

학번 : _____ 이름 : _____

1. 텍스트 마이닝에 대한 설명 중 틀린 것을 고르세요. 2

- ① 각 지식 콘텐츠의 특성을 파악해 그 내용 혹은 형태가 유사하거나 상호 관련성이 높은 콘텐츠들을 군집시켜 주는 기술도 포함한다.
- ② 텍스트 문서내에서 중요한 의미를 가지는 정보들을 자동으로 추출해 주는 기술은 문서요약이다.
- ③ 웹페이지, 블로그, 이메일 등 전자문서로 된 텍스트로 부터 유용한 정보를 추출하여 분석하기 위한 도구이다.
- ④ 문서 분류와 정보 추출도 텍스트 마이닝 분야에 속한다.
- ⑤ 문서 군집은 대상 문서의 언어학적 분석을 통해 차별화된 중요 특성들을 추출해 내고, 이를 다른 문서의 특성들과의 비교하여 그 유사도가 높은 문서들을 상호 묶어주는 방식으로 구현한다.

2. 텍스트 데이터를 분석하는 방법에 대한 설명 중 옳은 것을 고르세요. 1

- ① 정규화 작업에서 같은 의미이면서 표현이 다른 단어를 통합한다.
- ② 일반적으로 문법적, 관계적인 뜻을 나타내는 단어 또는 단어의 부분이 형태소이다.
- ③ 하나의 단어가 여러 품사를 갖을 수 있기 때문에, 품사의 모호성 (혹은 중의성)을 제거하는 과정이 토큰화 과정이다.
- ④ 키워드는 주로 불용어 중에서 선택된다.
- ⑤ 보통 문서 내에서 발생 빈도가 낮은 단어들을 키워드로 선정한다.

3. 다음 중 TF-IDF에 대한 설명 중 틀린 것을 고르세요. 3

- ① 주제어 빈도 분석은 특정 문서 집단 내에서 자주 언급되는 주제어를 추출하고 이들이 언급되는 빈도에 따라 중요도를 분석하는 방법이다.
- ② 각 문서 안에 등장하는 용어들의 출현 횟수를 행렬의 형태로 표현한 것을 TDM 이라고 한다.
- ③ 단어의 중요도를 측정하는 방법으로 단어빈도와 문서빈도의 곱으로 표현한 것이 TF-IDF 이다.
- ④ TF-IDF를 사용하는 목적 중 하나는 모든 문서에서 자주 출현하는 상투어를 걸러 내기 위함이다.
- ⑤ 모든 문서에서 자주 등장하는, 즉 그 단어가 흔하게 등장한다는 것을 의미하는 것을 DF로 판단할 수 있다.

4. 기계학습에 대한 설명 중 옳은 것을 고르세요. 2

- ① 입력(문제)-출력(답)의 데이터들로 부터 새로운 입력에 대한 출력을 결정할 수 있는 패턴을 추출하는 것을 강화 학습이라고 한다.
- ② 준지도 학습은 답이 있는 데이터와 답이 없는 데이터를 함께 사용하는 학습 방법이다.
- ③ 출력에 대한 정확한 정보를 제공하지는 않지만, 평가정보는 주어지는 문제에 대해 각 상태에서의 행동을 결정하여 학습하는 방법을 비지도학습 이라고 한다.
- ④ 일반적으로 비지도 학습 방법이 준지도 학습 방법에 비해 난이도가 낮다.
- ⑤ 회귀 (regression)은 비지도 학습 방법이다.

5. 앙상블 학습 방법에 대한 설명 중 틀린 것을 고르세요. 4

- ① 보팅과 배깅은 여러 개의 분류기가 투표를 통해 최종 예측 결과를 결정하는 방식이다.
- ② 대표적인 배깅 방식에는 랜덤 포레스트가 있다.
- ③ 여러 개의 분류기가 순차적으로 학습을 수행하는 방법을 부스팅이라고 한다.
- ④ 모델의 다양성과 관련 있는 것은 배깅 방식이다.
- ⑤ 하드 보팅을 이용한 분류는 다수결 원칙과 비슷한 개념이다.

