# Lý thuyết Bài 25: Làm quen với Học máy

**Lý thuyết Tin học 12 Bài 25: Làm quen với Học máy**  
**1. Tìm hiểu sơ lược về học máy**  
Học máy là lĩnh vực của trí tuệ nhân tạo (AI) phát triển thuật toán cho phép máy tính tự học từ dữ liệu để dự đoán hoặc quyết định mà không cần lập trình rõ ràng.  
Máy tính tự học từ dữ liệu, không cần lập trình chi tiết, giúp giải quyết các bài toán phức tạp và cải thiện khả năng phân loại và nhận diện.  
Học máy là một lĩnh vực của Al nghiên cứu và phát triển các thuật toán và mô hình đem lại khả năng học cho máy tính. Nó cho phép máy tính tự động tìm hiểu từ dữ liệu và tạo ra các mô hình dự đoán hoặc quyết định dựa trên dữ liệu mà không cần phải được lập trình cụ thể.  
**2. Phân loại và vai trò của học máy trong thực tế**  
a) Phân loại học máy:  
Trong Học máy, có hai loại dữ liệu đầu vào chính: dữ liệu có nhãn\*và dữ liệu không có nhãn.  
- Dữ liệu có nhãn: Dữ liệu đã được gán nhãn hoặc giá trị đích cụ thể. Ví dụ, hình ảnh đã được gán nhãn là "con ngựa" hoặc "không phải ngựa". Dữ liệu có nhãn quan trọng trong quá trình huấn luyện mô hình học máy vì nó giúp mô hình học và đưa ra dự đoán chính xác trên dữ liệu mới.  
- Dữ liệu không có nhãn: Dữ liệu không có nhãn hoặc giá trị đích cụ thể.  
Hai phương pháp học máy cơ bản tương ứng với hai loại dữ liệu đầu vào là:  
1.Học có giám sát: Sử dụng dữ liệu có nhãn để huấn luyện mô hình. Mô hình học mối quan hệ giữa đầu vào và đầu ra để dự đoán nhãn cho dữ liệu mới. Ví dụ: lọc thư rác, nhận dạng hình ảnh, nhận dạng chữ viết tay.  
Học có giám sát là phương pháp phổ biến nhất và có nhiều ứng dụng thực tế.  
2.Học không giám sát**:** là phương pháp học máy sử dụng dữ liệu không có nhãn. Phương pháp này giúp mô hình hóa cấu trúc hoặc thông tin ẩn trong dữ liệu dựa trên mối quan hệ tương tự hoặc khác biệt và xác suất đồng xuất hiện của các đối tượng.  
Ứng dụng của học không giám sát bao gồm:  
- Phân chia dữ liệu thành các nhóm  
- Xác định phân khúc khách hàng: Dựa trên lịch sử mua hàng để phân loại nhóm khách hàng.  
- Phát hiện bất thường.  
- Xác định chủ đề  
Học không giám sát giúp khai thác thông tin ẩn và phân nhóm dữ liệu mà không cần nhãn trước.  
b) Vai trò của học máy:  
Học máy đóng vai trò quan trọng trong nhiều ứng dụng thực tế nhờ khả năng khai thác và phân tích dữ liệu lớn và đa dạng. Dưới đây là một số ví dụ tiêu biểu:  
- Lọc thư rác: Học máy giúp phân loại thư điện tử thành thư rác hoặc thư thường dựa trên các đặc điểm của thư, cải thiện hiệu suất lọc qua việc học từ dữ liệu và cập nhật mô hình.  
 - Chẩn đoán bệnh: Học máy phân tích dữ liệu sức khoẻ và xét nghiệm để xây dựng mô hình chẩn đoán và dự báo tình trạng sức khoẻ, hỗ trợ bác sĩ đưa ra quyết định nhanh chóng và chính xác.  
- Phân tích thị trường\*\*: Học máy phân tích dữ liệu thị trường để xác định xu hướng, dự báo biến động giá cả và hỗ trợ chiến lược kinh doanh, giúp nhà đầu tư đưa ra quyết định dựa trên thông tin và phân tích kỹ thuật.  
- Nhận dạng tiếng nói: Học máy xây dựng mô hình âm thanh để nhận dạng và phân biệt âm thanh của các đơn vị tiếng, cải thiện khả năng nhận diện tiếng nói của người dùng khác nhau.  
- Nhận dạng chữ viết: Học máy sử dụng mô hình hình học để nhận diện chữ viết tay, với sự phát triển của học sâu giúp cải thiện đáng kể khả năng nhận diện chữ viết.  
-Dịch tự động: Học máy xây dựng mô hình dịch tự động từ dữ liệu bản dịch và bản gốc, giúp dịch văn bản và tiếng nói giữa các ngôn ngữ, giảm rào cản ngôn ngữ trong giao tiếp và phát triển hợp tác.  
Học máy giúp xử lý và phân tích dữ liệu lớn, tự động hoá các nhiệm vụ phức tạp, và không ngừng cập nhật mô hình để phản ứng với sự thay đổi của dữ liệu.  
->Hai phương pháp học máy cơ bản là học có giám sát và học không giám sát, tuỳ theo tập dữ liệu cung cấp cho mô hình học máy là dữ liệu có nhãn hay không có nhãn. Học máy giúp xử lí lượng lớn dữ liệu một cách nhanh chóng và hiệu quả, bao gồm cả các dữ liệu không ngừng thay đổi theo thời gian, trợ giúp các quá trình ra quyết định cũng như tự động hoá các nhiệm vụ phức tạp.  
**Trắc nghiệm Tin học 12 Bài 25: Làm quen với Học máy**  
**PHẦN I.** **Câu trắc nghiệm nhiều phương án lựa chọn. Thí sinh trả lời từ câu 1 đến câu 10. Mỗi câu hỏi thí sinh chỉ lựa chọn một phương án.**  
**Câu 1:** Học máy thuộc lĩnh vực nào trong trí tuệ nhân tạo?  
A. Xử lý ngôn ngữ tự nhiên  
B. Học sâu  
C. Học máy  
D. Tìm kiếm thông tin  
**Hiển thị đáp án**  
  
