1. <보기>의 서버 상태는 정상, 주의, 장애의 세 가지로 구분되고, 서버 상태의 변화는 동질 마르코프 체인(homogeneous Markov chain)을 따른다. 전이 확률 행렬(transition probability matrix)이 <보기>와 같을 때, 현재 서버가 정상 상태라면 정확히 2시간 뒤 시점에도 정상 상태일 확률로 가장 옳은 것은? (단, 상태 전이는 1시간마다 한 번씩 일어난다.)

<보기>

현재 상태 \ 다음 상태	정상	주의	장애
정상	0.7	0.2	0.1
주의	0.3	0.5	0.2
장애	0.1	0.2	0.7

① 0.49 ② 0.52
③ 0.56 ④ 0.71

2. L1 정칙화(regularization)와 L2 정칙화에 대한 설명으로 가장 옳은 것은?
① L1 정칙화에 사용하는 벌점(penalty)은 가중치 성분이 0인 점에서 미분 가능하며, 최적해를 항상 유일하게 만든다.
② L1 정칙화는 일부 가중치를 정확히 0으로 만들 수 있지만, L2 정칙화는 일반적으로 가중치의 크기를 줄이되 정확히 0으로 만들지는 않는다.
③ L1 정칙화는 가중치 제곱합을 사용하고, L2 정칙화는 가중치 절댓값의 합을 사용한다.
④ 정칙화 강도 λ가 커질수록 두 방법에서 정칙화의 영향은 약해진다.

3. 연소득[원]과 월 방문 횟수[회]를 유클리드 거리로 사용하는 k-최근접 이웃(k-nearest neighbors) 분류기에서, 두 특성의 척도 차이를 줄이기 위해 표준화를 적용할 때 가장 옳은 것은?
① 학습 자료에서 추정한 특성 스케일링 변환을 검증·시험 자료에 동일하게 적용한다.
② 학습 자료와 검증·시험 자료를 모두 합친 후, 각 특성의 평균과 표준편차로 표준화한다.
③ 학습 자료와 시험 자료를 각각 독립적인 평균과 표준편차로 표준화한다.
④ k-최근접 이웃 분류 방법은 학습 가중치가 없으므로 특성 스케일을 고려할 필요가 없다.

4. KL발산(Kullback-Leibler divergence)에 대한 설명으로 가장 옳지 않은 것은?
① 두 확률분포 사이의 차이를 측정하는 데 사용된다.
② 항상 0 이상이다.
③ 대칭성을 만족한다.($D_{KL}(P \| Q) = D_{KL}(Q \| P)$)
④ 두 분포가 같으면 0이다.

5. 최대 가능도 추정(maximum likelihood estimation, MLE), 최대 사후확률 추정(maximum a posteriori, MAP) 및 베이지안 예측(Bayesian prediction)을 비교한 설명으로 가장 옳은 것은?
① MAP는 가능도(likelihood)와 사전분포를 결합한 사후확률의 최댓값을 사용한다.
② MLE는 파라미터의 사후분포 전체에 대해 예측을 적분한다.
③ 베이지안 예측은 일반적으로 사전분포에 대하여 예측을 평균한다.
④ 베이지안 예측은 사전분포를 사용하므로 하이퍼파라미터(hyperparameter)가 존재하지 않는다.

6. 커널 밀도 추정(kernel density estimation)에 대한 설명으로 가장 옳은 것은?
① 커널의 대역폭이 너무 작으면 추정 분포가 뾰족해지고, 너무 크면 과도하게 평탄화된다.
② 커널의 대역폭이 0에 가까울수록 실제 분포를 정확하게 추정한다.
③ 커널 밀도 추정은 히스토그램보다 계단 모양이 심하다.
④ 고차원일수록 데이터 희소성 문제가 사라져 더 효과적이다.

7. 1차원 점 {0.00, 0.10, 0.20, 0.50, 0.60}에 DBSCAN(density-based spatial clustering of applications with noise)을 적용한다. ε=0.15이고 MinPts=3이며, ε-이웃에는 해당 점 자신과 그 점으로부터의 거리가 ε 이하인 점을 포함한다. 핵심점(core point), 경계점(border point), 잡음점(noise point)을 구분한 것으로 가장 옳은 것은?