6. 차원 축소 (Dimension Reduction)에 대한 설명 중 옳은 것을 고르세요. 5

- ① 저차원의 데이터를 정보의 손실을 최소화하면서 고차원으로 변환하는 것을 차원 축소라 한다.
- ② 차원이 늘어나면 해당 공간을 설명하기에 데이터가 부족하기 때문에 underfitting 문제가 발생한다.
- ③ 차원을 축소하는 방법 중 기존의 특성들에서 새로운 특성을 만드는 방법을 특성 선택이라고 한다.
- ④ 차원을 줄일 때는 자료의 특성을 가장 간략하게 표현할 수 있는 방향을 찾는 것이 중요하다.
- ⑤ 차원의 저주 문제를 완화시키기 위해 특성 선택 또는 특성 추출 등의 방법이 필요하다.

7. 배치학습과 온라인 학습에 대한 설명 중 옳은 것을 고르세요. 4

- ① 컴퓨팅 자원이 풍부한 경우 주로 온라인 학습을 수행한다.
- ② 배치 학습 방법은 시간과 자원을 적게 소모한다.
- ③ 온라인 학습은 주로 배치 크기가 크기 때문에 학습 단계가 빠르고 비용이 적게 든다.
- ④ 온라인 학습은 새로운 데이터 샘플을 학습하면, 학습이 끝난 데이터는 더 이상 필요하지 않기 때문에 보관하지 않아도 되므로 저장 공간을 많이 아낄 수 있다.
- ⑤ 배치 학습 방법은 실시간 시스템이나 메모리 부족의 경우 주로 사용한다.

8. 모델이 이전에 알고 있던 내용에 대한 성능이 현저하게 떨어지는 문제를 표현하는 용어로 올바른 것을 고르세요. 1

- ① Catastrophic forgetting
- ② Overfitting
- ③ DropOut
- ④ Online learning
- ⑤ Regularization

9. 인공 신경망의 특징에 대한 설명 중 틀린 것을 고르세요. 5

- ① 잡음 또는 불완전 데이터에 대해 robust한 결과를 보여준다.
- ② 처리 노드가 많기 때문에 몇 개의 노드나 연결에서의 결함이 심각한 결함 초래하지 않는다.
- ③ 입력변수와 예측변수들 간의 복잡한 비선형적인 모델을 추정하는데 유리하다.
- ④ 일반적으로 매우 복잡한 비선형 함수를 사용하므로 비수렴성의 문제가 발생할 수도 있다.
- ⑤ 추정된 가중치의 해석이나 입력변수의 중요성을 직관적으로 판단하기 쉽다.

10. 컨볼루션 신경망 (CNN)에 대한 설명 중 옳은 것을 고르세요. 4

- ① 반복적이고 순차적인 데이터 학습에 특화된 인공신경망의 한 종류이다.
- ② 이미지의 위치, 크기, 각도 변화 등에 매우 취약하다는 단점이 있다.
- ③ 컨볼루션 연산을 수행하여 분류하는 전반부, 특징을 추출하는 후반부로 구성되어 있다.
- ④ 미리 고정된 연결이 없으며, 대신 컨볼루션 과정을 통해 출력을 계산한다.
- ⑤ 특성맵의 수치가 높으면 그 위치가 필터와 유사도가 낮다는 의미이다.

11. 기울기 소멸 문제에 대한 설명 중 틀린 것을 고르세요. 4

- ① 은닉층이 많은 다층 퍼셉트론에서, 출력층에서 아래 층으로 갈 수록 전달되는 오차가 크게 줄어들어, 학습이 되지 않는 현상이다.
- ② 해결하는 방법 중 하나로 시그모이드나 쌍곡 탄젠트 활성화 함수 대신 ReLU (Rectified Linear Unit) 함수를 사용할 수 있다.
- ③ 기울기 소멸 문제를 해결하는 방법 중 하나로 PReLU (parameteric ReLU)나 ELU (exponential Linear Unit) 함수를 사용할 수도 있다.
- ④ 출력층에서 가까울 수록 (미분을 하면 할수록) 학습이 어려운 현상이다.
- ⑤ Vanishing gradient problem이라고 하며, 딥러닝에서 나타나는 문제이다.

