

1. Mục tiêu cốt lõi của học máy là gì?
 - A. Viết tay tất cả quy tắc ra mã
 - B. Học quy luật từ dữ liệu để dự đoán/ra quyết định
 - C. Chỉ tối ưu tốc độ xử lý
 - D. Chỉ lưu trữ dữ liệu
2. Trong mô tả nhiệm vụ học máy, “hiệu năng P” thường là:
 - A. Số lớp của mô hình
 - B. Tỷ lệ dự đoán đúng/độ lỗi định lượng
 - C. Dung lượng đĩa
 - D. Tốc độ mạng
3. Dự đoán rủi ro hồ sơ vay theo đặc trưng đầu vào là ví dụ của:
 - A. Gom cụm
 - B. Phân loại có giám sát
 - C. Hồi quy thời gian
 - D. Tăng cường
4. Bước nào thuộc tiền xử lý dữ liệu văn bản?
 - A. Chia tập kiểm thử
 - B. Chuẩn hoá hoa-thường/loại ký hiệu/giải mã viết tắt
 - C. Chọn thuật toán tối ưu
 - D. Triển khai mô hình
5. Chia train / validation / test có mục đích chính là:
 - A. Tăng kích thước mô hình
 - B. Huấn luyện / chọn mô hình / báo cáo kết quả
 - C. Nén dữ liệu
 - D. Gán nhãn dữ liệu
6. Học không giám sát thường dùng để:
 - A. Dự đoán nhãn có sẵn
 - B. Tìm cấu trúc/nhóm trong dữ liệu chưa gán nhãn
 - C. Tối ưu hàm phân thưởng
 - D. Tối ưu tốc độ suy luận
7. Học có giám sát được mô tả là tìm một hàm:
 - A. Tối đa entropy
 - B. Dự đoán đầu ra liên quan đến đầu vào mới
 - C. Giảm số chiều
 - D. Tạo dữ liệu mới
8. Học tăng cường mô hình hoá tác nhân tương tác với môi trường để:

- A. Cực tiểu lỗi huấn luyện
- B. Tối đa phần thưởng tích lũy
- C. Giảm số tham số
- D. Loại bỏ đặc trưng

9. Trong thành phần của một bài toán học máy, “hàm mục tiêu F” ánh xạ:

- A. Từ lớp sang đặc trưng
- B. Từ không gian đầu vào sang nhãn/lớp
- C. Từ trọng số sang lỗi
- D. Từ dữ liệu sang siêu tham số

10. Một lựa chọn biểu diễn cho F là:

- A. Bộ nhớ đệm
- B. Cây quyết định / tập quy tắc / đa thức / mạng nơ-ron
- C. Ổ đĩa SSD
- D. Bộ lập lịch CPU

11. Mục “tiền xử lý” trong tiếng Việt cần chú ý:

- A. Không có từ ghép
- B. Xử lý từ đơn/từ phức cho phù hợp
- C. Luôn bỏ dấu
- D. Chỉ dùng ký tự

12. Các môn học tiên quyết bao gồm?

- A. Hình học xạ ảnh
- B. Đại số tuyến tính, Xác suất thống kê, Giải tích vector, Lập trình Python
- C. Vật lý lượng tử
- D. Kinh tế vĩ mô

13. Trong mô tả nhiệm vụ theo khuôn T–E–P, “E” là:

- A. Tốc độ xử lý
- B. Dữ liệu/kinh nghiệm dùng để học
- C. Số tham số
- D. Dung lượng bộ nhớ

14. Nhiệm vụ “phân loại ký tự viết tay” là:

- A. Không giám sát
- B. Tăng cường
- C. Có giám sát
- D. Chuyển tiếp

15. Quy trình “tập dữ liệu → tiền xử lý → train/val/test” nhằm:

- A. Cho phép triển khai trực tiếp
- B. Tối ưu chọn mô hình & đánh giá công bằng
- C. Tăng dung lượng lưu trữ
- D. Thay thế mô hình

16. Một vấn đề dữ liệu thường gặp là:

- A. Thiếu GPU
- B. Nhiều/giá trị thiếu/mất cân bằng
- C. Trình biên dịch lỗi
- D. Hệ điều hành

17. Câu hỏi thuộc thuật toán học tập thường là:

- A. Chọn hệ điều hành
- B. Khi nào hội tụ? Thuật toán nào hợp điều kiện cụ thể?
- C. Chọn đĩa lưu trữ
- D. Tăng RAM

18. Học không giám sát khác học có giám sát ở điểm:

- A. Không cần dữ liệu
- B. Không có nhãn ở tập huấn luyện
- C. Dự đoán liên tục
- D. Không cần mô hình

19. Trong tiền xử lý, điền giá trị thiếu nhằm:

- A. Tăng số feature
- B. Giảm rò rỉ thông tin/ổn định học
- C. Tăng kích thước ảnh
- D. Tối ưu tốc độ GPU

20. Một biểu diễn khả dĩ cho F không phải là:

- A. Đa thức
- B. Ổ đĩa lưu trữ
- C. Cây quyết định
- D. Bộ quy tắc

21. Học tăng cường thường được mô hình hoá bằng khung:

- A. Đồ thị nhân
- B. MDP (trạng thái–hành động–phần thưởng–chuyển trạng thái)
- C. Sơ đồ quan hệ

