**CÂU HỎI ÔN TẬP**

**Môn học: NT522 – Phương pháp học máy trong An toàn thông tin**

**---**

**Cập nhật: 20.06.2023**

1. Phân biệt các loại AI khác nhau? Trình bày được định nghĩa của các loại học máy (Học giám sát, không giám sát, bán giám sát, học tăng cường)?
2. Các tiêu chí để đánh giá mô hình học máy? Cách chọn tiêu chí đánh giá phù hợp trong một số ngữ cảnh bài toán An toàn thông tin.
3. Cách huấn luyện mô hình, cách phân chia dữ liệu và tiền xử lí dữ liệu phổ biến?
4. Tác dụng của hàm kích hoạt (activation function) trong các mạng nơ-ron? Một số hàm activation phổ biến?
5. Phân biệt sự khác nhau của các dạng thuật toán học máy như Phân cụm (clustering), phát hiện sự dị thường (anomaly detection), phân loại (classification).
6. Học sâu (DL) là gì? Xác định được các lớp, đầu ra, đầu vào trong một số mô hình mạng nơ-ron nhân tạo, học sâu (DL)? Một số ứng dụng cơ bản trong lĩnh vực ATTT.
7. Nguyên tắc hoạt động của mạng sinh đối kháng? Một số ứng dụng của mô hình mạng sinh đối kháng (GANs) trong ATTT.
8. Học tăng cường là gì? Một số ứng dụng của Reinforcement Learning (RL) trong An toàn thông tin? Xác định được các khái niệm cơ bản, định nghĩa của RL trong các bài toán điển hình của ATTT.
9. Ảnh hưởng của tốc độ học, batch size và epoch lên quá trình huấn luyện mô hình học máy nói chung?
10. Cần làm gì khi dữ liệu bị mất cân bằng? Tiêu chí nào có thể dùng để đánh giá mô hình trong trường hợp dữ liệu bị mất cân bằng?
11. Học liên kết (Federated Learning - FL) là gì? Ưu và nhược điểm của nó? Công thức tổng hợp mô hình trong FL (thuật toán tổng hợp FedAvg)?
12. Xử lý đặc trưng (feature engineering) là gì? Mục đích áp dụng? Có những kỹ thuật nào để thực hiện?
13. MLOps là gì? Nó giải quyết vấn đề gì trong ngữ cảnh các ứng dụng có sử dụng Học máy?

TRẢ LỜI :

1. Các loại AI khác nhau:

- AI yếu (Weak AI): là một hệ thống thông minh được thiết kế để giải quyết các nhiệm vụ cụ thể và hạn chế trong phạm vi của nó. Ví dụ: trợ lý ảo, công cụ tìm kiếm, hệ thống thông tin khách hàng.

- AI mạnh (Strong AI): là một hệ thống thông minh tự học và có khả năng giải quyết các vấn đề đa dạng như con người. Hiện tại, AI mạnh vẫn chỉ ở mức độ nghiên cứu và chưa được phát triển.

- Học máy (Machine Learning): là một loại AI, có khả năng học hỏi từ dữ liệu mà không cần được lập trình cụ thể. Các loại học máy gồm: học giám sát, không giám sát, bán giám sát, và học tăng cường.

