Report paper
Maximum Likelihood Estimation and Maximum A Posteriori

Tóm tắt nội dung:

Bài report này là phần tóm tắt ý tưởng, nội dung chính của 2 cách ước lượng tham số Maximum Likelihood Estimation (MLE) và Maximum A Posteriori (MAP), đồng thời so sánh hai cách ước lượng này.

I. Introduction

Có 2 cách ước lượng tham số thường dùng trong các mô hình thống kê của machine learning.

- Maximum Likelihood Estimate (MLE): dựa vào dữ liệu trong tập huấn luyện để ước lượng bộ tham số θ của mô hình để xác suất dự đoán chính xác của tập huấn luyện là cao nhất.
- Maximum A Posteriori (MAP): ngoài tập huấn luyện ra, cách ước lượng này còn dùng những thông tin, giả thiết biết trước của các tham số để xây dựng bộ θ tối ưu. Ví dụ trong bài toán tung đồng xu, ta giả thiết xác suất của mặt ngửa là con số xấp xỉ 0.5, ta dựa vào đây để điều chỉnh bộ tham số θ.

II. Maximum Likelihood Estimation

1. Ý tưởng

Giả sử ta có cá điểm dữ liệu x1, x2,…, xN tuân theo một phân phối (mô hình) nào đó được mô tả bởi bộ tham số θ nào đó. Công việc của MLE là đi tìm bộ tham số θ sao cho xác suất đồng thời của các điểm dữ liệu đạt giá trị lớn nhất (GTLN):

$$\theta = \max_{\theta} p(x1 , . . . , xN|\theta) \qquad (2.0)$$

Trong đó, xác suất đồng thời p(x1 , . . . , xN|θ) gọi là likelihood
Giả sử các điểm dữ liệu x1,x2,,xN độc lập, p(x1,x2|θ) = p(x1|θ)p(x2|θ) (2.0) có thể biến đổi thành:

$$\theta = \max_{\theta} (\Pi\ p(x_i|\theta)(2.0)\ \theta))\ \text{với i: 1->N} \qquad (2.1)$$

hay

$$\theta = \max_{\theta} \Sigma \log(p(x_i|\theta)(2.0)\ \theta))\ \text{với i: 1->N}\ (2.2)$$

vì: - Tìm max của tổng đơn giản hơn max của tích
- Log của một tích bằng tổng các tích
- Log(x) là hàm đồng biến: giá trị θ để (2.1) đạt GTLN chính là giá trị θ để (2.2) đạt GTLN

2. Tóm tắt:

- Input: Bộ các điểm dữ liệu x1,x2,…,xN tuân theo một mô hình được mô tả bởi bộ tham số θ
- Output: Giá trị θ sao cho xác suất đồng thời của các điểm dữ liệu (độc lập) là maximum

III. Maximum A Posteriori

1. Ý tưởng

Trở lại với bài toán tung đồng xu ban đầu, nếu tung 5000 lần và kết quả có 1000 mặt ngửa thì ta có thể đánh giá xác suất của mặt ngửa là 0.2 là đáng tin vì số lượng mẫu tương đối lớn. Nhưng nếu chỉ tung 5 lần thì kết quả này không đáng tin cậy vì tập mẫu quá nhỏ. Nên cần dùng tới tham số giả thiết ban đầu của dữ liệu để dự đoán tham số. Ở TH này, ta có giả thiết giá trị mặt ngửa có xác suất là 0.5 - đây được gọi là prior của tham số θ (Trong bài toán đồng xu, θ chính là tham số lamda λ của phân phối Bernouli).

Ngược với MLE, MAP đi đánh giá tham số θ như xác suất có điều kiện của dữ liệu:

$$\theta = \underset{\theta}{\operatorname{argmax}}\ p(\theta|x1,...,xN) \qquad (3.0)$$

Trong đó, p(θ|x1,...,xN) được gọi là xác suất posterior của θ

Dựa vào quy tắc Bayes, ta có

$$p(\theta|x1,...,xN) = p(x1,...,xN\ |\theta)p(\theta)/p(x1, x2,\ldots,xN)$$ *với p(θ) gọi là xác suất prior.*

Biến đổi 3.0 thành:

$$\theta = \underset{\theta}{\operatorname{argmax}}\ [\Pi\ p(x\Pi\ p(x_i|\theta)(2.0)\ \theta)p(\theta)] \text{ với } i: 1\text{->}N \qquad (3.1)$$

Như vậy, ta có thể đưa ra nhận xét posteriori tit lệ thuận với tích của likelihood và prior

2. Lựa chọn prior - Conjugate prior

Nếu phân phối xác suất posterior p(θ|x1,...,xN) có cùng dạng với phân phối xác suất p(θ) thì prior và posterior được gọi là conjugate distributions và hàm prior p(θ) được gọi là conjugate prior cho hàm likelihood p(x1,...,xN|θ)

3. Tóm tắt:

- Input: Bộ các điểm dữ liệu x1, x2, … , xN tuân theo một mô hình được mô tả bởi bộ tham số θ, giả thiết prior cho bộ tham số θ.
- Output: Giá trị θ sao cho xác suất đồng thời của các điểm dữ liệu (độc lập) là maximum

IV. Nhận xét

- Nhiệm của 2 bài toán MLE và MAP đều là bộ tham số θ nên có cấu trúc giống nhau
- MLE và MAP đều dựa vào dữ liệu huấn luyện để dự đoán bộ tham số θ, tuy nhiên MAP sử dụng thêm giả thiết của tham số nên kết quả tìm được chính xác hơn hầu hết các TH đặc biệt là khi có ít dữ liệu huấn luyện

V. Reference

Machine Learning Cơ Bản - Vũ Hữu Tiệp