

Họ tên thí sinh:

Số báo danh:

Mã Đề: 006.

Lưu ý:

- Các số thực được hiển thị với dấu . phân tách phần nguyên và phần thực.
- Trong tất cả các câu hỏi, thí sinh chỉ cần chọn một đáp án duy nhất và chọn đáp án sai không bị trừ điểm.

Câu 1. Trong mạng nơ-ron, chuẩn hóa lô (Batch Normalization) thường được đặt ở đâu?

- A. Trước hàm kích hoạt ReLU và sau lớp tuyến tính (Linear)
- B. Sau ReLU
- C. Sau lớp bỏ ngẫu nhiên (dropout)
- D. Trước lớp đầu vào (input layer)

Câu 2. Thành phần nào sau đây là một tham số có thể học trong một lớp tích chập ?

- A. Kích thước của dữ liệu đầu vào
- B. Các giá trị trong bộ lọc
- C. Kích thước bước nhảy
- D. Kích thước padding

Câu 3. Nếu mất mát (loss) không giảm sau 5 vòng lặp (epoch) đầu tiên dù tỷ lệ học là 0.001, bạn nên làm gì đầu tiên?

- A. Dừng lại và chọn mô hình khác
- B. Kiểm tra lại quy trình dữ liệu (data pipeline), tăng cường dữ liệu (augmentation) và bộ tối ưu
- C. Tăng tỷ lệ học lên 0.1
- D. Tăng số lớp của mô hình

Câu 4. Hạn chế lớn nhất khi tinh chỉnh (fine-tune) toàn bộ mô hình GPT là gì?

- A. Không dùng được tiếng Việt
- B. Cần tài nguyên tính toán rất lớn
- C. Không dùng được API
- D. Không sinh được ảnh

Câu 5. Khi muốn trích xuất đặc trưng (features) từ ResNet50, bạn nên làm gì?

- A. Sử dụng bộ tối ưu Adam
- B. Sử dụng đầu ra từ lớp gần cuối (ví dụ: avgpool, penultimate, ...)
- C. Thêm nhiều lớp bỏ ngẫu nhiên (dropout)
- D. Thêm lớp kết nối đầy đủ siêu tốc (fully connected) mới

Câu 6. Giả sử bạn đang xây dựng một mô hình phân loại cảm xúc văn bản (positive/negative) bằng cách sử dụng biểu diễn Bag-of-Words (BoW) và PyTorch (phiên bản ≥ 1.6). Dưới đây là đoạn mã tiền xử lý đã được thực hiện:

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
import torch
import torch.nn as nn
import torch.optim as optim

texts = ["I love this movie", "I hate this product",
         "Amazing quality", "Terrible service"]
labels = [1, 0, 1, 0]

vectorizer = CountVectorizer()
X = vectorizer.fit_transform(texts).toarray()
y = torch.tensor(labels)

X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.25)

X_train = torch.tensor(X_train, dtype=torch.float32)
```

Phương án nào sau đây là phần mã đúng để huấn luyện mô hình phân loại nhị phân đơn giản với biểu diễn BoW, một tầng tuyến tính và hàm loss phù hợp.

A.

```
model = nn.Linear(X_train.shape[1], 1)
loss_fn = nn.BCEWithLogitsLoss()
optimizer = optim.Adam(model.parameters(), lr=0.01)
```

```
for epoch in range(100):
    outputs = model(X_train).squeeze()
    loss = loss_fn(outputs, y_train.float())
    loss.backward()
    optimizer.step()
    optimizer.zero_grad()
```

B.

```
model = nn.Sequential(
    nn.Linear(X_train.shape[1], 10),
    nn.ReLU(),
    nn.Linear(10, 2)
)
loss_fn = nn.NLLLoss()
optimizer = optim.Adam(model.parameters(), lr=0.01)
```

```
for epoch in range(100):
    outputs = model(X_train)
    loss = loss_fn(outputs, y_train)
    loss.backward()
    optimizer.step()
    optimizer.zero_grad()
```

C.

```
model = nn.Linear(X_train.shape[1], 2)
loss_fn = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=0.01)
```

```
for epoch in range(100):
    outputs = model(X_train)
    loss = loss_fn(outputs, y_train)
    loss.backward()
    optimizer.step()
    optimizer.zero_grad()
```

D.

```
model = nn.Linear(X_train.shape[1], 1)
loss_fn = nn.MSELoss()
optimizer = optim.Adam(model.parameters(), lr=0.01)
```