1. - Học giám sát (Supervised Learning): Đây là phương pháp học máy mà trong đó một mô hình học tập từ dữ liệu được gán nhãn. Dữ liệu được gán nhãn có nghĩa là chúng ta biết đầu ra mong muốn cho một tập dữ liệu và mô hình được đào tạo để tìm ra mối quan hệ giữa đầu vào và đầu ra. Ví dụ, nếu một bộ dữ liệu chứa các hình ảnh của mèo và chó, với nhãn tương ứng, mô hình học giám sát có thể được đào tạo để phân loại các hình ảnh đầu vào là mèo hoặc chó.
2. Học không giám sát (Unsupervised Learning): Đây là phương pháp học máy mà trong đó một mô hình học tập từ dữ liệu không được gán nhãn. Mô hình được đào tạo để phát hiện các mẫu hoặc cấu trúc trong dữ liệu mà không có thông tin đầu ra mong muốn cụ thể nào. Ví dụ, nếu chúng ta có một tập dữ liệu chứa các hình ảnh của động vật, mô hình học không giám sát có thể được đào tạo để phát hiện các nhóm động vật khác nhau trong tập dữ liệu.
3. Học bán giám sát (Semi-supervised Learning): Đây là phương pháp học máy mà trong đó một mô hình được đào tạo từ một tập dữ liệu lớn, trong đó chỉ một phần nhỏ các mẫu được gán nhãn. Mô hình được đào tạo để tìm hiểu các mẫu không được gán nhãn và phân loại chúng dựa trên thông tin của các mẫu được gán nhãn. Ví dụ, nếu chúng ta có một tập dữ liệu chứa các hình ảnh của mèo và chó, với chỉ một phần nhỏ các hình ảnh được gán nhãn, mô hình học bán giám sát có thể được đào tạo để phân loại các hình ảnh không được gán nhãn dựa trên thông tin của các hình ảnh được gán nhãn.
4. Học tăng cường (Reinforcement Learning): Đây là phương pháp học máy mà trong đó một mô hình tương tác với môi trường và học từ các phản hồi của môi trường trong quá trình tương tác đó. Mô hình được đào tạo để tìm ra các hành động tối ưu để đạt được một mục tiêu nào đó. Ví dụ, một mô hình học tăng cường có thể được đào tạo để chơi một trò chơi máy tính và học từ các phản hồi của trò chơi để tìm ra các hành động tối ưu để đạt được điểm số cao nhất.

2. Các tiêu chí để đánh giá mô hình học máy bao gồm:

1. Độ chính xác (Accuracy): Đây là tiêu chí đánh giá mức độ chính xác của mô hình, được tính bằng tỷ lệ giữa số lượng dữ liệu được dự đoán đúng và tổng số lượng dữ liệu. Tuy nhiên, tiêu chí này không phù hợp cho các bài toán có mất cân bằng dữ liệu hoặc các bài toán nhiễu.
2. Precision và Recall: Precision đánh giá tỉ lệ giữa số trường hợp dự đoán đúng và tổng số trường hợp được dự đoán là đúng, trong khi Recall đánh giá tỉ lệ giữa số trường hợp dự đoán đúng và tổng số trường hợp thực tế là đúng. Tiêu chí này phù hợp cho các bài toán mất cân bằng dữ liệu như phát hiện gian lận hoặc phát hiện virus.
3. F1-score: F1-score là trung bình điều hòa giữa Precision và Recall, giúp đánh giá mô hình học máy đối với các bài toán mất cân bằng dữ liệu.
4. AUC-ROC: AUC-ROC là diện tích dưới đường cong ROC (Receiver Operating Characteristic), đánh giá độ chính xác của mô hình dựa trên định tuyến giữa tỉ lệ True Positive Rate và False Positive Rate. Tiêu chí này thường được sử dụng trong các bài toán phân loại nhị phân.
5. MSE và RMSE: MSE (Mean Squared Error) đánh giá độ lỗi trung bình của mô hình dự đoán so với giá trị thực tế, trong khi RMSE (Root Mean Squared Error) là căn bậc hai của MSE. Tiêu chí này thường được sử dụng trong các bài toán dự đoán định lượng.

Trong bài toán An toàn thông tin, các tiêu chí đánh giá phù hợp sẽ phụ thuộc vào loại bài toán. Ví dụ, trong bài toán phát hiện xâm nhập, precision và recall sẽ là các tiêu chí quan trọng để đánh giá khả năng phát hiện xâm nhập, trong khi độ chính xác có thể không quan trọng bằng precision và recall. Trong bài toán phát hiện gian lận, precision là tiêu chí quan trọng để đánh giá khả năng phát hiện các giao dịch gian lận, trong khi recall là tiêu chí quan trọng để đánh giá khả năng phát hiện toàn bộ các giao dịch gian lận. Trong bài toán nhận dạng khuôn mặt, độ chính xác và AUC-ROC có thể là các tiêu chí quan trọng để đánh giá khả năng nhận dạng khuôn mặt đúng và giảm thiểu số lượng nhận dạng sai.