12. 추천시스템의 Collaborative Filtering에 관한 설명 중 옳은 것을 고르세요. 5

- ① Collaborative Filtering은 사용자가 관심 분야에 대해서 직접 입력한 정보나, 구매 내역, 평점 등을 기반으로 추천하는 방법이다.
- ② Collaborative Filtering의 모델 기반 접근 방법은 실제 데이터에 적용했을 때 성능이 떨어진다.
- ③ 아무도 구매를 하지 않아서 평가가 없거나 신규 아이템으로 평가가 없는 경우 유사도를 구할 수 없기 때문에 누군가 평가를 할 때까지 추천을 할수 없어 없는 상황을 Cold Start Problem이라고 한다.
- ④ 일관성이 없는 의견을 가진 사용자들의 데이터는 추천에 혼란을 줄 수 있으며, Shilling Attack 이라고 한다.
- ⑤ 한 사용자에 대한 충분한 데이터가 부족한 경우 선호도 예측 불가능 한 문제를 Cold Start Problem이라고 한다.

13. 추천 시스템에 관한 설명 중 틀린 것을 고르세요. 2

- ① 추천 시스템은 아이템에 대한 사용자의 선호도를 예측하는 시스템이다.
- ② 유저가 새로운 유형/견해의 콘텐츠에 노출되지 않고 특정 유형/견해를 갖는 콘텐츠에 고립되는 현상을 Grey Sheep Problem 이라고 한다.
- ③ 다양한 콘텐츠를 소비하는 사용자일수록 서서히 이탈율이 낮은 경향이 있다.
- ④ 과거 행동에서 유의한 순차적인 패턴을 찾고, 최근 아이템에 더 집중하여 이를 기반으로 추천하는 방법을 Sequential Recommendation 이라고 한다.
- ⑤ CTR을 사용하면 키워드 및 광고와 무료 등록정보의 실적을 파악 가능하다.

14. 시계열 데이터 처리가 가능하며, 번역, 자연어, 음성 처리에서 우수한 성능을 보이는 “encoder-decoder”로 구성된 모델은 무엇일까요? 1

- ① Transformer
- ② CNN
- ③ GAN
- ④ LSTM
- ⑤ RNN

15. 컴퓨터 비전에 관한 응용 분야가 아닌 것은? 3

- ① 자율주행
- ② 동작인식
- ③ Chatbot
- ④ OCR
- ⑤ 얼굴인식

16. 컴퓨터 비전에 관한 설명 중 옳은 것을 고르세요. 2

- ① 객체 검출은 비디오 영상 내의 객체의 위치 변화를 추적하는 기술이다.
- ② 객체 분할 방법 중 multi object가 존재하는 방식이 instance segmentation 이다.
- ③ 이미지의 상황을 텍스트로 설명할 수 있는 기술을 전처리라고 한다.
- ④ 이미지 또는 비디오 영상에서 객체를 식별해내는 기술을 객체 추적 이라고 한다.
- ⑤ 객체 분류에서 성능을 결정짓는 가장 중요한 요소는 True case의 데이터를 확보하는 데 있다.

17. 정보 보안의 원칙에 관한 설명 중 옳은 것을 고르세요. 4

- ① 상호 신뢰할 수 있는 합법적인 통신 상대 끼리만 시스템에 접근할 수 있어야 한다는 원칙이 무결성이다.
- ② 정보가 고의적이거나 비인가된 방법으로 변경되는 것으로부터 보호되어야 한다는 원칙이 기밀성이다.
- ③ 정보, 정보 시스템, 정보 보안 시스템 등에 사용자가 필요로 하는 시점에 접근할 수 있어야 한다는 원칙이 익명성이다.
- ④ 시스템을 마비시키는 DDoS 공격은 가용성을 해치는 공격이다.
- ⑤ 기밀성을 제공하는 방법으로는 해시 함수, 디지털 서명, 암호 알고리즘 등이 있다.

