**REPORT**

(CHAPTER 1~5)

로고, 상징, 등록 상표, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

|  |  |
| --- | --- |
| 이름 | 김민서 |
| 학과 | 컴퓨터인공지능공학부 |
| 학번 | 202213083 |
| 과목명 | 인공지능입문 |
| 교수님 | 진예지 교수님 |
| 분반 | 105분반 |
| 제출일 | 2025.04.17 |

**1 – 1. 인공지능의 뜻은 무엇인가? 인공지능의 장점과 단점을 인간의 지능과 비교하여 설명하시오.**

* 인공지능
  + 사람이 만든 지능
  + 사람이나 동물의 ‘지능’과 유사하게 동작하는 무엇이라고 할 수 있음
* 전체적인 밸런스 측면에서는 자연의 것들을 따라가기 힘들지만, 특정 기능에 한정한다면 자연의 기능을 능가하는 경우가 많음
* 인간 ‘지능’의 모든 기능을 가지고 있는 것은 아니지만, 대용량의 계산이나 정보의 취합, 정형화된 데이터의 예측과 분류라는 측면에서 인간을 추월해 발전함

**1 – 2. 인공지능이란 개념이 처음 등장한 시대가 언제인지, 그리고 해당 시기에 최초로 등장한 것들을 설명하시오.**

* 딥러닝의 원시 형태인 ‘퍼셉트론’은 1957년에 발표
* 기계학습에 사용되고 있는 ‘역전파 알고리즘’은 1980년대 중반에 캘리포니아 대학에서 발표
* 합성곱 신경망은 1979년에 NHK 연구원이 ‘네오코그니트론’이란 이름으로 발표
* 1980년대 중반 ~ 2010년대 중반 -> 다양한 인공지능 모델들이 발표됨
  + 볼츠만 머신(인공신경망)
  + 서포트 벡터 머신(데이터의 분포를 바탕으로 최적 초평면을 찾음)
  + 게임 트리(보드게임에서 가능한 경우의 수를 체크하여 최선의 수를 찾음)
  + 의사결정 트리(주어진 데이터에서 추출되는 각종 특징들의 경우의 수를 활용하여 분석)
  + 진화 알고리즘(솔루션의 각종 특징들을 자르고 이어 붙이며 솔루션을 자동으로 개선)
  + 랜덤 포레스트(의사결정 트리의 과적합 문제를 해결)

**3 – 4. 이산화란 무엇이며, 데이터를 이산화하는 이유에 대해 설명하시오.**

* 이산화 : 수치형 변수에 구간을 지정하여 범주형 데이터로 변환하는 것
  + 연속적인 데이터에 비해 범주형 데이터가 시각화하기 편함

**3 – 5. 상관계수 -1, 0, +1인 데이터의 분포를 예를 들어 설명하시오.**

* 상관계수 : 두 변수 간의 연관성을 알려주는 지표
  + -1 ~ 1의 값을 가짐
  + 크기가 클수록 두 변수 간 연관성이 높음
  + 상관계수의 부호는 연관성의 방향을 의미함

텍스트, 라인, 그래프, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

* 상관계수의 예
  + 키와 체중의 상관관계
  + 키가 클수록 체중이 높은 경우가 많으므로, 양의 상관관계를 가짐

**4 – 1. 데이터 전처리의 필요성에 대해 설명하시오.**

* 결측치 삭제
  + 결측치란 데이터 셋에서 알려지지 않거나 수집되지 않거나 잘못 입력된 값
  + 하나의 결측치로 인해 전체 데이터에 평균 등에 큰 지장이 가기 때문에 데이터 분석에 앞서 전처리 해야 함.

**4 – 2. 아래 위젯들이 속한 카테고리, 이름과 기능을 설명하시오.**

1. Impute 위젯의 기능
   * 데이터셋에 누락된 값이 있을 경우 이를 다양한 방법으로 보정하는 기능
2. Average/Most frequent 방법
   * 수치형은 평균값으로, 범주형은 최빈값으로 누락된 값을 채움
3. As a distinct value 방법
   * 결측값을 해당 속성에 새로운 고유한 값으로 치환
4. Fixed values 방법
   * 사용자가 직접 지정한 고정값으로 결측 데이터를 대체
5. Remove instances with unknown values 방법
   * 결측값이 포함된 행 전체를 제고

**4 – 5. 다음은 Select Rows와 Select Colums 위젯의 아이콘이다. 1) ~ 5)에 대해 설명하시오.**

1. Select Rows 위젯의 기능
   * 특정 조건을 만족하는 행들만 선택하거나 제외하는 기능
2. Select Columns 위젯의 기능
   * 전체 속성 중에서 필요한 속성만 선택하거나, 제거하는 기능
3. Is not 조건식
   * 해당 값이 지정한 값과 일치하지 않을 때 해당 행을 선택
4. Is between 조건식
   * 값이 특정 두 값 사이에 있는 경우 선택
5. Is outside 조건식
   * 값이 특정 범위 바깥에 있을 때 선택

**5 – 4. 계층적 군집 분류(Hierarchical Culstering)는 비지도 학습 방법 중 하나이다. 계층적 군집 분류에는 데이터 샘플 간의 거리를 계산하는 4가지 방법이 있다. 계층적 군집 분류에 대해 설명하고 인스턴스 간의 거리를 계산하는 방법에 대해 설명하시오.**

* 유클리드 거리
  + 실수 공간에서의 직선 거리
* 멘하탄 거리
  + 각 축별 거리의 합
* 코사인 거리
  + 원점에서의 각도 차이
* 자카드 거리
  + 두 집합에서 서로 다르게 나타나는 특징의 비율(자연어 처리에서 활용)

**5 – 5. MDS 위젯과 Scatter Plot 위젯의 차이에 대해 설명하시오.**

* MDS - 거리 기반의 차원 축소 시각화
  + 고차원 데이터의 구조를 유지한 채 2차원으로 차원 축소하여 시각화
  + 데이터 간의 유사성/거리에 기반하여 관계를 시각화
  + 데이터 분포의 전반적인 패턴 파악에 유용
  + 유사한 데이터 샘플들은 가까이에, 서로 다른 데이터 샘플들은 멀리 배치
* Scatter Plot – 속성 간의 직접 비교 시각화
  + 데이터의 두 속성을 선택하여 해당 값들을 좌표축에 직접 매핑하여 시각화
  + 변수 간 관계나 상관성 파악에 사용됨
  + 특정 특성에 따른 군집 확인에 사용됨