



I. Machine Learning (ML) là gì?

- Machine Learning là gì?
 - Một loại robot thông minh
 - Công nghệ giúp máy học từ dữ liệu mà không cần lập trình tường minh
 - Phần mềm xử lý văn bản
 - Công cụ thiết kế đồ họa
- Thành phần cơ bản của ML là gì?
 - Dữ liệu, mô hình, thuật toán
 - Chip xử lý và RAM
 - Màn hình và bàn phím
 - Website và email
- Mục tiêu của Machine Learning là gì?
 - Tăng tốc độ Internet
 - Tự động hóa việc ra quyết định từ dữ liệu
 - Vẽ biểu đồ nhanh hơn
 - Lưu trữ dữ liệu hiệu quả
- ML KHÔNG thể áp dụng cho loại dữ liệu nào sau đây?
 - Văn bản
 - Hình ảnh
 - Âm thanh
 - Không có dữ liệu
- Trong ML, “mô hình” là gì?
 - Một bảng tính Excel
 - Một đoạn phim hoạt hình
 - Một hàm số học từ dữ liệu để đưa ra dự đoán
 - Một người mẫu thời trang



II. Tại sao cần Machine Learning?

- ML quan trọng vì:
 - Giúp máy tính học từ dữ liệu và đưa ra quyết định tốt hơn
 - Giúp tăng thời lượng pin điện thoại
 - Giúp in tài liệu nhanh hơn
 - Tạo giao diện phần mềm đẹp hơn
- Một lý do chính khiến ML trở nên phổ biến là:
 - Máy tính rẻ hơn
 - Lượng dữ liệu lớn và rẻ
 - Mạng Internet yếu

- D. Nhiều người muốn làm AI
8. Trong thực tế, ML thường được dùng để:
- A. Chơi game giải trí
 - B. Phân tích cảm xúc khách hàng qua bình luận
 - C. Thiết kế banner
 - D. Lập trình robot đánh cờ
9. ML phù hợp với những vấn đề:
- A. Có quy tắc rõ ràng
 - B. Không thể diễn tả bằng quy tắc tường minh
 - C. Cần tính toán số học cơ bản
 - D. Có thể giải bằng giấy bút
10. ML giúp giảm chi phí vì:
- A. Loại bỏ hoàn toàn con người
 - B. Không cần sử dụng máy tính
 - C. Tự động hóa quy trình, giảm công sức thủ công
 - D. Tạo ra phần mềm miễn phí

III. Các nhóm bài toán trong ML

11. ML có mấy nhóm chính?
- A. 1
 - B. 2
 - C. 3
 - D. 4
12. Phân loại là bài toán thuộc nhóm nào?
- A. Reinforcement Learning
 - B. Supervised Learning
 - C. Unsupervised Learning
 - D. Clustering
13. Bài toán phân cụm (clustering) thuộc nhóm:
- A. Supervised Learning
 - B. Reinforcement Learning
 - C. Unsupervised Learning
 - D. Tăng cường
14. Trong Reinforcement Learning, tác nhân (agent) học như thế nào?
- A. Từ dữ liệu được gán nhãn
 - B. Từ phản hồi (thưởng/phạt) trong môi trường
 - C. Từ bản đồ
 - D. Từ bộ câu hỏi trắc nghiệm

15. Bài toán dự đoán giá nhà theo diện tích là:
- A. Phân loại
 - B. Phân cụm
 - C. Hồi quy (Regression)
 - D. Tăng cường