```
for epoch in range(100):
    outputs = model(X_train)
    loss = loss_fn(outputs, y_train)
    loss.backward()
    optimizer.step()
    optimizer.zero_grad()
```

Câu 7. Điều nào sau đây là ĐÚNG khi so sánh SVM (support vector machine) với k-NN?

- A. Cả hai phương pháp chỉ được sử dụng cho các bài toán phân loại (không phải hồi quy)
- B. Huấn luyện SVM có thể tốn kém về mặt tính toán, đặc biệt đối với các bộ dữ liệu lớn, trong khi huấn luyện k-NN không liên quan đến quy trình huấn luyện rõ ràng
- C. SVM luôn tốt hơn k-NN trên mọi tập dữ liệu
- D. Dự đoán của cả hai phương pháp đều chậm vì cả hai đều cần xử lý tất cả các mẫu dữ liệu huấn luyện để dự đoán mẫu dữ liệu mới

Câu 8. Đối với SVM (support vector machine) phi tuyến, để dự đoán, đáp án nào đúng.

- A. Cần sử dụng cùng hàm chuyển đổi tương tự với giai đoạn huấn luyện
- B. Sử dụng một hàm chuyển đổi khác với giai đoạn huấn luyện
- C. SVM phi tuyến luôn tốt hơn SVM tuyến tính trong mọi tập dữ liệu
- D. Không cần hàm chuyển đổi

Câu 9. Khi sử dụng `torchvision.models.resnet18(pretrained=True)` trong PyTorch, mục đích chính là gì?

- A. Huấn luyện mô hình từ đầu
- B. Tăng kích thước lô (batch size)
- C. Thử nghiệm mô hình mới
- D. Sử dụng trọng số huấn luyện trước (pre-trained weights) trên ImageNet để trích xuất đặc trưng

Câu 10. Khi các mô hình phát hiện vật thể hoạt động, chúng thường đề xuất nhiều hộp bao (bounding box) cho cùng một đối tượng trong ảnh. Nhiều hộp bao này có thể chồng lấn lên nhau, dẫn đến thông tin dư thừa vì mục tiêu của chúng ta là chỉ xác định một hộp bao duy nhất và chính xác nhất cho mỗi đối tượng. Thuật toán Non-Maximum Suppression (NMS) được thiết kế để giải quyết vấn đề này. Nó giúp lọc và loại bỏ các hộp bao dư thừa, chỉ giữ lại những hộp bao đại diện tốt nhất cho mỗi đối tượng.

Thuật toán NMS gồm các bước như sau:

- Sắp xếp: Sắp xếp tất cả các hộp bao trong tập P theo thứ tự giảm dần của điểm tin cậy (confidence score).
- Chọn hộp tốt nhất: Chọn hộp bao H có điểm tin cậy cao nhất từ tập P. Hộp bao này được xem là đại diện tốt nhất hiện tại.
- Giữ lại và loại bỏ: Chuyển hộp bao H vào danh sách kết quả cuối cùng (gọi là K) và loại bỏ nó khỏi tập P.
- So sánh và loại bỏ chồng lấn:
 - + Tính toán chỉ số Intersection over Union (IoU) giữa hộp H (vừa chọn) và tất cả các hộp bao còn lại trong tập P. Chỉ số IoU được tính theo công thức sau:

$$\text{IoU} = \frac{\text{Diện tích vùng giao của 2 hộp}}{\text{Diện tích vùng hợp của hai hộp}}.$$

+ Đối với mỗi hộp bao còn lại trong P, nếu giá trị IoU của nó với hộp H lớn hơn một ngưỡng xác định trước (IoU_threshold), thì loại bỏ hộp bao đó khỏi P. Lý do là vì nó chồng lấn quá nhiều với hộp H (vốn có điểm tin cậy cao hơn) và được coi là dư thừa cho cùng một đối tượng.

- Lặp lại: Quay lại Bước 2 và lặp lại quy trình (chọn hộp có điểm tin cậy cao nhất tiếp theo từ P, thêm vào K, loại bỏ các hộp chồng lấn khỏi P) cho đến khi tập P không còn hộp bao nào.
- Kết quả: Danh sách K sẽ chứa các hộp bao cuối cùng được giữ lại, mỗi hộp đại diện cho một đối tượng riêng biệt đã được phát hiện.

Áp dụng thuật toán NMS với ngưỡng IoU-threshold = 0.40 và giả sử tất cả các hộp cùng một lớp và thông tin các hộp đã sắp thứ tự theo độ tin cậy (confidence score) như sau:

ID	Toạ độ (x_1, y_1, x_2, y_2) (pixel)	Độ tin cậy
B_1	$(0, 0, 100, 100)$	0.95
B_2	$(10, 10, 90, 90)$	0.90
B_3	$(105, 105, 200, 200)$	0.85

Những hộp nào sẽ được giữ lại?

- A. B_1 và B_2
- B. Chỉ B_1
- C. B_1 và B_3
- D. B_2 và B_3

Câu 11. Trong Pandas, phương thức nào dùng để lấy 5 dòng đầu tiên của DataFrame?

- A. `df.head()`
- B. `df.take(5)`
- C. `df.top()`
- D. `df.first(5)`

Câu 12. Thuật toán học máy nào sau đây là phô biến và hiệu quả, dựa trên ý tưởng bagging?

- A. XGBoost
- B. Hồi quy tuyến tính (Linear Regression)
- C. Cây quyết định (Decision Tree)
- D. Rừng ngẫu nhiên (Random Forest)

Câu 13. Kỹ thuật nào sau đây được sử dụng để giảm ảnh hưởng của nhiễu và ngoại lệ trong tập dữ liệu?

- A. Phân tích thành phần chính (Principal Component Analysis - PCA)
- B. Chính quy hóa (Regularization)
- C. Xác thực chéo (Cross-validation)
- D. Trích xuất đặc trưng (Feature extraction)

Câu 14. Trong xử lý ngôn ngữ tự nhiên, đâu là thứ tự đúng của các bước xử lý cơ bản sau đây?

- 1. Tách từ (Tokenization)

- 2. Chuẩn hóa văn bản (Normalization)
 - 3. Rút gọn từ (Stemming)
 - 4. Gán nhãn từ loại (Part-of-speech tagging)
- A. $2 \rightarrow 1 \rightarrow 4 \rightarrow 3$ B. $2 \rightarrow 1 \rightarrow 3 \rightarrow 4$ C. $1 \rightarrow 3 \rightarrow 2 \rightarrow 4$ D. $1 \rightarrow 2 \rightarrow 4 \rightarrow 3$

Câu 15. Tại sao learning rate thích ứng lại hữu ích trong thực tế?

- A. Làm mô hình huấn luyện ngắn nhiên hơn
- B. Tự động điều chỉnh tốc độ học theo tham số cụ thể
- C. Tránh tràn số
- D. Hạn chế overfitting

Câu 16. An là một học sinh giỏi toán. Khi biết rằng các mô hình ngôn ngữ lớn (LLM) có thể giải được những bài toán phức tạp, An đã thử nghiệm nhưng kết quả không như mong đợi. Tuy nhiên, sau khi tìm hiểu và áp dụng kỹ thuật Chain-of-Thought, An nhận thấy mô hình bắt đầu giải đúng nhiều bài toán hơn. Vậy, kỹ thuật Chain-of-Thought là gì?

- A. Yêu cầu mô hình trả lời càng ngắn gọn càng tốt để tiết kiệm tài nguyên
- B. Huấn luyện mô hình dự đoán từ tiếp theo bằng dữ liệu song ngữ
- C. Yêu cầu mô hình sinh câu hỏi thay vì câu trả lời
- D. Thúc đẩy mô hình giải bài toán bằng cách liệt kê từng bước suy luận trung gian

Câu 17. Trọng số các thuộc tính trong k-NN được thực hiện bằng:

- A. Cập nhật giá trị thuộc tính của mẫu dữ liệu theo các trọng số khác nhau
- B. Thuộc tính quan trọng hơn được thêm vào bởi trọng số lớn hơn
- C. Điều chỉnh phép tính khoảng cách bằng cách nhân từng thuộc tính với trọng số
- D. Một ma trận trọng số được sử dụng để tính toán các hàng xóm

Câu 18. Chúng ta muốn phân 7 điểm dữ liệu vào trong 3 cụm sử dụng thuật toán K-means (với khoảng cách Euclid). Giả sử rằng sau vòng lặp đầu tiên, các cụm C1, C2 và C3 chứa các điểm dữ liệu sau (trong không gian 2 chiều): C1 chứa 2 điểm dữ liệu: (0,6), (6,0); C2 chứa 3 điểm dữ liệu: (2,2), (4,4), (6,6); C3 chứa 2 điểm dữ liệu: (5,5), (7,7). Tâm của 3 cụm sẽ là?

- | | |
|----------------------------------------|--------------------------------------|
| A. C1: (0,0), C2: (48,48), C3: (35,35) | B. C1: (3,3), C2: (4,4), C3: (6,6) |
| C. C1: (6,6), C2: (12,12), C3: (12,12) | D. C1: (3,3), C2: (6,6), C3: (12,12) |

Câu 19. GloVe là một phương pháp nhúng từ (word embedding). Phát biểu nào sau đây mô tả đúng cách mà nhúng từ GloVe được tạo ra?

- A. Được tạo ra trong quá trình dịch máy
- B. Sử dụng cơ chế chú ý (attention) để tạo vector
- C. Được huấn luyện từ ma trận đồng xuất hiện (co-occurrence matrix) toàn cục
- D. Một dạng vector gồm các số 0 và một số 1 (one-hot vector)

Câu 20. Mô hình BERT nhận đầu vào là gì?

- A. Vector ID, mặt nạ chú ý (attention mask), ID loại token
- B. Chỉ văn bản thô (raw text)
- C. Cặp câu với nhúng (embedding)
- D. Từ rời rạc

Câu 21. Nếu muốn sử dụng GPT để sinh câu trả lời dựa trên một đoạn văn, mô hình nào phù hợp nhất?

- | | |
|------------|-----------------------------------------|
| A. BERT | B. GPT-2 |
| C. GPT-Neo | D. GPT-3.5 hoặc GPT-4 với gợi ý phù hợp |

Câu 22. Nếu thêm bỏ ngẫu nhiên (dropout) vào mô hình nhưng độ chính xác kiểm tra (validation accuracy) giảm mạnh, bạn nên thử gì đầu tiên?

- | | |
|------------------------|----------------------------------------------|
| A. Dùng bộ tối ưu khác | B. Giảm xác suất bỏ ngẫu nhiên xuống nhỏ hơn |
| C. Tắt bỏ ngẫu nhiên | D. Tăng xác suất bỏ ngẫu nhiên lên 0.8 |

Câu 23. Trong bài toán phân loại văn bản, nếu mô hình học tốt các từ khóa rõ ràng nhưng không hiểu ngữ cảnh, phương pháp nào giúp cải thiện khả năng hiểu ngữ cảnh?

- A. Giảm số chiều nhúng (embedding)
- B. Dùng mô hình dựa trên chú ý (attention-based) như BERT
- C. Chuyển sang dùng TF-IDF
- D. Bỏ nhúng (embedding), dùng vector gồm các số 0 và một số 1 (one-hot vector)

Câu 24. Giả sử rằng bạn sử dụng phương pháp 1-láng giềng gần nhất (1-NN) để dự đoán nhãn lớp cho dữ liệu x , dựa trên tập huấn luyện D và thước đo khoảng cách d . 1-NN sẽ đưa ra dự đoán nào cho x ?

- A. y^* trong đó $(a^*, y^*) = \arg \min_{(a,y) \in D} d(x, y)$
- B. y^* trong đó $(a^*, y^*) = \arg \min_{(a,y) \in D} d(x, a)$
- C. a^* trong đó $(a^*, y^*) = \arg \min_{(a,y) \in D} d(x, a)$
- D. y^* trong đó $(a^*, y^*) = \min_{(a,y) \in D} d(x, a)$

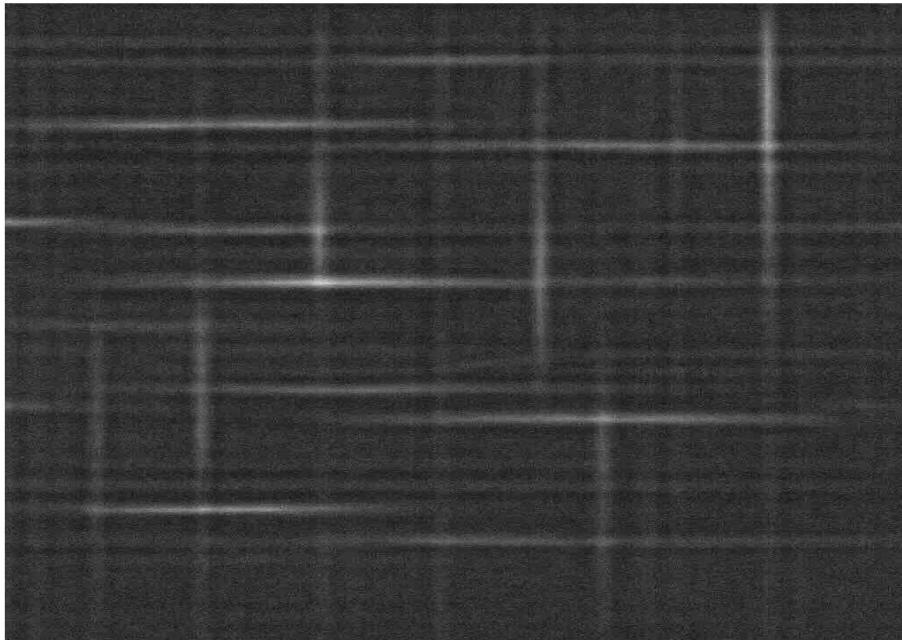
Câu 25. Xét tập huấn luyện gồm 8 mẫu dưới đây. Mỗi mẫu được mô tả bằng 3 đặc trưng số ($F1, F2, F3$) và thuộc về một trong ba lớp.

ID	F1	F2	F3	Lớp
S1	2	2	0	Đỏ
S2	1	3	1	Đỏ
S3	0	2	2	Đỏ
S4	8	7	7	Xanh dương
S5	9	6	6	Xanh dương
S6	7	7	8	Xanh dương
S7	5	2	5	Xanh lá
S8	6	1	4	Xanh lá

Sử dụng thuật toán K-láng giềng gần nhất (k-NN) với khoảng cách Euclid và $k = 3$, lớp nào sẽ được dự đoán cho điểm truy vấn $Q = (6, 2, 6)$?

- A. Xanh dương
- B. Đỏ
- C. Xanh lá
- D. Hòa thuật toán không thể quyết định

Câu 26. Hình dưới đây biểu thị bản đồ đặc trưng (feature map) thu được sau khi áp dụng một bộ lọc Conv2D lên ảnh đầu vào kích thước 128×128 .



Hãy cho biết bộ lọc này đang phát hiện đặc trưng gì nhất?

- A. Các mẫu bề mặt (texture patterns)
- B. Các đốm màu (color blobs)

C. Các cạnh theo hướng kết hợp ngang và dọc (cross directional edges)

D. cạnh theo hướng ngang (horizontal edges)

Câu 27. Cho đoạn mã dưới đây, kích thước của output_tensor là bao nhiêu?

