**Hồi quy mạnh (Robust Regression)** là một kỹ thuật thống kê được dùng để xử lý dữ liệu khi mô hình hồi quy tuyến tính thông thường (OLS – Ordinary Least Squares) không hoạt động hiệu quả, đặc biệt khi có ngoại lệ (outliers) hoặc vi phạm giả định phân phối chuẩn của sai số.

1. Lý thuyết về hồi quy mạnh

Hồi quy mạnh cố gắng giảm ảnh hưởng của các điểm dữ liệu bất thường (outliers) trong tập dữ liệu. Trong khi hồi quy OLS tìm cách tối thiểu hóa tổng bình phương sai số (∑(y\_i - ŷ\_i)2), thì hồi quy mạnh sử dụng hàm mất mát khác để giảm ảnh hưởng của những điểm có sai số lớn.

2. Cách hoạt động

Các bước chính trong hồi quy mạnh:

Bước 1: Chọn một hàm mất mát "mạnh" (robust loss function): ví dụ như Huber loss, Tukey’s biweight, hay Least Absolute Deviations (LAD) thay cho MSE.

Bước 2: Lặp lại việc ước lượng trọng số cho từng điểm dữ liệu: những điểm có sai số lớn sẽ được gán trọng số nhỏ, làm giảm ảnh hưởng của chúng.

Bước 3: Lặp lại cho đến khi hội tụ: quá trình này thường dùng thuật toán lặp lại như IRLS (Iteratively Reweighted Least Squares).

3. Các phương pháp hồi quy mạnh phổ biến

| Phương pháp | Đặc điểm chính |
| --- | --- |
| Huber Regression | Kết hợp giữa MSE và MAE |
| LAD (Least Absolute Deviations) | Tối thiểu hóa tổng trị tuyệt đối sai số |

4. Ví dụ minh họa

Cho tập dữ liệu như sau:

| X | Y |
| --- | --- |
| 1 | 2 |
| 2 | 4 |
| 3 | 6 |
| 4 | 8 |
| 5 | 10 |
| 6 | 30 ← outlier |

* OLS Regression sẽ bị kéo nghiêng về phía điểm outlier (6, 30), dẫn đến đường hồi quy không phản ánh đúng xu hướng chung.
* Hồi quy mạnh (ví dụ Huber) sẽ coi (6, 30) là điểm ngoại lệ và làm giảm trọng số của nó, từ đó giữ cho đường hồi quy gần hơn với các điểm còn lại.

Khi nào nên dùng hồi quy mạnh?

* Khi có ngoại lệ (outlier) trong dữ liệu.
* Khi dữ liệu không phân phối chuẩn.
* Khi nghi ngờ rằng một số quan sát có thể bị nhiễu hoặc sai lệch.
* 