|  |  |  |  |
| --- | --- | --- | --- |
| **3주. Machine Learning Concept** | | | |
| 학번 | 32191197 | 이름 | 김채은 |

|  |
| --- |
| Q1. 전통적인 SW 와 머신러닝을 적용한 SW 의 차이점을 설명하시오 |

전통적인 SW는 규칙을 인간이 알아내서 알고리즘의 형태로 구현한 것이다. 머신러닝은 인간이 규칙을 알아내는 방법을 제시하고 기계가 규칙을 알아낸다. 이 과정을 학습(훈련)이라고 한다.

|  |
| --- |
| Q2. 머신러닝에서 러닝(learning)의 실제적인 의미를 설명하시오. |

많은 양의 데이터를 통해 규칙을 찾아내고, 예측 모델을 만들어내는 것이다.

|  |
| --- |
| Q3. 머신러닝이 가능한 이유를 설명하시오. |

사람이 많은 데이터를 가지고 규칙을 학습할 수 있도록 프로그래밍하기 때문이다.

|  |
| --- |
| Q4. 회귀(regression) 와 분류(classification)의 차이점을 설명하시오 |

회귀는 반응변수가 수치로 나타나고, 분류는 범주로 나타난다.

|  |
| --- |
| Q5. 기후변화에 따른 연평균 기온을 예측하는 머신러닝 모델을 만들려고 한다. (2점)  1) 모델을 만들기 위해 필요한 것은 무엇인가  2) 이 모델은 회귀, 분류, 군집화, 강화학습중 어느 기술을 적용해야 하는가? 그 이유는 무엇인가 |

1) 이전 연평균 기온 데이터, 변화 추세

2) 기온이 반응변수이고 수치로 나타내기 때문에 회귀를 적용해야 한다.

|  |
| --- |
| Q6. 머신러닝 모델을 개발할 때 데이터셋을 training data 와 test data 로 나누는 이유는 무엇인가? 나누지 않는다면 어떤 문제가 발생하는가 |

모델이 훈련 셋에 적합하게 만들어지므로 훈련 셋으로 테스트를 하면 모델의 오차를 제대로 측정할수 있고, 따라서 실제로 모델을 적용할 때 오차가 클 수 있다.

|  |
| --- |
| Q7. scikit-learn 홈페이지([https://scikit-learn.org/stable/](http://scikit-learn.org/stable/))를 방문하여 scikit-learn에서 제공하는 군집화(clustering) 일고리즘에는 어떤 것들이 있는지 찾아서 제시하시오 |

Demo of DBSCAN clustering algorithm, K-means Clustering

|  |
| --- |
| Q8. Pandas 모듈을 이용하여 배포된 데이터셋중 cars 데이터셋을 읽어온 후 다음 문제를 해결하시오 (2점)  (1) 데이터셋의 위쪽 5행을 보이시오  (2) 데이터셋의 컬럼들 이름을 보이시오  (3) 데이터셋의 두 번째 컬럼의 값들만 보이시오.  (4) 데이터셋의 11~20행 자료중 speed 컬럼의 값들만 보이시오.  (5) speed 가 20 이상인 행들의 자료만 보이시오  (6) speed 가 10 보다 크고 dist 가 50보다 큰 행들의 자료만 보이시오.  (7) speed 가 15 보다 크고 dist 가 50보다 큰 행들은 몇 개인지 보이시오 |

**Source code :**

|  |
| --- |
| import pandas as pd  cars = pd.read\_csv('C:/Users/82104/Desktop/deeplearning/cars.csv')  # 1  print(cars.head(5))  # 2  print(cars.columns)  # 3  print(cars.iloc[:, 1])  # 4  print(cars.iloc[11:21, 0])  # 5  speed = cars['speed'] >= 20  print(cars[speed])  # 6  dist = cars['dist'] > 50  print(cars[speed & dist])  # 7  speed = cars['speed'] >= 15  print(len(cars[speed & dist])) |

**실행화면 캡쳐:**