```
1 import tensorflow as tf
2
3 input_tensor = tf.constant(tf.random.normal(shape=(1, 32, 32, 3)), dtype=
4   tf.float32)
5
6 conv_layer = tf.keras.layers.Conv2D(filters=32, kernel_size=(5, 5),
7   strides=(2, 2), padding='same')
8
9 output_tensor = conv_layer(input_tensor)
10
11 print(output_tensor.shape)
```

- A. (1, 16, 16, 32) B. (1, 16, 16, 3) C. (1, 14, 14, 32) D. (1, 32, 32, 32)

Câu 28. Mục đích chính của lấy mẫu phủ định (Negative Sampling) trong huấn luyện mô hình vectơ từ là gì?

- A. Tăng số lượng tham số của mô hình.
B. Giúp mô hình sinh ra văn bản dài hơn và tự nhiên hơn.
C. Giúp cải thiện độ chính xác mô hình.
D. Giảm chi phí tính toán khi huấn luyện trên tập từ vựng lớn.

Câu 29. Khi sử dụng bộ tối ưu Adam, nếu mất mát huấn luyện (training loss) ngừng giảm sút (plateau), bạn nên thử gì tiếp theo?

- A. Đặt lại toàn bộ mô hình
B. Tăng kích thước lô (batch size)
C. Bỏ chuẩn hóa lô (BatchNorm)
D. Giảm tỷ lệ học (learning rate) hoặc thử lại với SGD

Câu 30. Học tự giám sát (self-supervised learning) thường sử dụng phương pháp nào?

- A. Đầu phân loại (classification head)
B. Phân cụm (clustering)
C. Tăng cường dữ liệu (augmentation) và mất mát đối lập (contrastive loss)
D. Gắn nhãn thủ công

Câu 31. Chuẩn hóa rất quan trọng đối với k-NN vì

- A. Giúp tránh được vấn đề một thuộc tính có thể đóng vai trò quyết định, lấn át các thuộc tính khác.
B. Nó biến đổi các giá trị thuộc tính thành phạm vi [0, 1] để dễ dàng tính toán.
C. Nó cho phép so sánh và phân tích có ý nghĩa giữa các biến.
D. Nó cần thiết để tính toán khoảng cách giữa các mẫu dữ liệu.

Câu 32. Cho bản đồ đặc trưng sau. Giá trị ở vị trí (0,0) của bản đồ đặc trưng đầu ra sau khi áp dụng lớp gộp trung bình (Average Pooling) với kích thước cửa sổ 3×3 và bước nhảy 2 là:

```
1 [[10, 20, 30, 40],
2  [50, 60, 70, 80],
3  [90, 100, 110, 120],
4  [130, 140, 150, 160]]
```

- A. 55 B. 70 C. 60 D. 50

Câu 33. Chương trình sau thực hiện:

- Tải ResNet-50 pre-trained trên ImageNet.
 - Đóng băng toàn bộ các layer convolution.
 - Lấy output của lớp avgpool(shape(2048,1,1)) và flatten thành vector 2048.
- Bạn thiếu dòng nào dưới đây để trả về vec tơ đặc trưng (feature vector)?

