**Kỹ thuật ISOMAP**

Là 1 kỹ thuật giảm chiều dữ liệu phi tuyến tính dựa trên phương pháp Multi-Dimensional Scaling bằng cách sử dụng khoảng cách địa lý thay vì khoảng cách Euclidean. Khoảng cách địa lí là khoảng cách ngắn nhất đo được trên đồ thị, khác với khoảng cách Euclidean là khoảng cách thẳng (đường chim bay).

**Các bước thực hiện:**

* Kết nối mỗi điểm dữ liệu với các điểm lân cận gần nhất bằng một đồ thị có trọng số, trong đó trọng số là khoảng cách Euclidean.
* Sử dụng thuật toán Floyd-Warshall hoặc Dijkstra để tính khoảng cách địa lí giữa các cặp điểm trên đồ thị.
* Nhúng dữ liệu vào không gian chiều thấp hơn. Giải bài toán trị riêng trên ma trận khoảng cách để tìm tọa độ tối ưu trong không gian giảm chiều.

**Ưu điểm:**

* ISOMAP không chỉ bảo toàn quan hệ cục bộ giữa các điểm dữ liệu mà còn duy trì cấu trúc toàn cục trong không gian nhúng mới.
* Hiệu quả với nhiều dạng dữ liệu khác nhau, giúp giảm chiều mà vẫn giữ được thông tin quan trọng.
* Dễ dàng triển khai hơn so với các phương pháp như t-SNE hoặc UMAP, ISOMAP có một quy trình rõ ràng với ít hyperparameters hơn cần điều chỉnh. Khi chọn được số neighbors (k) phù hợp, ISOMAP hoạt động khá tốt mà không cần tinh chỉnh nhiều.

**Nhược điểm:**

* Nếu chọn neighbors quá nhỏ, đồ thị có thể bị rời rạc và mất thông tin. Nếu chọn quá lớn, khoảng cách geodesic sẽ gần như giống khoảng cách Euclidean, làm mất đi ưu thế của ISOMAP.
* ISOMAP phải tính toán tất cả khoảng cách giữa các cặp điểm dữ liệu, giải bài toán tìm trị riêng của ma trận, dẫn đến độ phức tạp cao (**O(N³)**), chi phí tính toán cao, khó mở rộng với dữ liệu lớn.
* Hoạt động kém với bộ dữ liệu chứa nhiễu và khó áp dụng vào bộ dữ liệu mới do cần phải tính toán lại toàn bộ đồ thị.

**Ứng dụng:**

* Thị giác máy tính: Nhận dạng khuôn mặt, nhận dạng vân tay.
* Y khoa: Phân tích cấu trúc protein, biểu hiện gene, sơ đồ mạch máu.
* Xử lý ngôn ngữ tự nhiên: Trích xuất đặc trưng từ văn bản.