IV. Xây dựng mô hình ML

16. Bước đầu tiên khi xây dựng ML model là:
- A. Chọn thuật toán
 - B. Huấn luyện mô hình
 - C. Thu thập và hiểu dữ liệu
 - D. Kiểm tra hiệu suất
17. Sau khi chia dữ liệu, ta thường có:
- A. Dữ liệu âm thanh
 - B. Dữ liệu học và dữ liệu kiểm tra
 - C. Dữ liệu video
 - D. Mô hình đã huấn luyện xong
18. Mục tiêu của quá trình huấn luyện mô hình là gì?
- A. Tăng tốc độ xử lý
 - B. Tạo ra file PowerPoint
 - C. Tìm ra hàm dự đoán tốt nhất từ dữ liệu
 - D. Vẽ biểu đồ
19. Sau khi huấn luyện, ta cần làm gì?
- A. Lưu mô hình lại
 - B. In kết quả
 - C. Xóa dữ liệu
 - D. Tạo slide
20. Nếu mô hình có kết quả kém trên tập test, ta nên:
- A. Xóa mô hình
 - B. Kiểm tra overfitting
 - C. Dừng học luôn
 - D. Nén dữ liệu lại

V. Overfitting và Underfitting

21. Overfitting là khi mô hình:
- A. Học rất tốt cả dữ liệu mới
 - B. Học thuộc dữ liệu huấn luyện nhưng đoán sai dữ liệu mới
 - C. Học quá nhanh

- D. Không học được gì
22. Underfitting xảy ra khi:
- A. Mô hình quá phức tạp
 - B. Dữ liệu thiếu chất lượng
 - C. Mô hình quá đơn giản, không học được quy luật
 - D. Mô hình bị virus
23. Dấu hiệu của overfitting là:
- A. Accuracy cao trên tập train, thấp trên test
 - B. Accuracy thấp trên cả train và test
 - C. Thời gian huấn luyện lâu
 - D. Mô hình không chạy
24. Giải pháp cho overfitting là:
- A. Bỏ hết dữ liệu
 - B. Dùng mô hình đơn giản hơn, regularization, hoặc nhiều dữ liệu hơn
 - C. Thêm RAM
 - D. Tăng batch size vô hạn
25. Mô hình lý tưởng là:
- A. Underfitting nhẹ
 - B. Overfitting nhẹ
 - C. Cân bằng giữa độ chính xác với tập huấn luyện và tập kiểm tra
 - D. Chạy nhanh nhất
-



VI. Tiền xử lý dữ liệu

26. Tiền xử lý dữ liệu là gì?
- A. Chỉnh sửa hình ảnh
 - B. Chuẩn bị dữ liệu sạch, đúng định dạng trước khi huấn luyện
 - C. Tạo giao diện cho dữ liệu
 - D. Mã hóa mạng
27. Kỹ thuật nào là tiền xử lý phổ biến?
- A. Gộp mô hình
 - B. Loại bỏ dữ liệu trùng
 - C. Tăng tần số mẫu
 - D. Viết code frontend
28. Xử lý dữ liệu bị thiếu thường dùng:
- A. Bỏ qua hoặc thay thế bằng trung bình
 - B. In dữ liệu
 - C. Sắp xếp lại ảnh

D. Dịch văn bản

29. Chuẩn hóa (Normalization) là gì?

A. Giảm chất lượng ảnh

B. Chuyển dữ liệu về cùng thang đo

C. Mã hóa ảnh

D. Làm tròn dữ liệu

30. Kỹ thuật giúp xử lý dữ liệu dạng chữ (text) là gì?

A. Normalization

B. Tokenization

C. Visualization

D. Classification



VII. Đo lường hiệu suất

31. Accuracy là gì?

A. Thời gian chạy

B. Số mẫu đúng chia tổng số mẫu

C. Mức độ phức tạp

D. Chi phí huấn luyện

32. Precision là gì?

A. Dự đoán đúng trong số tất cả dự đoán dương

B. Tổng số mẫu đúng

C. Số mẫu âm bị nhầm

D. Tốc độ dự đoán

33. Recall là gì?

A. Khả năng gọi lại dữ liệu

B. Dự đoán đúng trong số thực tế là dương

C. Số lỗi chia tổng số

D. Dự đoán âm

34. F1-score dùng để:

A. Tính trung bình cộng giữa precision và recall

B. Đo tốc độ

C. Hiển thị biểu đồ

D. Tính giá trị trung bình

35. Chỉ số nào phù hợp khi dữ liệu mất cân bằng?

A. Accuracy

B. F1-score

C. Loss

D. Time

