 충분한 특징을 찾아내지 못하고 머신러닝 모델을 학습할 때 발생한다.

Bias를 줄이고, Variance를 높인다. 즉, Variance가 높은 모델을 사용해본다.

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

답: 필요 이상의 특징을 발견해 복잡한 모델이 선택된다.

더 많은 학습데이터를 확보하고, 특징들의 수치 값을 정규화 한다.