Anomaly Detection

Là phương pháp học không giám sát nhằm **phát hiện các điểm dữ liệu bất thường**, khác biệt so với phần lớn các điểm khác trong tập huấn luyện.

Đặc điểm chính:

- **Dữ liệu huấn luyện không có nhãn**, chủ yếu gồm các ví dụ "bình thường" (non-anomalous)
- **Mục tiêu**: Sau khi học từ dữ liệu bình thường, nếu có điểm dữ liệu mới trông **khác biệt đáng kể**, thì gán nó là *bất thường*

Kỹ thuật chính: Density Estimation (Ước lượng mật độ xác suất)

Ý tưởng:

1. Học mô hình xác suất p(x) từ tập dữ liệu bình thường
2. Với điểm mới x_{test}, tính $p(x_{\text{test}})$
3. So sánh với ngưỡng ε:
    - Nếu $\varepsilon p(x_{\text{test}}) < \varepsilon$: → **anomaly**
    - Nếu $p(x_{\text{test}}) \geq \varepsilon$: → **ok**

Ứng dụng thực tế của Anomaly Detection

Phát hiện gian lận (Fraud Detection)

- Ví dụ: website theo dõi hoạt động người dùng
    - Đặc trưng có thể gồm:
        - Số lần đăng nhập, số trang truy cập
        - Tốc độ gõ bàn phím (số ký tự/giây)
- Nếu người dùng nào có hành vi rất khác biệt → có thể là **bot hoặc gian lận**

Giám sát sản xuất (Manufacturing)

- Phát hiện lỗi trên các sản phẩm:

- Vi mạch, smartphone, động cơ, bảng mạch, ...
- Mục tiêu: kiểm tra kỹ những sản phẩm bất thường trước khi gửi cho khách hàng

Giám sát hệ thống máy chủ

- Dữ liệu xix_ixi: sử dụng CPU, đĩa cứng, mạng
- Máy nào có thông số bất thường → có thể bị lỗi, hỏng phần cứng hoặc bị xâm nhập

Giám sát trạm phát sóng

- Phát hiện trạm phát tín hiệu di động hoạt động lạ → cử kỹ thuật viên kiểm tra

Gaussian Distribution

Định nghĩa:

Là phân phối xác suất phổ biến, thường gọi là **"phân phối hình chuông (bell-shaped curve)"**, được đặc trưng bởi:

- **Trung bình (mean):** μ
- **Độ lệch chuẩn (standard deviation):** σ
- **Phương sai (variance):** σ^2

Công thức phân phối chuẩn 1D:

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- μ: tâm của phân phối (điểm chính giữa)
- σ: độ rộng của đường cong (độ "phân tán")
- **Diện tích dưới đường cong luôn bằng 1**

Áp dụng vào Anomaly Detection

Ý tưởng:

- Học phân phối xác suất p(x) từ dữ liệu “bình thường”
- Với điểm mới x_{test}:
  - Tính $p(x_{\text{test}})$
  - Nếu $p(x_{\text{test}}) < \varepsilon$: → Flag là bất thường

Ưu điểm:

- Dễ triển khai, đặc biệt với 1 hoặc vài đặc trưng
- Có nền tảng xác suất rõ ràng
- Có thể mở rộng sang nhiều chiều (multivariate Gaussian)

Xây dựng Hệ thống Anomaly Detection nhiều đặc trưng

Đầu vào (Input):

- Tập huấn luyện: $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$
- Mỗi $x^{(i)} \in \mathbb{R}^n$: có **n đặc trưng**

Ý tưởng chính:

Xây dựng mô hình xác suất p(x)p(x)p(x) – tức xác suất xuất hiện của một vector đặc trưng xxx.

Nếu p(x)<εp(x) < \varepsilonp(x)<ε → đánh dấu là **anomaly** (bất thường)

Cho ví dụ mới x_{test}, thực hiện:

1. Tính:

$$p(x_{\text{test}}) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$$

2. So sánh với ngưỡng :
   - Nếu p(x)<ε → **Anomaly**
   - Ngược lại → **Bình thường**

Lưu ý

Cần đánh giá hệ thống trong quá trình phát triển

- Khi **thay đổi tính năng** hoặc **tinh chỉnh tham số (như ε)**, ta cần một **chỉ số đánh giá định lượng (real number evaluation)** để biết liệu thay đổi đó có cải thiện kết quả không.

Dùng nhãn để đánh giá, dù thuật toán là không giám sát

- Thuật toán vẫn **học từ tập huấn luyện không gán nhãn** (toàn là y=0y = 0y=0 – giả định là "bình thường").
- Nhưng để **đánh giá**, cần một **số ít ví dụ bất thường có nhãn** y=1y = 1y=1.

Cách đánh giá

- Huấn luyện mô hình Gaussian từ tập huấn luyện.
- Trên tập cross-validation hoặc test

Sử dụng các chỉ số đánh giá phù hợp với dữ liệu lệch

Nếu số anomaly quá ít, có thể bỏ tập test

- Dồn toàn bộ anomaly + engine thường còn lại vào **cross-validation**
- Sử dụng toàn bộ tập này để điều chỉnh ε và các đặc trưng

Feature Engineering

Biến đổi đặc trưng để có phân phối gần Gaussian

- Vì mô hình giả định phân phối **Gaussian (chuẩn)** cho từng đặc trưng ⇒ Nếu đặc trưng quá lệch hoặc phân phối không chuẩn thì cần **biến đổi để gần Gaussian hơn**.

Tạo thêm đặc trưng mới (feature engineering)

- Khi một điểm bất thường **trông giống điểm bình thường theo các đặc trưng hiện tại**, mô hình có thể **không phát hiện được**.

Tạo đặc trưng kết hợp (feature combinations)

- Khi các đặc trưng riêng lẻ không gây bất thường nhưng kết hợp lại thì có.