**Đáp án:** C  
**Giải thích:** Học máy là một lĩnh vực của trí tuệ nhân tạo (AI) nghiên cứu và phát triển các thuật toán giúp máy tính tự học từ dữ liệu để dự đoán hoặc quyết định mà không cần lập trình cụ thể.  
  
  
**Câu 2:** Trong học máy, loại dữ liệu nào được gán nhãn cụ thể để huấn luyện mô hình?  
A. Dữ liệu không có nhãn  
B. Dữ liệu không rõ  
C. Dữ liệu có nhãn  
 D. Dữ liệu phụ  
**Hiển thị đáp án**  
  
**Đáp án:** C  
**Giải thích:** Dữ liệu có nhãn là dữ liệu đã được gán nhãn hoặc giá trị đích cụ thể, rất quan trọng trong quá trình huấn luyện mô hình học máy vì nó giúp mô hình học và dự đoán chính xác hơn.  
  
  
**Câu 3:** Phương pháp học máy nào sử dụng dữ liệu không có nhãn để mô hình hóa cấu trúc hoặc thông tin ẩn?  
A. Học có giám sát  
B. Học không giám sát  
C. Học sâu  
D. Học tăng cường  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Học không giám sát sử dụng dữ liệu không có nhãn để khai thác thông tin ẩn, phân nhóm dữ liệu, và tìm hiểu cấu trúc nội tại của dữ liệu mà không cần nhãn trước.  
  
  
**Câu 4:** Một ứng dụng của học máy trong phân tích thị trường là gì?  
A. Lọc thư rác  
B. Nhận dạng chữ viết  
C. Dự báo biến động giá cả  
D. Chẩn đoán bệnh  
**Hiển thị đáp án**  
  
**Đáp án:** C  
**Giải thích:** Học máy phân tích dữ liệu thị trường để xác định xu hướng, dự báo biến động giá cả và hỗ trợ chiến lược kinh doanh.  
  
  
**Câu 5:** Trong các phương pháp học máy, cái nào sử dụng dữ liệu có nhãn để huấn luyện mô hình?  
A. Học không giám sát  
B. Học có giám sát  
C. Học tăng cường  
D. Học sâu  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Học có giám sát sử dụng dữ liệu có nhãn để huấn luyện mô hình, giúp mô hình học mối quan hệ giữa đầu vào và đầu ra để dự đoán nhãn cho dữ liệu mới.  
  
  
**Câu 6:** Công nghệ học máy nào được sử dụng để nhận diện và phân biệt âm thanh của các đơn vị tiếng?  
A. Nhận dạng chữ viết  
B. Dịch tự động  
C. Nhận dạng tiếng nói  
D. Phân tích thị trường  
**Hiển thị đáp án**  
  
**Đáp án:** C  
**Giải thích:** Học máy xây dựng mô hình âm thanh để nhận diện và phân biệt âm thanh của các đơn vị tiếng, giúp cải thiện khả năng nhận diện tiếng nói của người dùng khác nhau.  
  