18. 순환 신경망의 특징 중 옳은 것을 고르세요. 1

- ① 낱말의 등장 빈도 등과 같은 통계적 정보를 활용하기 어렵다.
- ② 문장에서 낱말의 위치에 따라 다르게 처리할 수는 있지만 해당 낱말들 간의 관계를 파악하기 쉽다.
- ③ 비교적 단순한 구조이지만 신경망을 구현하는데 있어 구축하는데 어려움이 있다.
- ④ 이전의 처리된 데이터를 반영하지 않고 현재의 데이터만을 처리한다.
- ⑤ 인간의 장기기억력과 연관된 신경망이다.

19. 과적합에 대한 설명 중 틀린 것을 고르세요. 1

- ① 해결 방법으로 파라미터가 더 많은 강력한 모델을 선택할 수 있다.
- ② 규제화를 통해 완화 가능하다.
- ③ 드롭아웃 (Drop Out)을 통해 성능 향상 및 과적합 완화가 가능하다.
- ④ 일반적으로 유의수준 α 를 미리 정해 놓고 β 를 최대화하는 검정의 방법을 사용한다.
- ⑤ 데이터의 차원이 늘어나면 과적합 문제가 발생할 수 있다.

20. 다음 중 옳은 것을 고르세요. 5

- ① 데이터의 다양성이 작을수록 더 복잡한 모델을 사용하면 좋은 성능을 얻을 수 있다.
- ② 일반적으로 더 큰 데이터셋 일수록 다양성이 높기 때문에 더 간단한 모델을 사용 가능하다.
- ③ 훈련 데이터셋은 검증데이터셋 없이 최대한 많이 학습하는 것이 성능 향상에 좋다.
- ④ 드롭아웃 (Drop Out)은 데이터의 양을 늘리기 위해 원본에 각종 변환을 적용하여 개수를 증강 방법이다.
- ⑤ 같은 데이터를 중복하거나 비슷한 데이터를 모으는 것은 다양성 증가에 도움이 되지 않는다.

21. 보안을 위협하는 악성 프로그램에 대한 설명 중 옳은 것을 고르세요. 3

- ① 자기 복제를 하며, 컴퓨터 시스템을 파괴하거나 작업을 지연 및 방해하는 것을 랜섬웨어라고 한다.
- ② 트로이목마는 실행 코드 자체로 번식하며, 주로 PC에서 실행된다.
- ③ 트로이목마는 유틸리티 프로그램에 악의적인 코드를 내장하거나 그 자체를 유틸리티 프로그램으로 위장하여 접근한다.
- ④ 시스템 설계자가 서비스 기술자의 접근 편의를 위해 일부러 만들어 놓은 시스템의 보안 구멍을 스파이웨어 라고 한다.
- ⑤ 다른 사람의 컴퓨터에 잠입하여 개인정보를 추적, 모니터 및 소유하며, 제3자에게 유출시키는 프로그램이 백도어이다.

22. 준지도학습에 대한 설명 중 틀린 것을 고르세요. 4

- ① 우리 주변에는 레이블이 없는 데이터가 대부분이며, 이를 사용해 분류기의 성능을 향상시키는 것을 목표로 한다.
- ② 모델 성능을 높이기 위해 정답이 있는 데이터와 정답이 없는 데이터를 함께 사용한다.
- ③ 특정 경우에는 준지도 학습이 오히려 성능이 저하시킬 수 있다.
- ④ Self-Training 방법은 정답이 있는 데이터로 모델을 학습 후 정답이 없는 데이터의 라벨을 예측 후 정답이 있는 데이터는 빼고 학습한다.
- ⑤ 일반적으로 지도 학습과 비교하여 정확도가 낮다.

