**Machine Learning**

**22-09(K) 빅데이터 분석을 활용한 예측&AI 모델링**

|  |  |  |  |
| --- | --- | --- | --- |
| **성명** | **이선무** | **점수** |  |

1. **다음은 k-최근접 이웃 알고리즘에 관한 설명이다. 바르지 않은 것은?( 3 )**

|  |  |
| --- | --- |
| ① | 새로운 데이터에 대한 예측은 가장 가까운 훈련 데이터의 포인트를 찾아서 결정함 |
| ② | 훈련 데이터에서 새로운 데이터 포인트에 가장 가까운 k개의 이웃을 찾음 |
| ③ | 찾은 이웃들의 클래스 중 가장 가까운 클래스를 예측값으로 사용 |
| ④ | K를 1로 지정할 수도 있음 |

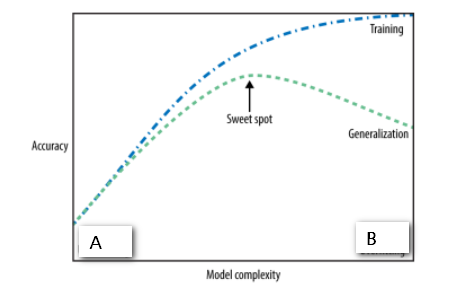
1. **다음은 회귀(Regression)에 관한 설명이다. 잘못된 것은?( 2 )**

|  |  |
| --- | --- |
| ① | 연속적인 숫자를 예측 |
| ② | 미리 정의된 가능성있는 여러 클래스 레이블 중 하나를 예측 |
| ③ | 대표적인 예로는 사람의 나이, 연간 소득 또는 농작물의 수확량 예측 등이 있음 |
| ④ | 출력값의 작은 차이는 문제가 되지 않음 |

1. **지도학습에서 모델이 처음 보는 데이터에 대해 정확하게 예측할 수 있다면 이를 훈련세트에서 테스트 세트로 “이것” 되었다고 한다. 이것은 다음 중 어느것인가? ( 1 )**

|  |  |
| --- | --- |
| ① | 일반화 |
| ② | 과대적합 |
| ③ | 과소적합 |
| ④ | 복잡화 |

1. **다음은 지도학습의 훈련과정을 나타낸 그래프이다. B에 들어갈 내용은? ( 4 )**



|  |  |
| --- | --- |
| ① | 과소적합 |
| ② | 일반화 |
| ③ | 복잡화 |
| ④ | 과대적합 |

1. **다음은 k-최근접 이웃 알고리즘에 관한 설명이다 나머지 셋과 다른 내용은? ( 2 )**

|  |  |
| --- | --- |
| ① | 이웃을 하나 선택함 |
| ② | 결정 경계가 부드러워짐 |
| ③ | 훈련 데이터의 정확도 증가 |
| ④ | 모델의 복잡도 증가 |

1. **다음은 k-최근접 이웃 알고리즘의 단점에 관한 내용이다. 바르지 않은 것은? ( 4 )**

|  |  |
| --- | --- |
| ① | 훈련 세트가 매우 크면 예측이 느려짐 |
| ② | k-nn 알고리즘을 사요할 때는 데이터의 전처리가 매우 중요 |
| ③ | 수백 개 이상의 많은 feature를 가진 데이터에는 잘 동작하지 않음 |
| ④ | k-nn은 이해하기 어려울 뿐만 아니라 많은 조정이 필요함 |

1. **회귀를 위한 이 모델은 feature가 하나일 때는 직선, 두 개일 때는 평면이 되며 더 높은 차원에서는 초평면이 되는 특성이 있다. 이 모델은 무엇인가?( 1 )**

|  |  |
| --- | --- |
| ① | 선형 모델 |
| ② | 이진 분류 모델 |
| ③ | 회귀 모델 |
| ④ | 다중 분류 모델 |

1. **회귀를 위한 선형 모델 중 L1 규제를 사용하며, 일부 feature들은 0이 되기도 함, 모델의 가장 중요한 feature가 무엇인지 드러내 주는 효과가 있음. 이는 다음중 어떤 방법인가?( 3 )**

|  |  |
| --- | --- |
| ① | 최소 제곱법 |
| ② | Ridge |
| ③ | Lasso |
| ④ | Classification |

1. **분류용 선형 모델인 LinearSVC에 관한 설명이다. 바르지 않은 것은?( 2 )**

|  |  |
| --- | --- |
| ① | 매개변수 C를 이용하여 규제를 가함 |
| ② | 매개변수 C의 값을 아주 크게 하면 일반화가 잘 됨 |
| ③ | 매개변수 C의 값을 작게 하면 규제가 많이 적용됨 |
| ④ | 낮은 차원에서의 데이터에서는 결정경계가 직선이거나 평면이 됨 |

1. **다음은 나이브 베이즈 분류기에 관한 설명이다. 잘못된 것은? ( 4 )**

|  |  |
| --- | --- |
| ① | 기존의 선형 분류기 보다 속도가 빠름 |
| ② | 각 feature를 개별로 취급, 파라미터를 학습함 |
| ③ | GaussianNB는 연속적인 어떤 데이터에도 적용할 수 있음 |
| ④ | BernoulliNB는 카운트 데이터에 사용 |

