TRƯỜNG THPT CHUYÊN TRẦN ĐẠI NGHĨA

---ଓ📖ଌ---

Chuyên đề: Maximum Likelihood

Giáo viên hướng dẫn: Th.s Hồ Ngọc Lâm

Lớp: 12CTin

Nhóm học sinh thực hiện:

27 – Nguyễn Hoàng Phúc Thịnh

32 – Phù Phương Tuấn

33 – Nguyễn Việt Tùng

35 – Trần Thái Vỹ

Thành phố Hồ Chí Minh, ngày 17 tháng 10 năm 2023

LỜI NÓI ĐẦU

"Những sự kiện có xác suất cao là những sự kiện có khả năng xảy ra hơn."

-Khuyết Danh-

Câu nói nói cũng như không này là khơi nguồn cho rất nhiều các thuật toán Machine Learning có liên quan đến xác suất.

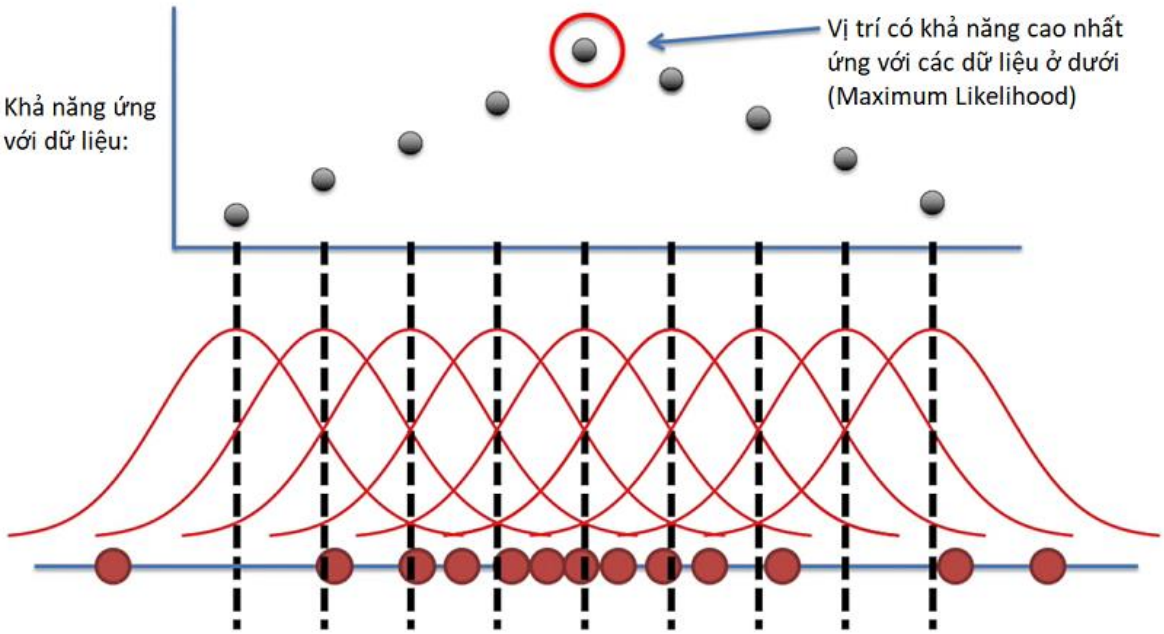
Thuật toán Maximum Likelihood (ML) là một trong những thuật toán được sử dụng phổ biến trong lĩnh vực thống kê và học máy. Khi xử lý các biến số có thể quan sát được hoặc không quan sát được, việc sử dụng trường hợp các biến số quan sát được là hoàn toàn có thể để học và dự đoán những biến số không quan sát được, thông qua các thuật toán và mô hình dựa trên những biến số quan sát được. Phương pháp này được gọi là xử lý dữ liệu thiếu. Vì thế, Maximum Likelihood được dùng để xử lý các tình huống khi dữ liệu bị thiếu một phần, và sử dụng một phần dữ liệu để dự đoán các dữ liệu còn thiếu. Ứng dụng của Maximum Likelihood rất đa dạng và phổ biến trong nhiều lĩnh vực như xử lý ngôn ngữ tự nhiên, nhận diện mẫu, dự đoán tài chính, ...Báo cáo này sẽ giúp hiểu rõ hơn về khái niệm và phương pháp của Maximum Likelihood.