```

1 import torch, torchvision.models as models
2
3 model = models.resnet50(weights='DEFAULT')
4
5 for p in model.parameters():
6     p.requires_grad_(False)
7
8
9 model.fc = torch.nn.Identity()
10
11 x = torch.randn(1, 3, 224, 224)
12 ... # <-- fil here
13 print(features.shape)

```

- A. features = model(x)
 C. features = model.avgpool(x)

- B. features = model.layer4(x)
 D. features = x

Câu 34. Bạn muốn sử dụng một mô hình Mạng Nơ-ron Tích chập (CNN) cho nhiệm vụ phân tích ảnh viễn thám (ảnh chụp từ vệ tinh, máy bay). Bạn có hai lựa chọn:

- Huấn luyện từ đầu: Xây dựng và huấn luyện một mô hình CNN hoàn toàn mới chỉ sử dụng bộ dữ liệu ảnh viễn thám (giả sử có kích thước trung bình).
- Tinh chỉnh (Fine-tuning): Lấy một mô hình CNN đã được huấn luyện trước trên một tập dữ liệu lớn gồm ảnh tự nhiên (ví dụ: ImageNet - ảnh chó, mèo, ô tô, v.v.) và điều chỉnh (tinh chỉnh) nó cho phù hợp với bộ dữ liệu ảnh viễn thám của bạn.

So với việc huấn luyện mô hình từ đầu, phương pháp tinh chỉnh mang lại nhiều lợi ích. Tuy nhiên, điều nào dưới đây **KHÔNG** phải là một ưu điểm điển hình của việc tinh chỉnh trong tình huống này?

- A. Giảm nguy cơ quá khóp (overfitting) vì có ít tham số cần cập nhật
 B. Khắc phục được hoàn toàn được vấn đề về sự khác biệt giữa đặc điểm ảnh tự nhiên và ảnh viễn thám.
 C. Mô hình hội tụ nhanh hơn do kế thừa trọng số đặc trưng cơ bản
 D. Tận dụng đặc trưng cấp thấp tổng quát

Câu 35. Một bộ lọc hình vuông, mỗi cạnh của nó dài k ô vuông. Bộ lọc này được dùng để quét qua một bức ảnh ban đầu có chiều ngang m ô vuông và chiều dọc n ô vuông ($m, n > k$). Mỗi lần quét, bộ lọc sẽ dịch chuyển đi 2 ô vuông theo cả chiều ngang lẫn chiều dọc. Quá trình quét này không thêm bất kỳ lớp đệm nào vào xung quanh ảnh gốc. Hỏi sau khi quét xong, bức ảnh kết quả (bản đồ đặc trưng đầu ra) sẽ có kích thước bao nhiêu ô vuông chiều ngang và bao nhiêu ô vuông chiều dọc?

- A. $(\frac{m}{2}, \frac{n}{2})$ B. $(\lfloor \frac{m-k}{2} \rfloor + 1, \lfloor \frac{n-k}{2} \rfloor + 1)$ C. $(\frac{m-k}{2} + k, \frac{n-k}{2} + k)$ D. $(\frac{m-k+1}{2}, \frac{n-k+1}{2})$

Câu 36. Khi dùng ViT để phân loại ảnh, lớp cuối cùng thường là gì?

- A. Khối chú ý (attention block)
 B. Bỏ ngẫu nhiên (dropout)
 C. Lớp kết nối đầy đủ (fully connected) và hàm Softmax
 D. LSTM

Câu 37. Dựa trên tập dữ liệu trong Bảng 1 dưới đây để xây dựng cây quyết định, hãy tính xấp xỉ entropy $H(Passed)$. Cây quyết định này dự đoán liệu sinh viên có qua môn hay không (T là có, F là không), dựa trên điểm CGPA (H: cao, M: trung bình, L: thấp) và việc có ôn tập hay không (T hoặc F).

CGPA	Ôn tập	Qua môn
H	F	T
H	T	T
M	F	F
M	T	T
L	F	F
L	T	T

- A. 0.66 B. 1.92 C. 0.92 D. 1.32

Câu 38. Một bức ảnh thuộc vào một trong hai lớp: 'chó' hoặc 'mèo'. Nhãn thực tế (ground truth label) của ảnh này được biểu diễn dưới dạng *one-hot encoding* là [0, 1], trong đó vị trí thứ nhất tương ứng với lớp 'chó' và vị trí thứ hai tương ứng với lớp 'mèo'. Mô hình của chúng ta đã dự đoán xác suất cho ảnh này là [0.3, 0.7], nghĩa là xác suất dự đoán là 0.3 cho lớp 'chó' và 0.7 cho lớp 'mèo'. Hãy tính giá trị của hàm mất mát *cross-entropy* cho dự đoán này, sử dụng logarit tự nhiên (\ln). Công thức hàm mất mát như sau:

- A. 0.105 B. 0.247 C. 0.357 D. 0.713

Câu 39. Khi tinh chỉnh (fine-tune) ResNet34, nếu suy luận (inference) chậm, bạn nên thử gì để tăng tốc độ?

- A. Chuyển sang dùng mô hình nhỏ hơn như MobileNet
- B. Thêm lớp bỏ ngẫu nhiên (dropout) vào suy luận
- C. Tăng số vòng lặp (epoch)
- D. Giảm số lớp (class)

Câu 40. Khi huấn luyện mạng nơ-ron đa tầng (MLP) bằng phương pháp tối ưu hóa theo lô nhỏ (mini-batch SGD), bạn cần làm gì sau mỗi vòng lặp (epoch) để đảm bảo mô hình học hiệu quả và tránh thiên lệch?

- A. Xáo trộn dữ liệu (shuffle) huấn luyện
- B. Sử dụng toàn bộ dữ liệu (full-batch) để cập nhật trọng số
- C. Đặt lại trọng số về giá trị ban đầu
- D. Chuyển sang sử dụng bộ tối ưu Adam

Câu 41. Với Lấy mẫu phủ định (Negative Sampling), các mẫu sẽ được chọn như thế nào?

- A. Là các từ có độ tương đồng ngữ nghĩa cao với từ mục tiêu.
- B. Là các từ gần nhất với từ ngữ cảnh trong văn bản.
- C. Là các từ có nhãn đúng trong tập huấn luyện.
- D. Được chọn ngẫu nhiên từ toàn bộ từ vựng, thường theo một phân phối cố định.

Câu 42. Cho trước tập dữ liệu không có nhãn gồm N điểm dữ liệu: $\{x_1, x_2, \dots, x_N\}$. Chúng ta chạy K-means với 50 lần khởi tạo ngẫu nhiên tâm cụm khác nhau (luôn với cùng số lượng tâm cụm K) và thu được 50 bộ tâm cụm khác nhau. Đâu là cách được gợi ý cho việc chọn 1 kết quả từ 50 kết quả trên để sử dụng?

- A. Chọn kết quả mà $\frac{1}{N} \sum_{i=1}^N \|x_i - m_{z_i}\|^2$ đạt giá trị nhỏ nhất trong 50 lần, với m_{z_i} là tâm cụm mà x_i được gán vào.
- B. Chọn lần chạy thứ mấy cũng có thể tốt
- C. Luôn chọn lần cuối cùng (thứ 50), vì lần này có khả năng đã hội tụ thành một giải pháp tốt
- D. Chỉ có cách duy nhất để chọn là yêu cầu dữ liệu phải có nhãn y_i

Câu 43. Sử dụng mô hình ResNet-18, với khoảng 11.7 triệu tham số, đã được tiền huấn luyện trên tập dữ liệu ImageNet làm điểm khởi đầu. Trong quá trình tinh chỉnh cho nhiệm vụ cụ thể, bạn áp dụng chiến lược đóng băng trọng số cho toàn bộ phần thân (backbone) của mô hình. Tuy nhiên, có một ngoại lệ đáng chú ý: bạn chỉ mở băng và cho phép cập nhật trọng số cho một khối duy nhất là BasicBlock thứ hai nằm trong lớp thứ tư (layer4), được gọi tắt là "block 4-2". Khối "block 4-2" này, bao gồm hai lớp tích chập 3×3 :