3. Cách huấn luyện mô hình, phân chia dữ liệu và tiền xử lí dữ liệu phổ biến:

- Huấn luyện mô hình: bao gồm việc chọn một kiến trúc mô hình phù hợp, tối ưu hóa các tham số và hàm mất mát, và đánh giá hiệu suất của mô hình.

- Phân chia dữ liệu: dữ liệu được chia thành hai tập: tập huấn luyện và tập kiểm tra hoặc tập xác nhận. Tập huấn luyện được sử dụng để huấn luyện mô hình, trong khi tập kiểm tra hoặc tập xác nhận được sử dụng để đánh giá hiệu suất của mô hình.

- Tiền xử lí dữ liệu: bao gồm các bước tiền xử lí như xóa các giá trị thiếu, chuẩn hóa dữ liệu, rút trích đặc trưng và mã hóa dữ liệu.

4. Tác dụng của hàm kích hoạt trong các mạng nơ-ron là để đưa ra quyết định về việc kích hoạt các đơn vị nơ-ron trong mạng. Một số hàm kích hoạt phổ biến bao gồm:

Hàm kích hoạt (activation function) là một phần quan trọng trong các mạng nơ-ron, nó được sử dụng để đưa ra quyết định cho mỗi nơ-ron và tạo ra đầu ra của mạng. Hàm kích hoạt giúp mô hình có khả năng học những mối quan hệ phi tuyến trong dữ liệu, giúp tăng độ chính xác của mô hình.

Một số hàm kích hoạt phổ biến trong các mạng nơ-ron bao gồm:

1. Hàm Sigmoid: Hàm sigmoid là một hàm phi tuyến, biểu diễn đầu ra của một nơ-ron dưới dạng một số từ 0 đến 1. Hàm sigmoid thường được sử dụng trong các mạng nơ-ron với kiến trúc cũ như Perceptron và Multi-Layer Perceptron (MLP).
2. Hàm Tanh: Tương tự như hàm sigmoid, hàm tanh cũng là một hàm phi tuyến, biểu diễn đầu ra của một nơ-ron dưới dạng một số từ -1 đến 1. Hàm tanh thường được sử dụng trong các mạng nơ-ron với kiến trúc cũ như MLP.
3. Hàm ReLU: Hàm ReLU (Rectified Linear Unit) được sử dụng rộng rãi trong các mạng nơ-ron hiện đại. Hàm này chỉ trả về giá trị dương của đầu vào và trả về 0 cho các giá trị âm. Hàm ReLU giúp giảm độ phức tạp tính toán và tăng tốc độ hội tụ của mô hình.
4. Hàm Leaky ReLU: Hàm Leaky ReLU là phiên bản cải tiến của hàm ReLU, nó giữ lại một phần độ dốc cho các giá trị âm. Hàm này giúp tránh tình trạng bị "chết ReLU" (ReLU dead) do các nơ-ron không hoạt động.
5. Hàm Softmax: Hàm softmax được sử dụng trong các mạng nơ-ron phân loại đa lớp và biểu diễn đầu ra của một nơ-ron dưới dạng một phân phối xác suất trên các lớp đầu ra.

5. Phân biệt sự khác nhau của các dạng thuật toán học máy:

- Phân cụm (clustering): tìm cách phân loại các điểm dữ liệu vào các nhóm dựa trên các đặc trưng tương tự.

- Phát hiện sự dị thường (anomaly detection): tìm cách phát hiện các điểm dữ liệu có đặc trưng khác biệt so với phần còn lại của dữ liệu.

- Phân loại (classification): tìm cách phân loại các điểm dữ liệu vào các nhóm đã được định nghĩa trước đó.