	핵심점	경계점	잡음점
①	{0.00, 0.10, 0.20}	없음	{0.50, 0.60}
②	{0.10}	{0.00, 0.20}	{0.50, 0.60}
③	{0.10, 0.50, 0.60}	{0.00, 0.20}	없음
④	없음	{0.00, 0.10, 0.20}	{0.50, 0.60}

8. 머신러닝 및 딥러닝의 하이퍼파라미터(hyperparameter) 최적화 기법에 대한 설명으로 가장 옳지 않은 것은?

① 하이퍼파라미터 최적화의 목적은 하이퍼파라미터의 최적 조합을 찾아 모델의 일반화 성능을 극대화하는 것이다.

② 베이지안 최적화(Bayesian optimization)는 대리 모델(surrogate model)을 사용하여 이전 시도의 결과를 토대로 탐색 공간을 설정한 뒤, 무작위 탐색과정을 거쳐 최적값을 찾아낸다.

③ 랜덤 서치(random search)는 탐색 범위 내에서 하이퍼파라미터를 무작위로 추출하여 제한된 시도 횟수 내에서 효율적으로 탐색한다.

④ 그리드 서치(grid search)는 지정한 후보 값들의 모든 조합을 전수 조사하므로, 탐색 공간이 커지면 연산 시간이 급격히 증가한다.

9. 선형 회귀 모델의 예측값과 손실함수가 <보기 1>과 같이 정의되어 있고, 3개의 학습 데이터는 <보기 2>와 같다. 현재 파라미터에서 손실함수의 그래디언트(gradient) $(\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b})$의 값으로 가장 옳은 것은? (단, 가중치 $w=2$, 편향 $b=-1$이다.)

<보기 1>

손실함수

$$L(w,b)=\frac{1}{n}\sum_{i=1}^{n}((wx_i+b)-y_i)^2$$

<보기 2>

학습 데이터($n=3$)

i	x_i	y_i
1	1	1
2	2	5
3	3	7

① $(\frac{8}{3}, \frac{20}{3})$ ② $(-\frac{8}{3}, -\frac{20}{3})$

③ $(\frac{20}{3}, \frac{8}{3})$ ④ $(-\frac{20}{3}, -\frac{8}{3})$

10. <보기>는 3개 클래스에 대한 단일 레이블 분류의 혼동 행렬(confusion matrix)이다. 거시 평균 F1 점수(macro-averaged F1)와 미시 평균 F1 점수(micro-averaged F1)를 차례로 짝지은 것으로 가장 옳은 것은? (단, 각 클래스의 F1은 해당 클래스를 양성 클래스로 나머지 클래스를 음성 클래스로 보아 계산하고, 행은 실제 클래스, 열은 예측 클래스이다.)

<보기>

	예측 A	예측 B	예측 C
실제 A	4	0	0
실제 B	0	1	1
실제 C	0	1	1

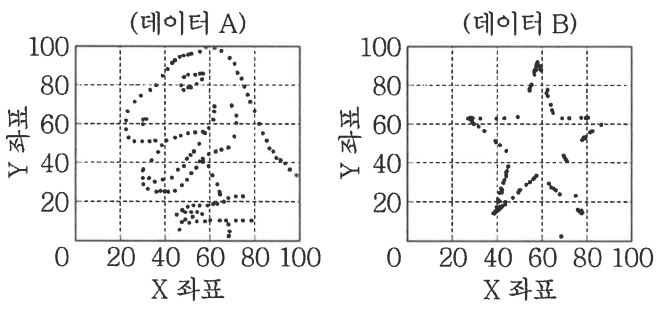
* 거시 평균 F1: 클래스별 F1의 산술평균
* 미시 평균 F1: 클래스별 TP, FP, FN을 각각 합산한 뒤 계산한 F1

① 1/2, 3/4 ② 2/3, 2/3

③ 2/3, 3/4 ④ 3/4, 2/3

11. <보기>는 데이터 세트 A와 B의 산점도(scatter plot) 및 두 데이터 세트의 동일한 요약 통계량(X와 Y의 평균(mean), X와 Y의 표준편차(SD), 피어슨 상관계수(Corr.))을 나타낸다. 데이터 시각화와 요약 통계량에 관한 설명으로 가장 옳지 않은 것은?