$$\text{Conv}_1 : 512 \xleftarrow[3 \times 3, 512, s=1]{} 512, \quad \text{Conv}_2 : 512 \xleftarrow[3 \times 3, 512, s=1]{} 512.$$

- Bỏ qua tham số và tính toán không cần thiết: Chúng ta sẽ không tính số lượng tham số bias và bỏ qua chi phí tính toán (FLOPs) của các lớp Batch Normalization (BN), hàm kích hoạt ReLU, và các kết nối tắt (skip connection/identity mapping).

- Định nghĩa FLOPs: Một phép toán dấu phẩy động (FLOP) được tính là một phép nhân và một phép cộng. Đối với lớp tích chập, tổng FLOPs xấp xỉ bằng 2 lần số phép tính nhân-cộng (MACs). ($FLOPs \approx 2 \times MACs$).

- Kích thước đầu vào cho khối: Khi mô hình xử lý một ảnh đầu vào gốc có kích thước 224×224 , tensor đặc trưng đi vào khối "block 4-2" có kích thước là (Batch=1, Channels=512, Height=7, Width=7).

Dựa trên giả sử trên, câu trả lời nào sau đây là đúng khi tính tổng số tham số có thể huấn luyện (*trainable parameters*) chỉ chứa trong khối "block 4-2" và tổng số FLOPs cần thiết để thực thi chỉ riêng khối "block 4-2" khi xử lý tensor đầu vào có kích thước (1,512,7,7).

- A. 11.7M; 1.8GFLOPs
- B. 4.7M; 0.46GFLOPs
- C. 0.50M; 0.05GFLOPs
- D. 8.4M; 0.82GFLOPs

Câu 44. Tỷ lệ học thích ứng (adaptive learning rate) như AdaGrad, RMSProp, Adam giúp gì cho mô hình?

- A. Giảm kích thước mô hình
- B. Tăng kích thước lô (batch size)
- C. Giảm số vòng lặp (epoch) cần thiết
- D. Tự động điều chỉnh tốc độ học cho từng tham số

Câu 45. Trong tối ưu hóa theo lô nhỏ (mini-batch SGD), nếu kích thước lô (batch size) quá nhỏ, hệ quả thường gặp là gì?

- A. Giảm thời gian huấn luyện
- B. Gradient quá nhiều, gây dao động
- C. Độ chính xác tăng nhanh
- D. Không ảnh hưởng

Câu 46. Những chiến lược nào có thể giúp giảm ván đề quá khít (overfitting) trong cây quyết định?

- i. Giới hạn độ sâu tối đa của cây
 - ii. Áp đặt số lượng mẫu tối thiểu tại các nút lá
 - iii. Cắt tia cây (pruning)
 - iv. Đảm bảo mỗi nút lá chứa duy nhất một lớp
- A. Không có lựa chọn nào đúng
 - B. Tất cả
 - C. (i), (ii) và (iii)
 - D. (i), (iii), (iv)

Câu 47. Một nơ-ron có 3 đầu vào với trọng số lần lượt là 1, 4 và 3, không có bias. Hàm truyền là hàm tuyến tính với hằng số tỷ lệ bằng 3. Các đầu vào lần lượt là 4, 8 và 5. Đầu ra sẽ là bao nhiêu?

- A. 162
- B. 153
- C. 139
- D. 160

Câu 48. Gradient của hàm số $2x^2 - 3y^2 + 4y - 10$ tại điểm $(0, 0)$ là gì?

- A. $1i + 10j$
- B. $2i - 3j$
- C. $-3i + 4j$
- D. $0i + 4j$

Câu 49. Khi huấn luyện mô hình phân loại nhiều lớp với tập dữ liệu mất cân bằng về nhãn lớp, nếu lớp hiếm (ít dữ liệu huấn luyện) gần như không được dự đoán, cách xử lý phù hợp là gì?

- A. Giảm số vòng lặp (epoch)
- B. Xóa lớp hiếm khỏi dữ liệu
- C. Tăng bỏ ngẫu nhiên (dropout)
- D. Sử dụng hàm mất mát có trọng số

Câu 50. Khi tinh chỉnh (fine-tune) MobileNetV2 trên tập dữ liệu nhỏ, bước đầu tiên bạn nên làm là gì để tối ưu hóa hiệu suất?

- A. Tăng số lớp
- B. Thay toàn bộ mô hình
- C. Đóng băng (freeze) các lớp đầu tiên
- D. Tăng tỷ lệ học (learning rate)

Câu 51. Khi huấn luyện bằng PyTorch, nếu GPU bị đầy bộ nhớ (Out-of-Memory - OOM), bạn nên thử gì đầu tiên?

- A. Chuyển sang dùng CPU
- B. Thêm bỏ ngẫu nhiên (dropout)
- C. Dùng mô hình lớn hơn
- D. Giảm kích thước lô (batch size) hoặc dùng tích lũy gradient (gradient accumulation)

Câu 52. Phương pháp nào sau đây KHÔNG thuộc nhóm học có giám sát?

- A. Cây quyết định
- B. Hồi quy tuyến tính với Ridge
- C. Naive Bayes
- D. K-means

Câu 53. Phát biểu nào sau đây là đúng về giải thuật học láng giềng gần nhất?

- A. Chỉ được sử dụng cho bài toán hồi quy
- B. Thuộc lớp bài toán học tham số
- C. Chỉ được sử dụng cho bài toán phân loại
- D. Được sử dụng cho cả bài toán phân loại và bài toán hồi quy

Câu 54. Mục đích chính của việc tăng cường dữ liệu trong huấn luyện mô hình học máy/học sâu là gì?

- A. Tăng tốc độ huấn luyện mô hình
- B. Giảm số lượng tham số của mô hình
- C. Giảm kích thước vật lý của tập dữ liệu gốc
- D. Cải thiện khả năng tổng quát hóa của mô hình

Câu 55. Khi sử dụng BertTokenizer từ Hugging Face, điều nào sau đây đúng?

- A. Bộ mã hóa không hỗ trợ lô (batch)
- B. tokenizer.encode_plus() trả về input_ids, attention_mask
- C. BERT chỉ dùng cho tiếng Anh
- D. Bộ mã hóa (tokenizer) không cần đệm (padding)

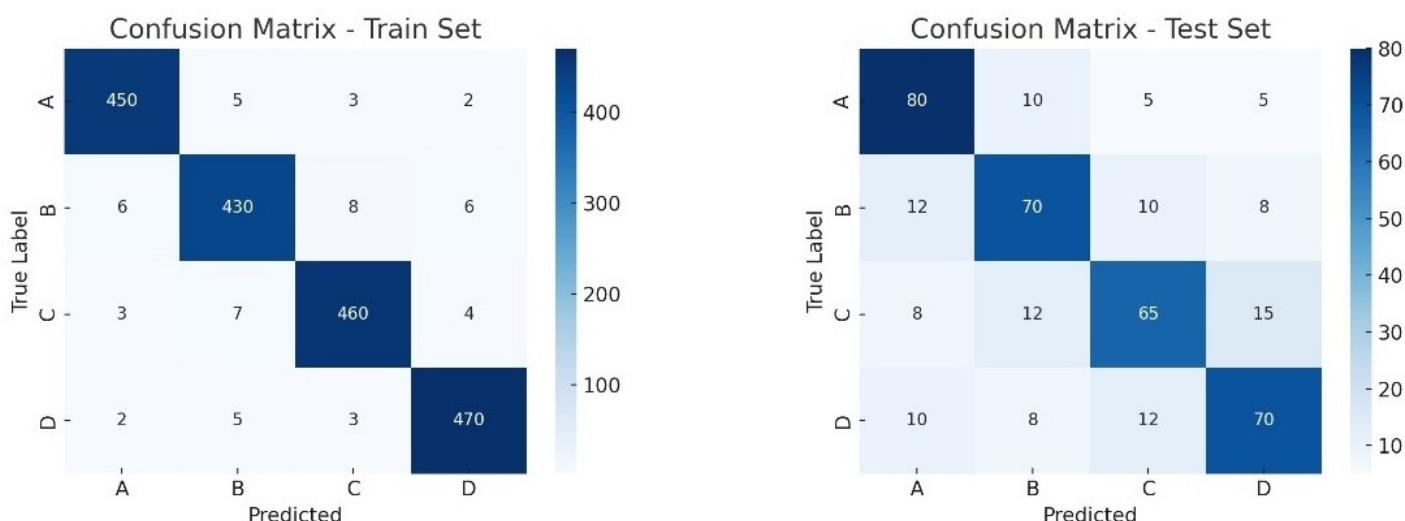
Câu 56. Dưới đây là một số lựa chọn để có thể thực hiện khi huấn luyện mạng nơ-ron. Đâu là trường hợp sẽ khiến mạng của bạn KHÓ đạt được độ chính xác cao trong tương lai?

- A. Đảo ngẫu nhiên lại dữ liệu khi bắt đầu mỗi epoch
- B. Khởi tạo tất cả bộ tham số bằng 0
- C. Sử dụng momentum
- D. Sử dụng Dropout

Câu 57. Trong mô hình SSD, hộp neo (anchor boxes) có vai trò gì?

- A. Định nghĩa trước các tỷ lệ và kích thước hộp
- B. Làm nhẹ mô hình
- C. Làm tăng số lớp
- D. Tăng độ phân giải ảnh

Câu 58. Quan sát hai ma trận nhầm lẫn (Confusion Matrix) dưới đây, nhận định nào sau đây là chính xác nhất trong số các lựa chọn được đưa ra?



- A. Mô hình có dấu hiệu quá khớp (Overfitting)