MỤC LỤC

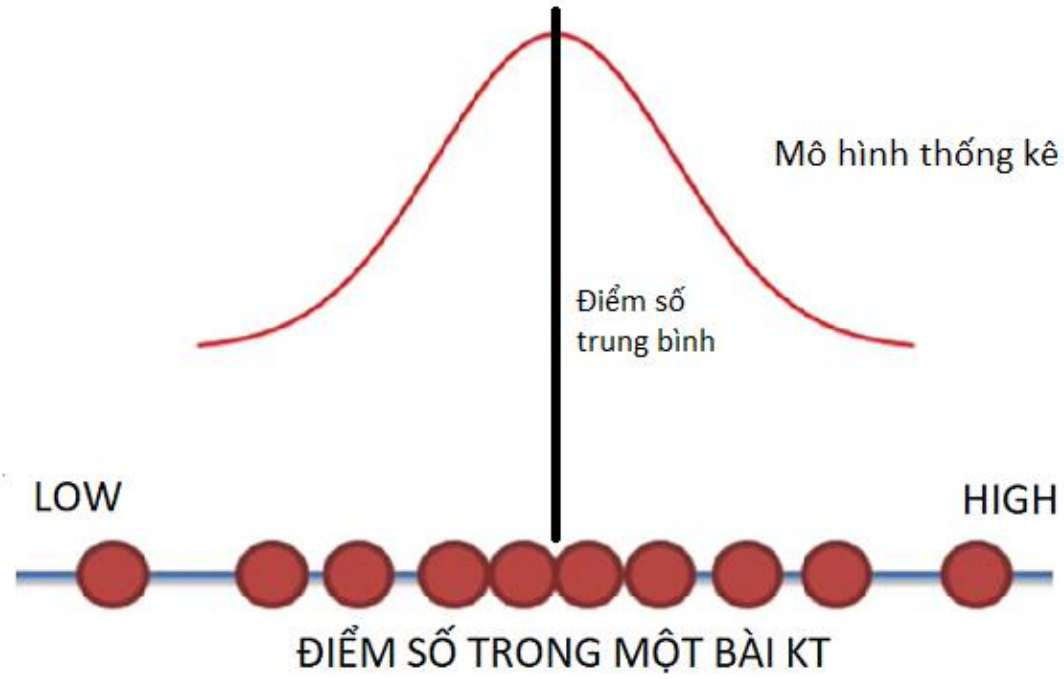
1 Maximum Likelihood (Hợp Lí Cực Đại) Là Gì?

Maximum Likelihood (cũng được gọi là giá trị hàm likelihood lớn nhất) đề cập đến giá trị của mô hình tham số mà khi được sử dụng trong hàm Likelihood, tạo ra xác suất lớn nhất cho dữ liệu quan sát được. Tức là, nó đại diện cho việc tối đa hóa cơ hội xảy ra của dữ liệu dựa trên giả định mô hình.

Qua quá trình MLE (Maximum Likelihood Estimation), ta tìm kiếm giá trị tham số tối ưu nhằm tối đa hóa hàm likelihood và xác định được mô hình phù hợp nhất cho dữ liệu.



2 Maximum Likelihood Estimation (Ước Lượng Hợp Lí Cực Đại)



⇒ ***Ước lượng được vị trí tập trung của phân phối dữ liệu***

2.1 Các Công Thức Để Ước Lượng

Giả sử trong phương pháp ước lượng hợp lý cực đại (MLE): Tất cả các giả thiết đều có giá trị xác suất trước như nhau: $P(h_i) = P(h_j), \forall h_i, h_j \in H$
Phương pháp MLE tìm giả thiết cực đại hoá giá trị $P(D \mid h)$ được gọi là khả năng có thể (Likelihood) của dữ liệu D đối với h.