<보기>



X mean: 54.26 Y mean: 47.83

X SD: 16.76 Y SD: 26.93

Corr. : −0.06

① 산점도 그래프와 같은 시각화 기법은 단순한 수치 데이터(요약 통계량)보다 데이터의 실제 공간적 분포와 형상을 훨씬 더 직관적으로 확인할 수 있게 해준다.

② 피어슨 상관계수는 공룡이나 별 모양과 같은 기하학적 형태의 종속성(dependency)까지 반영하여 측정하는 종합 지표이다.

③ 피어슨 상관계수가 −0.06이면 두 변수 사이에 직선 형태의 양 또는 음의 상관관계가 거의 없음을 나타낸다.

④ 데이터 분석에서 두 변수 X와 Y 사이에 높은 상관계수나 통계적 상관관계가 관찰되더라도, 이는 한 변수의 변화가 다른 변수를 직접적으로 변화시켰다는 인과관계(causality)를 입증하는 근거로 사용될 수 없다.

12. <보기>의 ㄱ~ㄹ 중 설명 가능한 인공지능(explainable artificial intelligence) 기술로 사용되는 것의 개수로 가장 옳은 것은?

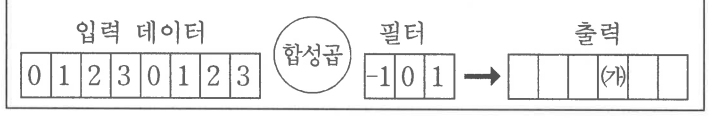
<보기>

ㄱ. LIME(local interpretable model-agnostic explanations)
ㄴ. SHAP(Shapley additive explanations)
ㄷ. 의사결정 트리 대리 모델(decision tree surrogate)
ㄹ. Grad-CAM(gradient-weighted class activation map)

① 1개 ② 2개
③ 3개 ④ 4개

13. <보기>의 1차원 합성곱 신경망(convolutional neural network)에서 입력 데이터와 필터(filter)를 교차상관(cross-correlation) 방식으로 합성곱(convolution)할 때, ㈎ 부분에 출력되는 값으로 가장 옳은 것은? (단, 합성곱 연산에서 보폭(stride)은 1이고, 패딩(padding)은 사용하지 않으며, 편향(bias)은 무시한다.)

<보기>



① −2 ② −1
③ 0 ④ 1

14. <보기> 중 강화학습에서 정책(policy)과 가치 함수(value function)에 대한 설명으로 옳은 것을 모두 고른 것은? (단, S는 모든 비종단 상태의 집합, t는 이산적 시간 간격을 의미한다.)

<보기>

ㄱ. 가치 함수 $v_\pi(s)$는 특정 정책의 상태 s에서 정책 π를 따른 이후 기대되는 누적보상(기대 이득(return))의 기댓값이다.
ㄴ. 최적의 정책 $\hat{\pi}$는 모든 상태에서 가치 함수를 최소로 만드는 정책이다.($\hat{\pi} = \arg\min_\pi v_\pi(s), \forall s \in S$)
ㄷ. 할인율(γ)이 적용된 누적보상액은 $r = \sum_k \gamma^{k-1} r_{t+k}$이며, 이때 $\gamma > 1$이다.

① ㄱ ② ㄱ, ㄴ
③ ㄱ, ㄷ ④ ㄴ, ㄷ

15. <보기> 중 자연어 처리 모델에 대한 설명으로 옳은 것을 모두 고른 것은?

<보기>

ㄱ. Word2Vec의 Skip-gram은 중심 단어로 주변 단어를 예측한다.
ㄴ. 문장 "the cat sat on the mat"에서 bigram(2-gram)을 만들 때 생성되는 bigram의 총 개수는 5개이다. (단, 토큰은 공백 단위로 나눈다.)
ㄷ. Transformer는 순환 구조 없이 셀프 어텐션(self-attention)으로 문맥을 처리한다.
ㄹ. BERT는 양방향 문맥을, GPT는 좌→우 단방향 문맥을 주로 학습한다.

