MLE và MAP

Nguyen Nguyen

Ngày 1 tháng 8 năm 2019

Outline

1. Bài toán mở đầu
2. Ước lượng hợp lý cực đại
3. Ước lượng cực đại hậu nghiệm

Bài toán ngược

Lưu lượng xe lưu thông trong TPHCM được biểu diễn bởi một biến ngẫu nhiên X phụ thuộc vào một phân phối D. Sau năm 2018, X cho ta 365 dữ liệu $x_1, ..., x_{365}$.

Bài toán ngược

Lưu lượng xe lưu thông trong TPHCM được biểu diễn bởi một biến ngẫu nhiên X phụ thuộc vào một phân phối D. Sau năm 2018, X cho ta 365 dữ liệu $x_1, \ldots, x_{365}$.
Câu hỏi đặt ra: Bây giờ giả sử ta quan sát được 365 dữ liệu này mà ta không biết phân phối D, hỏi có cách nào khôi phục hoặc dự đoán gần chính xác phân phối D không?

Bài toán mở đầu

Bài toán (Tìm phân phối xác suất)

Cho một phân phối xác suất D chưa biết và X là một biến ngẫu nhiên $X \sim D$. Cho $x_1, x_2, ..., x_n$ là n giá trị thu thập được từ X. Hãy khôi phục phân phối D từ n giá trị này.

Tham số hóa phân phối

- Phân phối của chúng ta có thể thuộc một họ các phân phối xác suất có cùng cấu trúc.
- Họ các phân phối này thường phụ thuộc vào các **tham số** mà khi xác định các tham số này, ta xác định chính xác phân phối.

Tham số hóa phân phối

- Phân phối của chúng ta có thể thuộc một họ các phân phối xác suất có cùng cấu trúc.
- Họ các phân phối này thường phụ thuộc vào các **tham số** mà khi xác định các tham số này, ta xác định chính xác phân phối.

Vậy bài toán của chúng ta cơ bản chỉ là xác định tham số hợp lý nhất so với dữ liệu, nếu ta đã biết phân phối cho trước thuộc họ phân phối thế nào.

Review một tí

Một số họ phân phối ta đã được học:

- Phân phối Bernoulli: $B(1, \lambda)$.
- Phân phối nhị thức: $B(n, p)$.
- Phân phối Poisson: $P(\lambda)$.
- Phân phối hình học: $Geo(p)$, siêu hình học $H(N, K, n)$.
- Phân phối đều: $U(a, b)$.
- Phân phối chuẩn: $N(\mu, \sigma^2)$.

Ví dụ

Một cầu thủ sút penalty vào lưới 100 quả thì được 25 quả. Hỏi xác suất sút trúng của cầu thủ này là bao nhiêu?

Ví dụ

Một cầu thủ sút penalty vào lưới 100 quả thì được 25 quả. Hỏi xác suất sút trúng của cầu thủ này là bao nhiêu?

- Bài toán yêu cầu xác định phân phối Bernoulli $B(1, \theta)$ ứng với việc sút trúng/không trúng của cầu thủ.
- Cơ sở toán học nào để ta ước lượng gần chính xác việc này từ dữ liệu đã cho?

Giải quyết

- Nếu không có thông tin gì ngoài dữ liệu, ta sử dụng phương pháp **Ước lượng hợp lý cực đại (Maximum Likelihood Estimation**
- Nếu có thông tin tiền nghiệm về phân phối xác suất, ta sử phương pháp **Ước lượng cực đại hậu nghiệm (Maximum a Posteriori Estimation)**.

Giải quyết

- Nếu không có thông tin gì ngoài dữ liệu, ta sử dụng phương pháp **Ước lượng hợp lý cực đại (Maximum Likelihood Estimation**
- Nếu có thông tin tiền nghiệm về phân phối xác suất, ta sử phương pháp **Ước lượng cực đại hậu nghiệm (Maximum a Posteriori Estimation)**.

Hai phương pháp này có gì khác nhau?

MLE

Xét họ phân phối $D(\theta)$ và một biến ngẫu nhiên $X \sim D(\theta)$ và n giá trị $x_1, x_2, ..., x_n$ có được từ n phép thử của X.