- B. Mô hình có dấu hiệu kém khớp (Underfitting).
- C. Độ chính xác trên tập kiểm thử và huấn luyện là gần như tương đương nhau.
- D. Sai số chỉ do phân bố lớp khác nhau giữa hai tập.

Câu 59. Entropy cao có nghĩa là các phần phân chia trong thuật toán cây quyết định ID3 thì

- A. Thuần khiết (Pure): Các điểm dữ liệu ở mỗi nhánh của cây quyết định tập trung đa số vào một lớp
- B. Không có ý nghĩa gì
- C. Không thuần khiết (Not pure): Các điểm dữ liệu ở mỗi nhánh của cây quyết định phân bố tương đối đều vào các lớp
- D. Có thể suy ra độ đo F1-score cao

Câu 60. Mục đích của việc thêm nhiều Gaussian vào dữ liệu đầu vào khi huấn luyện là gì?

- A. Giảm số lớp cần thiết
- B. Tăng tính ổn định và khả năng chống nhiễu
- C. Giảm thời gian huấn luyện
- D. Tăng độ chính xác

Câu 61. Khi huấn luyện mạng GAN, nếu bộ tạo (generator) tạo ra ảnh toàn màu xám, nguyên nhân có thể là gì?

- A. Mất mát (loss) quá nhỏ
- B. Tỷ lệ học (learning rate) quá nhỏ
- C. Kích thước lô (batch size) quá lớn
- D. Mất cân bằng giữa bộ phân biệt (discriminator) và bộ tạo (generator)

Câu 62. Trong bài toán phân loại văn bản, nếu mô hình học tốt các từ khóa rõ ràng nhưng không hiểu ngữ cảnh, phương pháp nào giúp cải thiện khả năng hiểu ngữ cảnh?

- A. Dùng mô hình dựa trên chú ý (attention-based) như BERT
- B. Bỏ nhúng (embedding), dùng vector gồm các số 0 và một số 1 (one-hot vector)
- C. Chuyển sang dùng TF-IDF
- D. Giảm số chiều nhúng (embedding)

Câu 63. Nếu mô hình bị học quá mức (overfitting), phương pháp nào nên thử đầu tiên để giảm hiện tượng này?

- A. Tăng số vòng lặp (epoch)
- B. Thêm bỏ ngẫu nhiên (dropout) hoặc tăng suy giảm trọng số (weight decay)
- C. Tăng tỷ lệ học (learning rate)
- D. Giảm kích thước lô (batch size)

Câu 64. Khi huấn luyện, nếu độ chính xác huấn luyện (training accuracy) tăng đều nhưng độ chính xác kiểm tra (validation accuracy) dao động mạnh và không cải thiện, nguyên nhân có thể là gì?

- A. Học quá mức (overfitting) hoặc dữ liệu kiểm tra chưa được xáo trộn kỹ
- B. Không sử dụng bỏ ngẫu nhiên (dropout)
- C. Mô hình quá nhỏ
- D. Số vòng lặp (epoch) quá ít

Câu 65. Mục tiêu chính khi tinh chỉnh (fine-tune) mô hình BERT cho bài toán phân loại văn bản là gì?

- A. Tạo nhúng (embedding)
- B. Thay lớp cuối bằng một lớp phân loại
- C. Dùng mô hình sinh
- D. Tạo một bộ mã hóa (tokenizer) mới

Câu 66. Nếu tỷ lệ học (learning rate) quá cao, điều gì có thể xảy ra trong quá trình huấn luyện?

- A. Mô hình dao động và không hội tụ
- B. Tăng khả năng regularization
- C. Hàm mất mát giảm đều đặn
- D. Mô hình hội tụ nhanh hơn

Câu 67. Sự khác biệt giữa tập kiểm tra (test set) và tập xác thực (validation set) là gì?

- A. Tập xác thực là không cần thiết trong học máy.
- B. Tập xác thực dùng để điều chỉnh siêu tham số, còn tập kiểm tra dùng để đánh giá hiệu suất của mô hình.
- C. Tập xác thực và tập kiểm tra là một.
- D. Tập xác thực dùng để đánh giá hiệu suất mô hình trong quá trình huấn luyện, trong khi tập kiểm tra dùng để đánh giá sau khi huấn luyện.

Câu 68. Xét một Random Forest gồm K cây với nhiệm vụ hồi quy. Mỗi cây quyết định i có thể biểu diễn một hàm $T_i(x)$ của đầu vào x . Random Forest này biểu diễn hàm nào sau đây?

- A. $y(x) = \sum_{i=1}^K T_i(x)$
- B. $y(x) = \max_{i \in 1, \dots, K} T_i(x)$
- C. $y(x) = \sum_{i=1}^K T_i(\frac{x}{K})$
- D. $y(x) = \frac{1}{K} \sum_{i=1}^K T_i(x)$

Câu 69. Để tải nhúng từ (embedding) Word2Vec đã được huấn luyện trước trong thư viện gensim, bạn sử dụng lệnh nào?

- A. gensim.models.load("word2vec")
- B. import word2vec.load_model(path)
- C. KeyedVectors.load_word2vec_format(path)
- D. spacy.load_word2vec(path)

Câu 70. Khi sử dụng câu lệnh nn.CrossEntropyLoss trong PyTorch, bạn nên đưa gì vào đối số đầu tiên?

- A. Logits do lớp cuối cùng của mạng tạo ra
- B. Vector gồm các số 0 và một số 1 (one-hot vector) biểu diễn các lớp mục tiêu
- C. Xác suất thu được sau khi áp dụng softmax lên đầu ra của mạng
- D. Log-xác suất sau khi áp dụng log_softmax lên đầu ra của mạng

Câu 71. Trong huấn luyện, nếu mất mát huấn luyện (training loss) giảm nhưng mất mát kiểm tra (validation loss) tăng, điều này cho thấy gì?

- A. Mô hình đang học quá mức (overfitting)
- B. Mô hình đang hội tụ
- C. Cần tăng tỷ lệ học (learning rate)
- D. Mô hình học chưa đủ (underfitting)

Câu 72. Trong các bài toán phân đoạn ảnh (image segmentation), một thách thức phổ biến là sự mất cân bằng nghiêm trọng giữa các lớp. Ví dụ, diện tích của đối tượng cần phân đoạn (lớp tiền cảnh - foreground) có thể rất nhỏ so với phần còn lại của ảnh (lớp hậu cảnh - background). Khi gặp tình huống mất cân bằng lớp như vậy, hàm mất mát (loss function) nào trong số các lựa chọn dưới đây thường được xem là hiệu quả hơn và được ưu tiên sử dụng thay cho hàm Cross-Entropy (CE) tiêu chuẩn?

A. Mean Squared Error (MSE)

C. L_1 Loss

B. Hinge Loss

D. Dice Loss hoặc Focal Loss

Câu 73. Mạng nơ-ron ResNet sử dụng một kỹ thuật quan trọng gọi là kết nối tắt (skip connection) để giải quyết hiện tượng biến mất đạo hàm (Vanishing Gradient) trong quá trình huấn luyện. Dựa trên đoạn mã của khối identity_block dưới đây, hãy liệt kê các thành phần chính của khối theo đúng thứ tự xuất hiện và chỉ ra dòng mã thực hiện phép kết nối tắt.

