

Mối quan hệ giữa học máy, đại số tuyến tính, vector và ma trận

Đại số tuyến tính

Có 2 loại bài toán mà đại số tuyến tính thường được sử dụng để giải quyết:

- **Bài toán xác định giá (price discovery):**

Một người đi mua táo và chuối hai lần, mỗi lần mua với số lượng khác nhau và tổng chi phí được biết. Mục tiêu là tìm ra giá của từng loại trái cây bằng cách giải hệ phương trình tuyến tính. Khi có nhiều loại hàng hóa và nhiều lần mua sắm, việc giải hệ phương trình bằng tay sẽ trở nên khó khăn, và ta cần đến thuật toán máy tính – đây chính là một ví dụ điển hình của bài toán Đại số tuyến tính.

- **Bài toán tối ưu hóa (fitting data):**

Đây là bài toán tìm phương trình phù hợp nhất với tập dữ liệu đã cho (ví dụ như một biểu đồ histogram mô tả dân số). Trong học máy, không chỉ cần tìm tham số tối ưu mà còn để máy tính xác định luôn cả cấu trúc của phương trình phù hợp. Điều này giúp mô tả dữ liệu hiệu quả mà không cần lưu trữ toàn bộ dữ liệu gốc, đồng thời hỗ trợ bảo mật quyền riêng tư.

Vector

Vector là:

- Một danh sách các con số.
- Vị trí trong ba chiều không gian và trong một chiều thời gian.
- Một cái gì đó di chuyển trong một không gian của các thông số phù hợp.

Lấy ví dụ với bài toán **fitting dữ liệu**, sử dụng một **phân bố chuẩn (Gaussian distribution)** với hai tham số:

- **μ (mu):** trung bình (vị trí trung tâm của phân bố),
- **σ (sigma):** độ lệch chuẩn (độ rộng của phân bố).

Mục tiêu là tìm cặp giá trị μ và σ sao cho đường cong khớp tốt nhất với dữ liệu thực tế. Để làm điều này, ta đánh giá **độ tốt của phép khớp** (goodness of fit)

bằng cách tính **tổng bình phương sai số** giữa giá trị dự đoán và dữ liệu thật.

Việc điều chỉnh tham số có thể được hình dung như **di chuyển trong không gian tham số** (parameter space), nơi mỗi bước di chuyển được biểu diễn bằng một **vector**. Khi đó:

- **Vectors** biểu diễn hướng thay đổi (cải thiện hay không),
- **Tính toán đạo hàm, gradient** sẽ giúp tìm hướng giảm nhanh nhất trong "địa hình" độ sai lệch – một kỹ thuật cốt lõi của **tối ưu hóa và học máy**.

⇒ **Hiểu vectors và phép tính đạo hàm trên chúng** là tiền đề để giải các bài toán tối ưu, fitting và học máy.