① ㄱ, ㄴ ② ㄷ, ㄹ
③ ㄱ, ㄴ, ㄷ ④ ㄱ, ㄴ, ㄷ, ㄹ

16. <보기>의 ㈎에 들어갈 용어로 가장 옳은 것은?

<보기>

[㈎]은(는) 딥러닝 모델의 예측 결과를 설명해 주는 방법 중 하나로, 입력 이미지의 일부를 가렸을 때, 모델의 예측 결과가 얼마나 변하는지 측정하여 모델이 예측에서 가장 중요하게 여기는 영역이 어디인지 찾아내는 기법이다.

① CAM(class activation map)
② Grad-CAM(gradient-weighted class activation map)
③ Occlusion sensitivity
④ tf-explain

17. 트랜스포머(transformer)에서 <보기>의 스케일드 닷-프로덕트 어텐션(scaled dot-product attention)에 대한 설명으로 가장 옳지 않은 것은? (단, Q, K, V는 각각 query, key, value 행렬(또는 벡터)이며, d_k는 key의 차원 수이다.)

<보기>

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V$$

① Q와 K의 내적으로 쿼리-키 유사도를 계산한다.
② $\sqrt{d_k}$로 나누는 것은 내적 값이 커져 소프트맥스 그래디언트가 포화되는 것을 완화한다.
③ 소프트맥스 출력은 V에 대한 가중 평균의 가중치로 쓰인다.
④ d_k가 커질수록 소프트맥스 함수의 기울기가 줄어들어 스케일링 없이도 안정적으로 학습할 수 있다.

18. 생성형 인공지능에서 사용되는 생성 모델에 대한 설명으로 가장 옳은 것은?

① 변이형 자기 부호화기(variational autoencoder, VAE)는 생성자와 판별자를 서로 경쟁시키는 적대적 학습을 사용한다.

② 자기회귀 모델(autoregressive model)은 데이터의 결합확률을 조건부 확률의 곱으로 분해한다.

③ 확산 모델(diffusion model)의 순방향 과정은 잡음으로부터 데이터를 복원하며, 역방향 과정은 실제 데이터에 잡음을 점진적으로 추가한다.

④ 생성적 적대 신경망(generative adversarial network, GAN)은 데이터의 확률밀도 $p(x)$를 반드시 명시적으로 계산하고, 그 우도를 직접 최대화하여 학습한다.

19. 지식 증류(knowledge distillation)에 대한 설명으로 가장 옳은 것은?

① 학생 모델은 교사 모델의 가중치를 동일한 위치에 그대로 복사해야 하므로, 두 모델은 반드시 같은 구조를 가져야 한다.

② 지식 증류가 완료된 후에도 예측을 수행할 때에는 교사 모델과 학생 모델을 항상 함께 실행해야 한다.

③ 학생 모델은 실제 정답뿐 아니라 교사 모델이 출력한 클래스별 확률분포를 학습에 활용하여, 교사 모델이 학습한 정보를 전달받을 수 있다.

④ 학습이 완료된 모델에서 값이 작은 가중치를 제거하여 모델의 연산량을 줄이는 방법이다.

20. 학습된 인공지능 모델을 모바일/임베디드 디바이스에 배포하기 위한 최적화 및 경량화 기술에 대한 설명으로 가장 옳지 않은 것은?

① 컴퓨팅 자원이 제한된 모바일/임베디드 장치에서 정확도를 크게 낮추지 않으면서도 효율적인 경량 모델 배포를 지원하는 기술이다.

② 가중치의 수치 정밀도를 기존 32비트 부동소수점(FP32)에서 16비트 부동소수점(FP16)으로 낮추는 반정밀도(half precision)를 사용하면 모델 예측 정확도의 손실을 최소화하면서 모델 용량을 약 1/2 수준으로 줄일 수 있다.

③ 양자화(quantization) 기법을 사용해 FP32 가중치를 INT8 정수로 변환하여 모델 용량을 약 1/4 수준으로 대폭 줄일 수 있다.

④ 모바일 디바이스용 연산 정밀도 축소 및 양자화 기술의 궁극적인 목적은 모바일 기기 내부에서 수행되는 역전파(backpropagation) 기반의 학습(training) 연산을 가속화하기 위함이다.