  
**Câu 7:** Học máy có thể giúp cải thiện khả năng phân loại và nhận diện trong lĩnh vực nào sau đây?  
A. Quản lý dự án  
B. Dịch tự động  
C. Phát triển phần mềm  
D. Lập trình cơ sở dữ liệu  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Học máy xây dựng mô hình dịch tự động từ dữ liệu bản dịch và bản gốc, giúp dịch văn bản và tiếng nói giữa các ngôn ngữ, giảm rào cản ngôn ngữ trong giao tiếp.  
  
  
**Câu 8:** Để xây dựng một mô hình chẩn đoán bệnh, học máy sử dụng loại dữ liệu nào?  
A. Dữ liệu không có nhãn  
B. Dữ liệu có nhãn  
C. Dữ liệu giả lập  
D. Dữ liệu tham khảo  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Để xây dựng mô hình chẩn đoán bệnh, học máy thường sử dụng dữ liệu có nhãn từ các xét nghiệm và dữ liệu sức khoẻ để dự đoán và chẩn đoán tình trạng sức khoẻ.  
  
  
**Câu 9:** Học máy giúp xử lý và phân tích dữ liệu lớn như thế nào?  
A. Tăng cường độ chính xác  
B. Tự động hoá các nhiệm vụ phức tạp  
C. Giảm kích thước dữ liệu  
D. Cải thiện tốc độ xử lý phần cứng  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Học máy giúp tự động hoá các nhiệm vụ phức tạp bằng cách xử lý và phân tích lượng lớn dữ liệu, tạo ra các mô hình dự đoán và quyết định dựa trên dữ liệu.  
  
  
**Câu 10:** Để cải thiện khả năng nhận diện chữ viết tay, học máy sử dụng mô hình nào?  
A. Mô hình phân tích thị trường  
B. Mô hình hình học  
C. Mô hình phân nhóm  
D. Mô hình nhận dạng tiếng nói  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Học máy sử dụng mô hình hình học để nhận diện chữ viết tay, với sự phát triển của học sâu giúp cải thiện khả năng nhận diện chữ viết tay một cách chính xác.  
  
  
**PHẦN II. Câu trắc nghiệm đúng sai.** **Thí sinh trả lời từ câu 1 đến câu 2. Trong mỗi ý a), b), c), d) ở mỗi câu, thí sinh chọn đúng hoặc sai**  
**Câu 1:** Phát biểu sau đúng hay sai về học máy?  
a) Học máy giúp máy tính tự xây dựng bộ lọc để phân loại thư điện tử mà không cần lập trình chi tiết.  
b) Học máy chỉ hoạt động tốt với dữ liệu đã được lập trình cụ thể.  
c) Một trong những ứng dụng của học có giám sát là nhận dạng hình ảnh.  
d) Học không giám sát là phương pháp phổ biến nhất trong học máy.  
**Hiển thị đáp án**  
  
a) Đúng: Học máy cho phép máy tính tự học từ dữ liệu để xây dựng bộ lọc phân loại thư điện tử, thay vì phải lập trình cụ thể cho từng loại thư.  
b) Sai: Học máy có thể học từ dữ liệu mà không cần lập trình cụ thể, tạo ra các mô hình dự đoán hoặc quyết định dựa trên dữ liệu mà không cần các quy tắc rõ ràng.  
c) Đúng: Học có giám sát sử dụng dữ liệu có nhãn để huấn luyện mô hình, bao gồm các ứng dụng như nhận dạng hình ảnh, nơi mà dữ liệu đầu vào có nhãn là hình ảnh và nhãn là danh mục tương ứng.  
d) Sai: Học có giám sát là phương pháp phổ biến nhất trong học máy, do nó sử dụng dữ liệu có nhãn để huấn luyện mô hình, dễ dàng đánh giá và đo lường hiệu quả.  
  