D. Lược đồ dữ liệu

22. “Học nửa giám sát” được đề cập trong Chương 1 như:

- A. Bỏ qua vì hiếm dùng
- B. Một kiểu học bên cạnh có/không giám sát
- C. Thuật toán tối ưu
- D. Bước tiền xử lý

23. “Học chuyển tiếp” hữu ích khi:

- A. Dữ liệu mục tiêu rất lớn
- B. Dữ liệu mục tiêu hạn chế nhưng liên quan tới nhiệm vụ đã học
- C. Không có mô hình nguồn
- D. Chỉ dùng cho ảnh

24. Một kết quả học tập kỳ vọng sau học phần là khả năng:

- A. Thay mainboard
- B. Mô hình hoá vấn đề, chọn kỹ thuật, đánh giá định lượng
- C. Viết driver nhân hệ điều hành
- D. Thiết kế CPU

25. Công cụ thường dùng:

- A. Trình chỉnh sửa ảnh
- B. Thư viện học máy phổ biến (vd. sklearn, PyTorch, TensorFlow, OpenCV)
- C. Trình duyệt
- D. Hệ quản trị tệp

26. Một “cạm bẫy” xã hội–đạo đức khi áp dụng ML là:

- A. Tăng bộ nhớ
- B. Rủi ro riêng tư/giả mạo nội dung/ tai nạn ngoài ý muốn
- C. Tốc độ mạng chậm
- D. Hết pin

27. Trong “các thành phần bài toán ML”, ngoài dữ liệu còn có:

- A. Trình điều khiển thiết bị
- B. Hàm mục tiêu, cách biểu diễn, thuật toán học
- C. Module hiển thị
- D. Bộ nhớ ảo

28. “Ví dụ đào tạo” cần tương thích với:

- A. Hệ điều hành
- B. Tình huống sử dụng trong tương lai
- C. Ổ cứng
- D. Cạc mạng

29. Mục tiêu tối ưu hoá trong khung tăng cường là:

- A. Tối đa số lần thăm trạng thái
- B. Tối đa tổng phần thưởng tích lũy dài hạn
- C. Tối thiểu số hành động
- D. Giảm số trạng thái

30. Với bài toán không giám sát, một tiêu chí đánh giá điển hình là:

- A. Độ chính xác nhân thật
- B. Độ cô kết/độ tách giữa các nhóm
- C. MSE trên nhân thật
- D. Tốc độ GPU

31. Khi tiền xử lý văn bản tiếng Việt, điều gì có thể gây lỗi tách từ?

- A. Ảnh mờ
- B. Từ phức cần tách–gộp đúng ngữ cảnh
- C. Không có dấu câu
- D. Dung lượng file nhỏ

32. Một vấn đề trong quy trình học là:

- A. Chọn vỏ máy
- B. Chiến lược thứ tự dừng ví dụ huấn luyện
- C. Chọn tản nhiệt
- D. Giá linh kiện

33. Câu mô tả đúng về biểu diễn F:

- A. Chỉ có một dạng biểu diễn
- B. Có thể là cây/quy tắc/đa thức/mạng nơ-ron
- C. Phụ thuộc phân cứng
- D. Luôn tuyến tính

34. Chọn phát biểu không đúng về quy trình:

- A. Tiền xử lý trước huấn luyện
- B. Chia tập để tránh thiên lệch đánh giá
- C. Gộp test vào train để tăng dữ liệu
- D. Dùng validation để chọn mô hình

35. Khi ví dụ huấn luyện bị mất cân bằng, điều gì có thể xảy ra?

- A. Mô hình luôn tổng quát tốt
- B. Hiệu năng ảo cao ở lớp lớn, kém ở lớp hiếm
- C. Không ảnh hưởng
- D. Không thể huấn luyện

36. Trong khung thành phần bài toán, “thuật toán học” có nhiệm vụ:

- A. Lưu trữ dữ liệu
- B. Xấp xỉ/học hàm mục tiêu F từ dữ liệu
- C. Cấp phát bộ nhớ
- D. Nén mô hình

37. Một mốc lịch sử được nêu nói chung về các họ mô hình là:

- A. Chỉ còn mô hình tuyến tính
- B. Các họ như SVM, mạng tích chập, mạng hồi quy...
- C. Chỉ các mô hình đồ thị
- D. Chỉ mô hình cây

38. Trong Chương 1, kiến thức hỗ trợ bao gồm:

- A. Thiết kế mạch số
- B. Đại số tuyến tính, xác suất-thống kê, giải tích vector, lập trình
- C. Tối ưu phân cứng
- D. Quản trị dự án

39. Khi mô hình hoá bài toán thực tế, nên:

- A. Bỏ qua kiến thức miền
- B. Tận dụng kiến thức miền để đơn giản hoá & định hình bài toán
- C. Chỉ tăng kích thước mô hình
- D. Bỏ qua dữ liệu

40. “Học nửa giám sát” phù hợp nhất khi:

- A. Không có dữ liệu
- B. Ít nhãn nhưng nhiều dữ liệu chưa gán nhãn
- C. Dư thừa nhãn
- D. Chỉ dùng cho ảnh