1. 6. Học sâu (Deep Learning) là một lĩnh vực của trí tuệ nhân tạo (AI) dựa trên các mô hình mạng nơ-ron nhân tạo sâu với nhiều lớp ẩn để xử lý và trích xuất thông tin từ dữ liệu đầu vào. Các mô hình học sâu được sử dụng để giải quyết các bài toán phức tạp đòi hỏi khả năng học, phân tích và tìm ra các mối quan hệ phi tuyến từ dữ liệu.

Một số ứng dụng của AI trong lĩnh vực An toàn thông tin bao gồm:

- Phát hiện xâm nhập (Intrusion Detection): sử dụng các mô hình học máy để phát hiện các hành vi xâm nhập vào hệ thống máy tính.

- Phát hiện gian lận (Fraud Detection): sử dụng các mô hình học máy để phát hiện các hành vi gian lận trong các giao dịch tài chính và thẻ tín dụng.

- Phân loại email rác (Spam Classification): sử dụng các mô hình học máy để phân loại email vào hai nhóm: email hợp lệ và email rác.

- Nhận dạng khuôn mặt (Facial Recognition): sử dụng các mô hình học sâu để nhận dạng khuôn mặt trong các hệ thống an ninh và giám sát.

- Phân tích tấn công mạng (Network Attack Analysis): sử dụng các mô hình học máy để phân tích các cuộc tấn công mạng và đưa ra các giải pháp bảo vệ hệ thống.

Một số thách thức của AI trong lĩnh vực An toàn thông tin bao gồm:

- Bảo mật dữ liệu: AI có thể được sử dụng để tấn công và đánh lừa các hệ thống bảo mật, do đó cần có các biện pháp bảo mật dữ liệu để đảm bảo tính an toàn và riêng tư.

- Giải thích mô hình: các mô hình học máy và học sâu có thể rất phức tạp và khó hiểu, do đó cần có các giải pháp để giải thích quyết định đưa ra bởi các mô hình.

- Khả năng tấn công mạng: AI có thể được sử dụng để phát triển các công cụ tấn công mạng và đánh lừa các hệ thống bảo mật, do đó cần có các giải pháp để phát hiện và ngăn chặn các tấn công này.

- Thiếu dữ liệu: các mô hình học máy và học sâu đòi hỏi một lượng lớn dữ liệu để huấn luyện chính xác, tuy nhiên trong một số trường hợp, dữ liệu không đủ để đảm bảo tính chính xác và độ tin cậy của các mô hình.

7. Mạng sinh đối kháng (Generative Adversarial Networks - GANs) là một mô hình học sâu có thể tạo ra các dữ liệu mới thông qua quá trình đối kháng giữa hai mạng: mạng sinh và mạng phân biệt. Mạng sinh được huấn luyện để tạo ra các dữ liệu giống như dữ liệu thật, trong khi mạng phân biệt được huấn luyện để phân biệt dữ liệu thật và dữ liệu giả. Qua các vòng lặp, mạng sinh và mạng phân biệt cùng tiến hóa để tạo ra các dữ liệu giả chất lượng cao. Một số ứng dụng của GANs trong an toàn thông tin bao gồm: tạo ra dữ liệu giả để đào tạo các mô hình phát hiện gian lận, tạo ra dữ liệu giả để đào tạo các mô hình phân loại email rác, và tạo ra các hình ảnh giả để kiểm tra tính an toàn của các hệ thống nhận dạng khuôn mặt.

8. Học tăng cường (Reinforcement Learning - RL) là một mô hình học máy, trong đó một tác nhân tương tác với một môi trường và học cách tối đa hóa một phần thưởng được cung cấp bởi môi trường. Một số ứng dụng của RL trong an toàn thông tin bao gồm: tìm kiếm chiến lược tốt nhất để phòng chống các cuộc tấn công mạng, tối ưu hóa các hệ thống giao thông để đảm bảo an toàn, và tìm kiếm chiến lược tốt nhất để quản lý rủi ro trong các hệ thống tài chính. Trong các bài toán điển hình của ATTT, RL được sử dụng để tối ưu hoá các chiến lược bảo mật, quản lý rủi ro và tối ưu hóa các thuật toán phát hiện xâm nhập.