1. **다음은 나이브 베이즈 분류기의 장단점에 관한 설명이다. 바르지 않은 것은? ( 1 )**

|  |  |
| --- | --- |
| ① | alpha가 크면 더 완만해지고 모델의 복잡도는 높아짐 |
| ② | BernoulliNB, MultinomiaNB 는 모델의 복잡도를 조절하는 alpha 매개변수 하나를 가짐 |
| ③ | GaussianNB 은 대부분 매우 고차원적인 데이터셋에 사용 |
| ④ | MultinomiaNB 는 보통 0이 아닌 feature가 비교적 많은 데이터셋에서 BernoulliNB 보다 성능이 높음 |

1. **다음은 결정트리의 과대적합을 막는 방법에 대한 설명이다. 바르지 않은 것은?( 3 )**

|  |  |
| --- | --- |
| ① | 트리 생성을 일찍 중단하는 방법은 사전 가지치기 방법이다 |
| ② | 사전 가지치기 방법은 트리의 최대 깊이나 리프의 최대 수를 제한하는 방법이 있음 |
| ③ | Scikit-learn은 사전/사후 가지치기 모두 지원함 |
| ④ | 트리를 만든 후 특정 노드를 삭제하거나 병합하는 방법도 있음 |

1. **다음은 결정트리의 장단점에 관한 설명이다. 바르지 않은 것은? ( 4 )**

|  |  |
| --- | --- |
| ① | max\_depth, max\_leaf\_nodes, min\_samples\_leaf 중 하나만 지정해도 과대적합을 막는 데 충분함 |
| ② | 만들어진 모델을 쉽게 시각화할 수 있어서 비전문가도 이해하기 쉬움 |
| ③ | 결정 트리에서는 feature의 정규화나 표준화 같은 전처리 과정이 필요 없음 |
| ④ | 사전 가지치기를 사용하면 일반화 성능이 좋아짐 |

1. **다음은 커널 서포트 벡터머신(SVM)에 관한 설명이다. 바르지 않은 것은? ( 4 )**

|  |  |
| --- | --- |
| ① | 일반적으로 훈련 데이터의 일부만 결정 경계를 만드는 데 영향을 줌 |
| ② | 위와 같은 데이터 포인트를 서포트 벡터(support vector) 라 함 |
| ③ | 서포트 벡터의 중요도는 학습과정에서 결정되지 않음 |
| ④ | 데이터 포인트 사이의 거리는 가우시안 커널에 의해 계산됨 |

1. **다음은 SVM 매개변수 튜닝에 관한 설명이다. 바르지 않은 것은?( 2 )**

|  |  |
| --- | --- |
| ① | gamma 매개변수는 가우시안 커널 폭의 역수에 해당 |
| ② | 가우시안 커널의 반경이 클수록 훈련 샘플의 영향 범위가 작아짐 |
| ③ | gamma 매개변수가 하나의 훈련 샘플이 미치는 영향의 범위를 결정 |
| ④ | C 매개변수는 선형 모델에서 사용한 것과 비슷한 규제 매개변수 |

1. **다음은 비지도 학습에 관한 설명이다. 바르지 않은 것은?( 1 )**

|  |  |
| --- | --- |
| ① | 비지도 학습은 지도학습과 함께 사용되지 않는다. |
| ② | 데이터의 비지도 변환과 군집은 비지도 학습이다 |
| ③ | 특성의 차원을 축소하는 차원 축소도 비지도 학습이다. |
| ④ | 군집 알고리즘은 데이터를 비슷한 것끼리 그룹으로 묶는 것 |

1. **다음의 데이터 전처리 방법 중 특성 벡터의 길이는 상관 없고, 데이터의 방향만이 중요할 때 주로 사용되는 기법은?( 2 )**

|  |  |
| --- | --- |
| ① | StandardScaler |
| ② | Normalizer |
| ③ | MinMaxScaler |
| ④ | RobustScaler |

1. **다음은 k-평균 군집 알고리즘에 관한 설명이다. 바르지 않은 것은? ( 4 )**

|  |  |
| --- | --- |
| ① | 널리 사용하는 군집 알고리즘 |
| ② | 데이터의 어떤 영역을 대표하는 클러스터 중심을 찾는다 |
| ③ | 클러스터에 할당된 데이터 포인트의 평균으로 클러스터 중심을 다시 지정하는 과정을 반복한다 |
| ④ | 사용자가 지정한 반복횟수를 완료하면 알고리즘이 종료 |

1. **다음은 병합군집 알고리즘에 관한 설명이다. 바르지 않은 것은?( 4 )**

|  |  |
| --- | --- |
| ① | 군집 알고리즘의 모임을 말함 |
| ② | 기본값인 ward 연결은 모든 클러스터 내의 분산을 가장 작게 증가시키는 두 클러스터를 합침 |
| ③ | average 연결은 클러스터 포인트 사이의 평균 거리가 가장 짧은 두 클러스터를 합침 |
| ④ | linkage 옵션에서 가장 비슷한 클러스터를 측정하는 방법을 지정, 이 측정은 항상 k개의 클러스터 사이에서 이루어짐 |

1. **다음은 DBSCAN 알고리즘에 관한 설명이다. 바르지 않은 것은?( 1 )**

|  |  |
| --- | --- |
| ① | 특성 공간의 밀집지역에서 가장 멀리있는 포인트를 찾음 |
| ② | 밀집 지역에 있는 포인트를 핵심 샘플이라 함 |
| ③ | 한 데이터 포인트에서 eps 거리 안에 데이터가 min\_samples 수 만큼 들어 있으면 이 데이터 포인트를 핵심 샘플로 분류 |
| ④ | eps보다 가까운 핵심 샘플은 DBSCAN에 의해 동일한 클러스터로 합쳐짐 |