1. **Những vấn đề thách thức của machine learning là gì? Nêu ví dụ cho mỗi loại.**

Những vấn đề thách thức: bad data, bad algorithm

Ví dụ của bad data: thiếu hụt dữ liệu, đặc tính dữ liệu không phù hợp

Ví dụ của bad algorithm: Overfitting, Underfitting

1. **Nếu model của bạn hoạt động tốt trên dữ liệu huấn luyện nhưng kém hiệu quả trên dữ liệu mới thì hiện tượng gì có thể đã xảy ra?**

Overfitting là hiện tượng model hoạt động tốt trên dữ liệu huấn luyện nhưng kém hiệu quả trên dữ liệu mới.

1. **Tìm hiểu những cách xử lý overfitting.**

Tăng thêm dữ liệu để tăng độ phong phú đa dạng của dữ liệu

Giảm độ phức tạp của model

Use Regularization: thêm vào hàm loss 1 đại lượng lamda để tối ưu model mô hình giảm độ phức tạp.

Use Dropout: bỏ đi random p% node của layer => giúp cho mô hình bớt phức tạp

Early stoping: dừng việc training để không tốn tài nguyên

1. **Test set là gì? Tại sao cần dùng test set để đánh giá model mà không dùng training set**.

Test set: tập hợp dữ liệu được sử dụng để thử nghiệm một chương trình học máy sau khi đã được đào tạo về một bộ dữ liệu huấn luyện ban đầu

Để kiểm nghiệm được độ chính xác của chương trình, người ta dùng tập Testing set. Khác với Training set dùng để huấn luyện hệ thống, Testing set chỉ gồm các giá trị mà không có các giá trị kết quả. Máy tính sẽ nhận những giá trị này, và xử lý các giá trị, sau đó đưa ra kết quả tương ứng.

1. **Hyperparameter tuning là gì? Validation set là gì? Tại sao cần dùng nó.**

Hyperparameter tuning là quá trình chọn model tốt nhất cho quá trình training diễn ra trước training

Validation set là tập các giá trị input đi kèm với giá trị output và được dùng để kiểm thử độ chính xác của mô hình máy học trong quá trình huấn luyện.