9. Tốc độ học, batch size và epoch là các siêu tham số quan trọng trong quá trình huấn luyện mô hình học máy. Tốc độ học quyết định sự nhanh chậm của quá trình học, batch size quyết định số lượng dữ liệu được sử dụng trong mỗi lần cập nhật trọng số, và epoch quyết định số lần huấn luyện trên toàn bộ tập dữ liệu. Tốc độ học quá cao có thể dẫn đến hiện tượng quá khớp, trong khi tốc độ học quá thấp có thể dẫn đến quá trình học chậm hoặc không hội tụ. Batch size quá nhỏ có thể dẫn đến hiện tượng dao động, trong khi batch size quá lớn có thể dẫn đến sự mất mát tính toán và thời gian huấn luyện dài hơn. Epoch quá ít có thể dẫn đến khả năng tổng quát hóa kém, trong khi epoch quá nhiều có thể dẫn đến hiện tượng quá khớp.

10. Khi dữ liệu bị mất cân bằng, có thể sử dụng các kỹ thuật tái cân bằng dữ liệu như undersampling, oversampling hoặc kết hợp cả hai để cân bằng số lượng các lớp dữ liệu. Tiêu chí có thể được sử dụng để đánh giá mô hình trong trường hợpdữ liệu bị mất cân bằng là precision, recall và F1-score. Precision là tỷ lệ giữa số lượng dự đoán đúng và tổng số lượng dự đoán trong một lớp, recall là tỷ lệ giữa số lượng dự đoán đúng và tổng số lượng thực tế trong một lớp, và F1-score là trung bình điều hòa giữa precision và recall. Các tiêu chí này cần được sử dụng để đánh giá hiệu quả của mô hình trên cả các lớp dữ liệu thiểu số và đa số.

11. Học liên kết (Federated Learning - FL) là một mô hình học máy phân tán, trong đó các thiết bị địa phương thực hiện quá trình huấn luyện trên dữ liệu của mình và gửi thông tin trọng số được cập nhật đến một trung tâm để tổng hợp. Ưu điểm của FL là đảm bảo tính riêng tư của dữ liệu, giảm thiểu chi phí truyền dữ liệu và tăng tốc độ huấn luyện. Tuy nhiên, FL cũng có nhược điểm là cần có một cơ chế đáng tin cậy để giám sát quá trình huấn luyện và đảm bảo tính toàn vẹn của dữ liệu. Công thức tổng hợp mô hình trong FL được gọi là thuật toán tổng hợp FedAvg, trong đó trung tâm cập nhật trọng số bằng cách tính trung bình có trọng số của các trọng số được gửi từ các thiết bị địa phương.

12. Xử lý đặc trưng (feature engineering) là quá trình chọn lọc và biến đổi các đặc trưng của dữ liệu để tạo ra các đặc trưng mới có tác dụng cải thiện hiệu quả của mô hình học máy. Mục đích của việc áp dụng xử lý đặc trưng là giảm thiểu sự ảnh hưởng của đặc trưng nhiễu và cải thiện khả năng dự đoán của mô hình. Các kỹ thuật xử lý đặc trưng phổ biến bao gồm: loại bỏ các đặc trưng không quan trọng, chuẩn hóa dữ liệu và tạo ra các đặc trưng mới bằng cách kết hợp các đặc trưng hiện có.

13. MLOps là viết tắt của Machine Learning Operations, là quá trình quản lý và triển khai các hệ thống học máy. MLOps giải quyết vấn đề của việc triển khai các mô hình học máy trong các ứng dụng thực tế, bao gồm quản lý phiên bản mô hình, triển khai mô hình, giám sát hiệu suất và tự động hóa quy trình huấn luyện và triển khai mô hình. MLOps giúp cải thiện tính ổn định và tin cậy của các ứng dụng sử dụng học máy và giảm thiểu tác động tiêu cực của các lỗi liên quan đến mô hình học máy.