```
1 def identity_block(X, f, filters, stage, block):
2
3     conv_name_base = 'res' + str(stage) + block + '_branch'
4     bn_name_base = 'bn' + str(stage) + block + '_branch'
5
6     F1, F2, F3 = filters
7
8     X_shortcut = X
9
10    X = Conv2D(filters = F1, kernel_size = (1, 1), strides = (1,1),
11                padding = 'valid', name = conv_name_base + '2a', kernel_initializer =
12                glorot_uniform(seed=0))(X)
13    X = BatchNormalization(axis = 3, name = bn_name_base + '2a')(X)
14    X = Activation('relu')(X)
15
16    X = Conv2D(filters = F2, kernel_size = (f, f), strides = (1,1),
17                padding = 'same', name = conv_name_base + '2b', kernel_initializer =
18                glorot_uniform(seed=0))(X)
19    X = BatchNormalization(axis = 3, name = bn_name_base + '2b')(X)
20    X = Activation('relu')(X)
21
22    X = Conv2D(filters = F3, kernel_size = (1, 1), strides = (1,1),
23                padding = 'valid', name = conv_name_base + '2c', kernel_initializer =
24                glorot_uniform(seed=0))(X)
25    X = BatchNormalization(axis = 3, name = bn_name_base + '2c')(X)
26
27    X = Add()([X_shortcut, X])
28    X = Activation('relu')(X)
29
30    return X
```

- A. Ba cặp Conv2D–BatchNorm–ReLU; kết nối tắt ở dòng ở dòng 8
- B. Hai cặp Conv2D–BatchNorm–ReLU; kết nối tắt nằm trong BatchNorm ở dòng 11, 15, 19
- C. Ba cặp Conv2D–BatchNorm–ReLU; Không có cơ chế kết nối tắt ở dòng
- D. Ba cặp Conv2D–BatchNorm–ReLU; kết nối tắt ở dòng 21

Câu 74. Trong mô hình Transformer, cơ chế chú ý (attention) giúp mô hình làm gì?

- A. Tự động sinh từ tiếp theo
- B. Tập trung vào phần quan trọng của câu khi tính toán
- C. Xác định vị trí từ trong câu
- D. Chuẩn hóa đầu vào

Câu 75. Trong các kiến trúc mạng phân đoạn ảnh dạng bộ mã hóa - bộ giải mã (encoder-decoder), ví dụ như U-Net, các "kết nối tắt" (skip connections) đóng vai trò quan trọng. Chúng kết hợp thông tin đặc trưng từ các lớp ở bộ mã hóa (encoder path) với thông tin tương ứng ở bộ giải mã (decoder path) sau khi được phóng đại (upsampling). Điều này giúp mô hình giữ lại các chi tiết không gian có độ phân giải cao bị mất trong quá trình mã hóa.

Giả sử bạn đang xây dựng một lớp trong phần bộ giải mã của mạng U-Net bằng Python, sử dụng một framework học sâu như TensorFlow/Keras hoặc PyTorch. Bạn có hai tensor:

- *encoder_output*: Tensor chứa các đặc trưng từ một lớp tương ứng ở bộ mã hóa.

- *decoder_input*: Tensor đầu vào cho lớp hiện tại ở bộ giải mã, đã được upsample để có cùng kích thước chiều cao (H) và chiều rộng (W) với *encoder_output*.

