

(Sinh viên không được sử dụng tài liệu)

HỌ VÀ TÊN SV:
MSSV:
STT:
PHÒNG THI:

CÁN BỘ COI THI

CÂU HỎI TỰ LUẬN

Câu 1 (2 điểm) (G2)

Hãy trình bày hai phương pháp đánh giá mô hình máy học **train-test split** và **K-fold cross validation**. Cách trình bày tốt là cách có hình ảnh minh họa, số liệu cụ thể.

So sánh hai phương pháp này bằng cách đánh giá các ưu, nhược điểm và tính huống sử dụng của từng phương pháp. Cách trình bày tốt là kê bảng đối sánh 2 phương pháp.

Câu 2 (3 điểm) (G2)

a) Theo bạn các loại dữ liệu ảnh, văn bản, âm thanh là dữ liệu có cấu trúc hay không có cấu trúc. Tại sao?

b) Trong mạng CNN, thông tin nào nào là **parameter (tham số)**, thông tin nào là **hyper-parameter (siêu tham số)**?

c) Theo bạn, để tune (tinh chỉnh) siêu tham số của mạng CNN thì chúng ta có thể sử dụng phương pháp nào là phù hợp.

Câu 3 (2 điểm) (G1)

Hoàn tất mã nguồn sử dụng mô hình Logistic Regression để phân loại trên tập dữ liệu Iris và tune các tham số của mô hình

```
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LogisticRegression
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Load dữ liệu
X, y = load_iris(return_X_y=True)
```

```

normalizer = MinMaxScaler()
# Sử dụng normalizer.fit_transform/transform và hàm train_test_split
# để tiên xử lý bộ dữ liệu X, y nhằm chuẩn bị dữ liệu train và test.
# Lưu ý thứ tự thực hiện nào trước? Gợi ý: khoảng 3 dòng code.

...
...
...

# Sử dụng phương pháp grid search trên hai siêu tham số C và fit_intercept.
# Sử dụng hàm fit, predict và accuracy_score để dự đoán kết quả và
# đánh giá hiệu quả của mô hình sử dụng độ đo accuracy một cách khác quan.
# Từ đó, cho biết bộ siêu tham số nào là tốt nhất.
# Gợi ý: LR = LogisticRegression(C=0.01, fit_intercept = True,
solver='liblinear')
# Gợi ý: khoảng 8-10 dòng code.

...
...
...

```

Câu 4 (3 điểm) (G3)

Bạn được yêu cầu thiết kế một giải pháp để dự đoán kết quả tập luyện của một gymer (người tập thể hình). Trong đó, thông tin đầu vào sẽ bao gồm:

- cân nặng hiện tại (float),
- tỉ lệ mỡ hiện tại (float),
- tỉ lệ cơ hiện tại (float),
- nghề nghiệp (string)
- loại bài tập (string). Ví dụ: kéo xà, hít đất, bài tập squat
- cường độ luyện tập (string): nặng, nhẹ, trung bình

Hình trình bày quy trình xây dựng mô hình để thuyết phục khách hàng cho phép thực hiện dự án này. Lưu ý cần: chứng minh tính hợp lý và hiệu quả của mô hình, bắt đầu từ khâu thu thập dữ liệu đến đánh giá độ chính xác của mô hình.