  
**Câu 2:** Xác định tính đúng/sai của các phát biểu sau về các ứng dụng của học máy:  
a) Học máy giúp phân loại thư điện tử thành thư rác hoặc thư thường dựa trên các đặc điểm của thư đúng hay sai?  
b) Học máy không thể được sử dụng trong chẩn đoán bệnh.  
c) Học máy không có khả năng xử lý và phân tích dữ liệu lớn.  
d) Một trong những ứng dụng của học máy là dịch tự động từ ngôn ngữ này sang ngôn ngữ khác.  
**Hiển thị đáp án**  
  
a) Đúng: Học máy có khả năng phân loại thư điện tử thành thư rác hoặc thư thường bằng cách học từ các đặc điểm của thư trong quá trình huấn luyện.  
b) Sai: Học máy được sử dụng trong chẩn đoán bệnh bằng cách phân tích dữ liệu sức khỏe và xét nghiệm, hỗ trợ bác sĩ trong việc đưa ra các quyết định chính xác.  
c) Sai: Học máy có khả năng xử lý và phân tích lượng lớn dữ liệu, tự động hóa các nhiệm vụ phức tạp và liên tục cập nhật mô hình để đáp ứng sự thay đổi của dữ liệu.  
d) Đúng: Học máy được sử dụng để xây dựng các mô hình dịch tự động, giúp dịch văn bản và tiếng nói giữa các ngôn ngữ khác nhau, giảm rào cản ngôn ngữ trong giao tiếp.  
  
  
**PHẦN III. Câu trả lời ngắn**. **Thí sinh trả lời từ câu 1 đến câu 3**  
**Câu 1:** Học máy (Machine Learning) là gì và nó hoạt động như thế nào?  
**Hiển thị đáp án**  
  
**Đáp án:** Học máy là một lĩnh vực của trí tuệ nhân tạo (AI) chuyên nghiên cứu và phát triển các thuật toán và mô hình cho phép máy tính tự học từ dữ liệu mà không cần lập trình chi tiết. Máy tính sử dụng các thuật toán này để phân tích dữ liệu, nhận diện mẫu, và tạo ra các dự đoán hoặc quyết định dựa trên dữ liệu đã học được**.**  
**Giải thích:**Học máy hoạt động bằng cách sử dụng dữ liệu để huấn luyện các mô hình, sau đó áp dụng các mô hình này để phân tích và đưa ra dự đoán cho các dữ liệu mới. Nó giúp máy tính tự động hóa các nhiệm vụ phức tạp mà con người thường làm, như phân loại thư điện tử, chẩn đoán bệnh, và dịch ngôn ngữ.  
  
  
**Câu 2:** Phân biệt giữa học có giám sát và học không giám sát trong học máy?  
**Hiển thị đáp án**  
  
**Đáp án: -** Học có giám sát: Sử dụng dữ liệu có nhãn (được gán giá trị cụ thể) để huấn luyện mô hình. Mô hình học cách dự đoán nhãn cho dữ liệu mới dựa trên mối quan hệ giữa đầu vào và đầu ra từ dữ liệu huấn luyện.  
- Học không giám sát: Sử dụng dữ liệu không có nhãn. Mô hình tìm hiểu cấu trúc hoặc mẫu ẩn trong dữ liệu mà không có giá trị đích cụ thể, thường để phân chia dữ liệu thành các nhóm hoặc xác định sự bất thường.  
**Giải thích:** Học có giám sát yêu cầu dữ liệu phải có nhãn để mô hình có thể học cách phân loại hoặc dự đoán chính xác. Trong khi đó, học không giám sát không yêu cầu nhãn và được sử dụng để tìm hiểu các cấu trúc ẩn trong dữ liệu, như phân cụm khách hàng hoặc phát hiện sự bất thường.  
  
  
**Câu 3:** Nêu vai trò của học máy trong các ứng dụng thực tế?  
**Hiển thị đáp án**  
  
**Đáp án:** Học máy đóng vai trò quan trọng trong nhiều ứng dụng thực tế, bao gồm:  
Lọc thư rác: Phân loại thư điện tử thành thư rác hoặc thư thường.  
Chẩn đoán bệnh: Phân tích dữ liệu y tế để hỗ trợ chẩn đoán và điều trị.  
Phân tích thị trường: Dự báo xu hướng và biến động giá cả.  
Nhận dạng tiếng nói: Cải thiện nhận diện tiếng nói của người dùng.  
Dịch tự động: Dịch văn bản và tiếng nói giữa các ngôn ngữ.  
**Giải thích:** Học máy giúp tự động hóa và cải thiện hiệu suất trong các lĩnh vực khác nhau nhờ khả năng phân tích dữ liệu lớn, học từ dữ liệu để tạo ra các mô hình dự đoán, và liên tục cập nhật mô hình để thích ứng với sự thay đổi của dữ liệu.  
  
  
Xem thêm