```
1  
2 # encoder_output.shape = (batch, H, W, C1) in TF/Keras or (batch, C1, H, W  
    ) in PyTorch  
3 # decoder_input.shape = (batch, H, W, C2) in TF/Keras or (batch, C2, H, W)  
    in PyTorch  
4  
5 merged_features = some_concatenation_operation([decoder_input,  
    encoder_output], axis=...)
```

Thao tác *some_concatenation_operation* và tham số *axis* phù hợp nhất để thực hiện kết nối tắt (skip connection) kiểu U-Net là gì?

- A. Phép nối (Concatenation) và axis dọc theo chiều batch
- B. Phép nối (Concatenation) và axis dọc theo chiều kênh (channel dimension)
- C. Phép cộng element-wise và axis không quan trọng
- D. Phép nhân element-wise và axis không quan trọng

Câu 76. Mô hình ngôn ngữ lớn (Large Language Model - LLM) là gì?

- A. Một công cụ tìm kiếm dựa trên quy tắc
- B. Một mô hình học sâu được huấn luyện trên lượng lớn dữ liệu văn bản để dự đoán từ tiếp theo trong một chuỗi
- C. Một hệ thống học tăng cường chuyên cho bài toán xử lý ngôn ngữ
- D. Một cơ sở dữ liệu ngữ nghĩa với các quan hệ giữa từ

Câu 77. Stable Diffusion thuộc loại mô hình nào trong các lựa chọn sau?

- A. Bộ chuyển đổi (Transformer)
- B. Mô hình khuếch tán (Diffusion Model)
- C. GAN
- D. Bộ mã hóa tự động (autoencoder)

Câu 78. Để áp dụng kỹ thuật dừng sớm (Early Stopping) trong huấn luyện, bạn cần theo dõi chỉ số nào để quyết định dừng huấn luyện?

- A. Mất mát kiểm tra (validation loss) hoặc độ chính xác kiểm tra (validation accuracy)
- B. Tỷ lệ học (learning rate)
- C. Số vòng lặp (epoch count)
- D. Mất mát huấn luyện (training loss)

Câu 79. Lệnh nào dùng để chuyển mô hình sang GPU?

- A. model.gpu()
- B. model.to('cuda')
- C. model.cuda.enable()
- D. model.device('GPU')

Câu 80. Hàm chi phí sử dụng MSE (mean squared error) từ dữ liệu sau là bao nhiêu?

Giá trị kỳ vọng	15	17	10	26	14	12	11	13
Giá trị thực tế	12	19	15	24	13	14	8	11

- A. 8.5
- B. 6.5
- C. 5.5
- D. 7.5

Câu 81. Thư viện LangChain được sử dụng để làm gì?

- A. Phân tích âm thanh
- B. Dịch máy
- C. Sinh văn bản ngẫu nhiên
- D. Kết nối và xây dựng quy trình cho tác nhân ngôn ngữ lớn (LLM agent pipeline)

Câu 82. Trong mô hình hồi quy logistic sử dụng scikit-learn, thuộc tính nào chứa trọng số đã học của mô hình?

- A. model.weights
- B. model.intercept
- C. model.coefficients
- D. model.coef

Câu 83. Bạn đang phát triển một hệ thống phân tích ảnh vệ tinh để xác định các loại cây trồng khác nhau trong một khu vực nông nghiệp rộng lớn. Sau khi tiền xử lý và trích xuất đặc trưng từ ảnh đa phỏ, bạn áp dụng một

mô hình phân cụm (clustering) dựa trên thuật toán K-means để phân đoạn các vùng có khả năng chứa cùng một loại cây trồng.

Giả sử sau khi chạy thuật toán K-means với K=3, bạn thu được ba cụm C_1 , C_2 , C_3 đại diện cho ba loại cây trồng tiềm năng. Để đánh giá chất lượng phân cụm, bạn quyết định sử dụng một biến thể của Silhouette Coefficient.

Silhouette Coefficient cho một điểm dữ liệu i được định nghĩa là: $s(i) = \frac{b(i)-a(i)}{\max(a(i), b(i))}$, trong đó $a(i)$ là khoảng cách trung bình từ điểm i đến tất cả các điểm khác trong cùng cụm. $b(i)$ là khoảng cách trung bình từ điểm i đến tất cả các điểm trong cụm lân cận gần nhất (khác với cụm của i).

Trong trường hợp đang xét, mỗi pixel trong ảnh sau khi trích xuất đặc trưng có thể được coi là một điểm dữ liệu trong không gian đặc trưng nhiều chiều. Giả sử bạn chọn ngẫu nhiên một pixel p thuộc cụm C_1 . Sau khi tính toán, bạn có các giá trị sau:

- Khoảng cách trung bình từ pixel p đến tất cả các pixel khác trong cụm C_1 là $a(p) = 0.35$.
 - Khoảng cách trung bình từ pixel p đến tất cả các pixel trong cụm C_2 là $d(p, C_2) = 0.60$.
 - Khoảng cách trung bình từ pixel p đến tất cả các pixel trong cụm C_3 là $d(p, C_3) = 0.45$.

Biết rằng $b(p)$ là khoảng cách trung bình nhỏ nhất từ pixel p đến các điểm trong một cụm khác với cụm chứa p . Hãy tính Silhouette Coefficient $s(p)$ cho pixel p .

- A.** 0.0 **B.** 0.222 **C.** 0.125. **D.** 0.308

Câu 84. Nếu khởi tạo trọng số (weight initialization) không phù hợp, hiện tượng nào có thể xảy ra trong quá trình huấn luyện?

- A. Mô hình học nhanh hơn
 - B. Không ảnh hưởng vì bộ tối ưu sẽ điều chỉnh
 - C. Học quá mức nhẹ (overfitting)
 - D. Gradient biến mất (vanishing) hoặc nổ (exploding)

Câu 85. Để đánh giá mô hình phân loại ảnh với các lớp không cân bằng, chỉ số nào nên dùng thay vì chỉ độ chính xác (accuracy)?

- A.** AUC (Area Under the Curve)
B. Điểm F1 trung bình (Macro F1-score)
C. RMSE (Root mean square error)
D. Độ chính xác cao nhất (Top-1 accuracy)

Câu 86. Sau khi huấn luyện mô hình phân loại với độ chính xác 90%, khách hàng muốn triển khai thực tế. Việc cần làm tiếp là gì?

- A. Nén ảnh đầu vào
 - B. Chuyển mô hình sang TensorRT/ONNX để triển khai (deploy)
 - C. Tăng thêm số vòng lặp (epoch) để đạt 95%
 - D. Thay mô hình sang BERT

Câu 87. Khi xử lý dữ liệu ảnh, nếu một số ảnh bị hỏng (không mở được), cách tốt nhất để xử lý khi huấn luyện là gì?

- A.** Tăng kích thước lô (batch size) để bù lại
C. Bắt lỗi khi tải ảnh và bỏ qua ảnh bị hỏng

B. Bỏ qua toàn bộ thư mục chứa ảnh đó
D. Dừng toàn bộ huấn luyện

Câu 88. Trong PyTorch, để thêm lớp bỏ ngẫu nhiên (dropout) với xác suất 0.5 vào mạng nơ-ron, bạn sử dụng lệnh nào?

- A.** F.dropout(0.5) **B.** nn.dropout(0.5) **C.** nn.Dropout(p=0.5) **D.** nn.Dropout2d(0.5)

Câu 89. Hãy xem xét một công cụ phát hiện các gói tin chứa mối đe dọa trong trường hợp chỉ có một số lượng nhỏ gói là mối đe dọa. Yêu cầu là công cụ cần phát hiện các gói đe dọa mà không bỏ sót gói nào. Đâu là độ đo quan trọng nhất để đánh giá công cụ?

- ### A. Accuracy B. F1 C. Recall D. Precision

Câu 90. Phương pháp nào dưới đây mà việc chuẩn hóa các thuộc tính dữ liệu đầu vào không ảnh hưởng đến kết quả dự đoán?

C. Soft-margin SVM

D. k-NN

Câu 91. Mục tiêu huấn luyện của mô hình CBOW (Continuous Bag-of-Words) là gì?

- A. Sử dụng toàn bộ văn bản để dự đoán một từ bất kỳ
- B. Sử dụng vị trí của từ trong câu để dự đoán nghĩa của câu
- C. Sử dụng các từ xung quanh để dự đoán từ trung tâm
- D. Sử dụng từ trung tâm để dự đoán các từ xung quanh

Câu 92. Mục tiêu chính của phương pháp học tương phản (contrastive learning) trong lĩnh vực học tự giám sát (self-supervised learning) là gì?

- A. Tối thiểu hoá khoảng cách Euclidean giữa mọi cặp ảnh trong batch
- B. Khôi phục ảnh gốc từ ảnh đã bị thêm nhiễu Gaussian
- C. Đưa các ảnh được biến đổi từ cùng một ảnh gốc thông qua các phép biến đổi cơ bản tới gần nhau, đầy các mẫu lấy từ các ảnh khác nhau xa nhau trên không gian biểu diễn (embedding space)
- D. Tối ưu hoá hàm cross-entropy có nhãn đầy đủ

Câu 93. Điểm khác biệt chính giữa Stochastic Gradient Descent (SGD) và Mini-Batch Gradient Descent là gì?

- A. SGD luôn hội tụ nhanh hơn
- B. Mini-Batch Gradient Descent là một thuật toán hoàn toàn khác
- C. SGD không dùng được trong mạng nơ-ron
- D. SGD cập nhật tham số sau mỗi mẫu dữ liệu, Mini-Batch thì sau một nhóm mẫu

Câu 94. Giả sử các từ được biểu diễn bởi các vector 4 chiều như sau:

$$w1 = [0.8, 0.6, 0.0, 0.2]$$

$$w2 = [0.9, 0.5, 0.1, 0.3]$$

$$w3 = [1.0, 0.1, 0.0, 0.0]$$

$$w4 = [0.0, 0.1, 0.9, 0.3]$$

Hãy tính độ tương đồng cosine giữa w1 và các từ còn lại (w2, w3, w4), sau đó chọn từ gần nhất với w1.

Công thức tính cosine similarity giữa hai vector A và B là:

$$\text{cosine - similarity}(A, B) = \left(\frac{A \cdot B}{|A| \times |B|} \right)$$

Với: $A \cdot B$ là tích vô hướng giữa hai vector, $|A|$ là độ dài của vector A, tính bằng căn bậc hai tổng bình phương các thành phần

- A. Có nhiều hơn một từ gần nhất với w1
- B. w3
- C. w4
- D. w2

Câu 95. Mạng GAN bao gồm hai mạng chính nào?

- A. Bộ phát hiện (Detector) và Bộ phân đoạn (Segmentor)
- B. Bộ chuyển đổi (Transformer) và Cơ chế chú ý (Attention)
- C. Bộ mã hóa (Encoder) và Bộ giải mã (Decoder)
- D. Bộ tạo (Generator) và Bộ phân biệt (Discriminator)

Câu 96. Quá khớp (Overfitting) có thể do

- A. Lựa chọn mô hình không phù hợp
- B. Độ phức tạp của bài toán học
- C. Một lỗi nào đó trong quá trình huấn luyện
- D. Nhiều trong dữ liệu

Câu 97. Khi huấn luyện bộ mã hóa tự động (autoencoder), nếu ảnh đầu ra bị mờ, nguyên nhân có thể là gì?

- A. Bỏ ngẫu nhiên (dropout) quá thấp
- B. Bộ tối ưu sai
- C. Lớp ẩn quá nhỏ hoặc chính quy hóa (regularization) quá mạnh
- D. Kích thước lô (batch size) lớn

Câu 98. Mô hình DALL·E có khả năng đặc biệt nào?

- A. Phân đoạn vật thể
- B. Sinh ảnh từ mô tả văn bản
- C. Nén ảnh thành vector
- D. Sinh mô tả từ ảnh

Câu 99. Trong xử lý ngôn ngữ tự nhiên (NLP), mục đích chính của việc loại bỏ từ dừng (stopword) khỏi văn bản là gì?

- A. Để giảm số lượng từ trong tập huấn luyện
- B. Để tạo ra các câu hoàn chỉnh và rõ nghĩa hơn
- C. Để làm giảm độ phức tạp và tập trung vào các từ mang nội dung quan trọng
- D. Để giữ lại tất cả các từ giúp cải thiện độ chính xác

Câu 100. Cho các tham số của mô hình SVM (support vector machine) đã huấn luyện và bảng dữ liệu dưới đây, dự đoán cho chỉ số 0 là gì?

Vector trọng số: $w = [2, -3]$ Độ lệch: $b = 1$

Chỉ số	X1	X2
0	1	2
1	-1	-1
2	-1	2
3	4	5

- A. không xác định được
- B. không phân loại
- C. +1
- D. -1

----HẾT---