Định nghĩa (Ước lượng cực đại hợp lý)

Ước lượng cực đại hợp lý là ước lượng giá trị θ_0 thỏa mãn:

$$\theta_0 = \arg\max_{\theta} P(x_1, x_2, ..., x_n | \theta)$$

- Nói cách khác, *MLE* là phương pháp ước lượng độ hợp lý của dữ liệu xét trên một phân phối cho trước.
- Phân phối nào cho ta độ hợp lý chính xác nhất thì ta sẽ chọn phân phối đó.

Giả thiết độc lập, hàm log và cách tính MLE

Giả thiết các dữ liệu là độc lập:

$$\theta_0 = \arg\max_{\theta \sim P} P\left(x_1, x_2, ..., x_n | \theta\right)$$
$$= \arg\max_{\theta \sim P} \prod_{i=1}^{n} P(x_i|\theta).$$

Giả thiết độc lập, hàm log và cách tính MLE

Giả thiết các dữ liệu là độc lập:

$$\begin{aligned}\theta_0 &= \arg\max_{\theta \sim P} P\left(x_1, x_2, ..., x_n | \theta\right) \\ &= \arg\max_{\theta \sim P} \prod_{i=1}^{n} P(x_i|\theta).\end{aligned}$$

Giá trị này **rất gần 0** nếu số lượng data quá lớn, nên ta nên chuyển về tổng bằng hàm log.

$$\begin{aligned}\theta_0 &= \arg\max_{\theta \sim P} log\left(\prod_{i=1}^{n} P(x_i|\theta)\right). \\ &= \arg\max_{\theta \sim P} \sum_{i=1}^{n} log\left(P(x_i|\theta)\right).\end{aligned}$$

Ví dụ

Một cầu thủ sút penalty vào m lần trên tổng số n lần với $0 < m < n$. Ước lượng xác suất để cầu thủ sút penalty vào mỗi lần.

Ví dụ

Một cầu thủ sút penalty vào m lần trên tổng số n lần với $0 < m < n$. Ước lượng xác suất để cầu thủ sút penalty vào mỗi lần.

Xét X là biến ngẫu nhiên nhận hai gía trị $0, 1$ ứng với sút vô và sút trượt. Khi đó $X \sim B(1, \theta)$ là một phân phối Bernoulli. Ta có:

Ví dụ

Một cầu thủ sút penalty vào m lần trên tổng số n lần với $0 < m < n$. Ước lượng xác suất để cầu thủ sút penalty vào mỗi lần.

Xét X là biến ngẫu nhiên nhận hai gía trị $0, 1$ ứng với sút vô và sút trượt. Khi đó $X \sim B(1, \theta)$ là một phân phối Bernoulli. Ta có:

$$P(0|\theta) = 1 - \theta, P(1|\theta) = \theta$$

Như vậy ta có

$$P(x_1, ..., x_n|\theta) = \prod_{i=1}^{n} P(x_i|\theta) = \theta^m (1 - \theta)^n$$

Ví dụ

Một cầu thủ sút penalty vào m lần trên tổng số n lần với $0 < m < n$. Ước lượng xác suất để cầu thủ sút penalty vào mỗi lần.

Xét X là biến ngẫu nhiên nhận hai gía trị $0, 1$ ứng với sút vô và sút trượt. Khi đó $X \sim B(1, \theta)$ là một phân phối Bernoulli. Ta có:

$$P(0|\theta) = 1 - \theta, P(1|\theta) = \theta$$

Như vậy ta có

$$P(x_1, ..., x_n|\theta) = \prod_{i=1}^{n} P(x_i|\theta) = \theta^m (1 - \theta)^n$$

Như vậy ước lượng hợp lý cho θ là $\frac{m}{n}$. Nghĩa là xác suất sút trúng vào lưới của cầu thủ là $\frac{m}{n}$.

Nhược điểm của MLE

- MLE có đặc điểm chính là phụ thuộc rất nhiều vào dữ liệu. Khi dữ liệu có hiện tượng thiên vị hoặc do số lượng data chưa đủ lớn, mô hình dễ đưa ra các đánh giá sai, dẫn đến hiện tượng **overfitting**
- Để khắc phục vấn đề này, ta cần phải thêm một thông tin để làm công cụ đánh giá, dẫn đến phương pháp khác là MAP.

