

1. Overfitting (Quá khớp)

Overfitting xảy ra khi một mô hình học quá tốt trên tập huấn luyện, nhưng lại không thể tổng quát hóa tốt đối với dữ liệu mới (tập kiểm tra hoặc dữ liệu thực tế). Mô hình học được các chi tiết và nhiễu trong dữ liệu huấn luyện quá mức, dẫn đến việc nó quá "phức tạp" và không thể hoạt động tốt với dữ liệu chưa thấy.

Dấu hiệu của Overfitting:

- Mô hình có độ chính xác cao trên tập huấn luyện nhưng độ chính xác thấp trên tập kiểm tra hoặc tập xác thực.
- Mô hình học quá chi tiết về dữ liệu huấn luyện và không có khả năng khái quát.

Cách cải thiện Overfitting:

1. **Giảm độ phức tạp của mô hình:**
 - Dùng các mô hình đơn giản hơn (chẳng hạn giảm số lớp hoặc số neuron trong mạng neural).
 - Sử dụng các thuật toán học có tính khái quát tốt hơn.
2. **Sử dụng regularization (Điều chuẩn):**
 - **L1/L2 regularization:** Thêm một yếu tố phạt vào hàm mất mát để giảm độ lớn của các trọng số.
 - **Dropout (Đối với mạng neural):** Ngẫu nhiên bỏ qua một phần các neuron trong mỗi bước huấn luyện để tránh mô hình quá phụ thuộc vào một số đặc trưng nào đó.
3. **Tăng kích thước tập huấn luyện:**
 - Thu thập thêm dữ liệu huấn luyện để giúp mô hình học được nhiều đặc trưng và từ đó giảm overfitting.
 - **Data augmentation:** Tăng cường dữ liệu, đặc biệt hữu ích trong các bài toán xử lý hình ảnh, bằng cách tạo ra các biến thể khác nhau từ dữ liệu gốc.
4. **Cross-validation (Kiểm tra chéo):**
 - Sử dụng cross-validation để đánh giá hiệu suất mô hình trên nhiều phần dữ liệu khác nhau và giúp xác định xem mô hình có bị overfitting hay không.
5. **Early stopping (Dừng sớm):**
 - Dừng huấn luyện khi độ chính xác trên tập xác thực không còn cải thiện, thay vì tiếp tục huấn luyện cho đến khi đạt đến số epoch nhất định.

2. Underfitting (Thiếu khớp)

Underfitting xảy ra khi mô hình học không đủ khả năng để học được các mối quan hệ quan trọng trong dữ liệu huấn luyện, dẫn đến việc mô hình không thể đạt được độ chính xác cao trên cả tập huấn luyện và tập kiểm tra.

Dấu hiệu của Underfitting:

- Mô hình có độ chính xác thấp trên cả tập huấn luyện và tập kiểm tra.

- Mô hình quá đơn giản, không đủ khả năng để học các đặc trưng phức tạp trong dữ liệu.

Cách cải thiện Underfitting:

1. Tăng độ phức tạp của mô hình:

- Sử dụng các mô hình phức tạp hơn như mạng neural với nhiều lớp (deep learning), cây quyết định phức tạp hơn, hoặc các thuật toán học mạnh mẽ hơn như Random Forest, XGBoost, v.v.

2. Giảm regularization:

- Nếu đang sử dụng regularization quá mạnh, có thể khiến mô hình không đủ linh hoạt để học các đặc trưng trong dữ liệu. Cần giảm bớt mức độ regularization.

3. Tăng thời gian huấn luyện:

- Đảm bảo mô hình huấn luyện đủ lâu để có thể học các mối quan hệ trong dữ liệu. Nếu mô hình dừng huấn luyện quá sớm, nó có thể không học được đầy đủ các đặc trưng.

4. Thu thập thêm dữ liệu:

- Đôi khi thiếu dữ liệu có thể làm cho mô hình không thể học được đầy đủ thông tin. Mở rộng tập dữ liệu huấn luyện giúp mô hình học được các mẫu tốt hơn.