23. 애플리케이션 보안에 대한 설명 중 옳은 것을 고르세요. 4

- ① SSO는 비밀번호 분실이나 취약한 비밀번호의 위험도를 높인다.
- ② 바이러스 백신은 바이러스의 감염을 사전에 막을 수 있다.
- ③ 허가된 사용자만 접근할 수 있도록 만드는 제한 기술이 디지털 포렌식 이다.
- ④ 지정된 플랫폼 하나에서 인증되면 이후 매번 로그인/아웃을 반복할 필요없이 다양한 서비스를 사용 가능한 것이 SSO 이다.
- ⑤ PGP는 무결성, 인증, 부인방지를 제공하지 않기 위해 디지털 서명을 사용한다.

24. 개인화 추천 엔진이 유저의 과거 이력에 기반한 추천 결과를 내보내면서 유저가 새로운 유형/견해의 콘텐츠에 노출되지 않고 특정 유형/견해를 갖는 콘텐츠에 고립되는 현상을 나타내는 용어를 고르세요. 5

- ① Truecolor
- ② Underfitting
- ③ 특징 추출
- ④ calibration
- ⑤ 필터 버블 이슈

25. 오차 역전파 (Back propagation)에 대한 설명 중 틀린 것을 고르세요. 5

- ① 역전파 학습은 미분의 “Chain Rule”을 따른다.
- ② 가중치들은 선형대수에서 배우는 행렬 형태를 따른다.
- ③ 경사 하강법을 이용해 바로 앞 가중치를 오차가 작아지는 방향으로 업데이트 한다.
- ④ Multi Layer Perceptron (MLP) 학습을 위해 오차 역전파가 사용된다.
- ⑤ 단일 퍼셉트론에서 반드시 필요한 알고리즘 이다.

26. 장단기 기억망에 대한 구성 중 과거 정보를 얼마나 유지할 지를 결정하는 모듈의 명칭을 고르세요.
1

- ① forget gate
- ② XOR gate
- ③ output gate
- ④ input gate
- ⑤ convolution gate

27. 어텐션 (Attention)에 대한 설명 중 틀린 것을 고르세요. 5

- ① 전체 입력 문장을 전부 다 동일한 비율로 참고하는 것이 아니라 , 해당 시점에서 예측해야 할 단어와 연관이 있는 입력 단어 부분을 좀 더 집중하는 개념이다.
- ② Transformer에서 사용하는 기법이다.
- ③ 디코더에서 출력 단어를 예측하는 매 시점 (time 마다 , 인코더에서의 전체 입력 문장을 다시 한번 참고하는 아이디어 이다.
- ④ Seq2Seq 모델의 단점을 극복할 수 있다.
- ⑤ 2020년도에 소개된 개념으로 encoder-decoder 로 구성된다.

28. 언어 표현을 사전 학습시키는 방법으로 Transformer의 인코더 부분을 발전시킨 모델을 고르세요. 2

- ① GPT
- ② BERT
- ③ CNN
- ④ CTRL
- ⑤ ChatGPT

29. 강화학습에 대한 설명 중 옳은 것을 고르세요. 3

- ① 준지도 학습 방법 중 하나이다.
- ② 지도학습이 경험을 통해서 학습을 하는 것이라면 강화학습은 배움을 통해서 학습한다.
- ③ 게임에서의 강화학습은 게이머의 판단력이 정책 (policy)에 해당된다.
- ④ 행동의 결과가 자신에게 유리한 것이었다면 벌을 받고 , 불리한 것이었다면 재학습한다.
- ⑤ 에이전트가 환경을 관찰하고, 행동을 실행 후 피드백을 통해 보상이 줄어드는 방향으로 학습한다.

30. 생체 인식에 관한 설명 중 틀 것을 고르세요. 5

- ① 각 개인이 갖고 있는 특징을 인식하여 보안을 유지하는 기술이다.
- ② 신체적 특성으로는 지문, 홍채, 얼굴, 정맥 등이 있으며 행동적 특성으로는 목소리, 서명 등이 있다.
- ③ 지문인식은 현재 가장 대중적인 방법 중 하나이다.
- ④ 사물인터넷(IoT)과 모바일, 웨어러블 디바이스와 융합된 제품과 서비스로 확대될 것으로 전망이다.
- ⑤ 정맥인식은 원격지에서도 전화를 이용하여 신분 확인이 가능하다.