1. Bài toán mở đầu
2. Ước lượng hợp lý cực đại
3. Ước lượng cực đại hậu nghiệm

MAP

Định nghĩa (Ước lượng cực đại hậu nghiệm)

Cũng xét ví dụ như trên. Giả sử rằng $X \sim D(\theta)$ và θ là một biến ngẫu nhiên có phân phối tiền nghiệm $\theta \sim P$. Ước lượng cực đại hậu nghiệm của θ là giá trị θ_0 sao cho xác suất:

$$P(\theta = \theta_0 | (X_1 = x_1), ..., (X_n = x_n))$$

là lớn nhất.

Biểu thức trên là **độ hợp lý hậu nghiệm** của $D(\theta_0)$.

Cách ước lượng cực đại hậu nghiệm

Ta cần ước lượng giá trị θ_0 như sau:

$$\theta_0 = \underset{\theta \sim P}{\arg\max}\, P(\theta | x_1, x_2, ..., x_n)$$

Từ định lý Bayes:

$$\begin{aligned}
\theta^* &= \underset{\theta \sim P}{\arg\max}\, P\left(\theta | x_1, x_2, ..., x_n\right) \\
&= \underset{\theta \sim P}{\arg\max}\, \frac{P\left(x_1, x_2, ..., x_n | \theta\right) . P\left(\theta\right)}{P\left(x_1, x_2, ..., x_n\right)} \\
&= \underset{\theta \sim P}{\arg\max}\, P\left(x_1, x_2, ..., x_n | \theta\right) . P\left(\theta\right).
\end{aligned}$$

Giả thiết độc lập, hàm log

Giả sử các dữ liệu độc lập nhau:

$$\begin{aligned}\theta^* &= \arg\max_{\theta\sim P} P\left(\theta|x_1, x_2, ..., x_n\right)\\ &= \arg\max_{\theta\sim P} \prod_{i=1}^{n} P(x_i|\theta).P(\theta)\end{aligned}$$

Logarithm hóa:

$$\theta^* = \arg\max_{\theta\sim P} \sum_{i=1}^{n} \log(P(x_i|n)) + \log(P(\theta)).$$

Ví dụ

Xét ví dụ ở MLE, giả sử một câu thủ sút 100 quả pen chỉ vô 25 quả. Tuy nhiên bình thường cầu thủ này rất xuất sắc và luôn sút trúng với xác suất là một phân phối đều trong khoảng $[0.7, 0.8]$. Ta có ước lượng bằng MAP:

$$\begin{aligned}
\theta^* &= \underset{\theta \sim U(0.7;0.8)}{\arg\max} \; P\left(x_1, x_2, ..., x_n | \theta\right).P\left(\theta\right) \\
&= \underset{\theta \sim U(0,7;0.8)}{\arg\max} \prod_{i=1}^{n} P(x_i|\theta).10 \\
&= \underset{\theta \sim U(0,7;0.8)}{\arg\max} \; \theta^{25}(1-\theta)^{75} = 0.7.
\end{aligned}$$

Ví dụ

Xét ví dụ ở MLE, giả sử một câu thủ sút 100 quả pen chỉ vô 25 quả. Tuy nhiên bình thường cầu thủ này rất xuất sắc và luôn sút trúng với xác suất là một phân phối đều trong khoảng $[0.7, 0.8]$. Ta có ước lượng bằng MAP:

$$
\begin{aligned}
\theta^* &= \underset{\theta \sim U(0.7;0.8)}{\arg\max} \; P\left(x_1, x_2, ..., x_n|\theta\right).P\left(\theta\right) \\
&= \underset{\theta \sim U(0,7;0.8)}{\arg\max} \prod_{i=1}^{n} P(x_i|\theta).10 \\
&= \underset{\theta \sim U(0,7;0.8)}{\arg\max} \; \theta^{25}(1-\theta)^{75} = 0.7.
\end{aligned}
$$

Kết luận: Đẳng cấp là mãi mãi.

Ý nghĩa của MAP

- Khắc phục nhược điểm về sự phụ thuộc dữ liệu của MLE.

Ý nghĩa của MAP

- Khắc phục nhược điểm về sự phụ thuộc dữ liệu của MLE.
- Có ý nghĩa trong bài toán phân loại Bayes.

Ý nghĩa của MAP

- Khắc phục nhược điểm về sự phụ thuộc dữ liệu của MLE.
- Có ý nghĩa trong bài toán phân loại Bayes.
- Cả hai phương pháp đều giải quyết câu hỏi xác định phân phối xác suất, bài toán có ý nghĩa trong ML.