Giả thiết hợp lý cực đại (Maximum Likelihood Hypothesis) được xác định như sau:

$$h_{ML} = \arg\max_{h \in H} P(D \mid h)$$

Tương tự giả sử mô hình được mô tả bởi bộ tham số θ, các "quan sát" (hay điểm dữ liệu) là $x_1, x_2, x_3, \ldots x_N$. Khi đó chúng ta cần tìm:

$$\theta = \arg\max P(x_1, x_2, x_3, \ldots x_N \mid \theta) \quad (1)$$

$P(x_1, x_2, x_3, \ldots x_N \mid \theta)$ là xác suất để các sự kiện $x_1, x_2, \ldots x_N$ xảy ra đồng thời, được gọi là Likelihood. Chính vì vậy mà phương pháp này được gọi là Maximum Likelihood.

Tuy nhiên, việc giải trực tiếp bài toán (1) thường là khó khăn. Chúng ta có thể đơn giản hoá bài toán bằng việc giả sử các điểm dữ liệu xảy ra độc lập với nhau. Khi đó, (1) trở thành:

$$\theta = \arg\max \prod P(x_N \mid \theta) \quad (2)$$

Chúng ta có thể khiến việc tính toán dễ dàng hơn bằng cách biến đổi về bài toán Maximum Log-likelihood:

$$\theta = \arg\max \sum \log\big(P(x_N \mid \theta)\big) \quad (3)$$

2.1.1 Ví Dụ 1

Tập H bao gồm 2 giải thiết có thể:
h_1: Anh ta chơi đá banh
h_2: Anh ta không chơi đá banh
D: Tập dữ liệu (các ngày) mà trong đó thuộc tính $Trời$ có giá trị $Nắng$ và thuộc tính $Gió$ có giá trị $Mạnh$

Tính 2 giá trị khả năng xảy ra (Likelihood Values) của dữ liệu D đối với 2 giải thiết: $P(D \mid h_1)$ và $P(D \mid h_2)$

$$P(Trời = Nắng, Gió = Mạnh \mid h_1) = \frac{1}{8}$$
$$P(Trời = Nắng, Gió = Mạnh \mid h_2) = \frac{1}{4}$$

Giả thiết MLE $h_{MLE} = h_1$ nếu $P(D \mid h_1) \geq P(D \mid h_2)$; và ngược lại thì $h_{MLE} = h_2$
Vì $P(Trời = Nắng, Gió = Mạnh \mid h_1) < P(Trời = Nắng, Gió = Mạnh \mid h_2)$
⇒ Hệ thống sẽ kết luận rằng: Anh ta sẽ không chơi đá banh.

2.1.2 Ví Dụ 2

Giả sử bài toán là có 5 học sinh làm bài kiểm tra được số điểm lần lượt là: 3, 6, 5, 9, 8. Để mô hình hoá điểm của các học sinh này, ta giả thiết các điểm dữ liệu được phân bố theo phân phối Gaussian:

$$P(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

Để dự đoán bộ tham số của phân phối chuẩn, ta sử dụng phương pháp MLE:

$$\mu, \sigma = \arg\max_{\mu, \sigma} \left[\prod_{i=1}^{N} P(x_i \mid \mu, \sigma^2)\right]$$
$$= \arg\max_{\mu, \sigma} \left[\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\right] \exp\left(-\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{2\sigma^2}\right)$$
$$= \arg\max_{\mu, \sigma} \left[-N\log(\sigma) - \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{2\sigma^2}\right]$$

Để tìm μ và σ sao cho biểu thức trong ngoặc vuông đạt giá trị cực đại, chúng ta đạo hàm biểu thức theo từng biến và giải phương trình khi giá trị đó bằng 0.

$$\frac{\partial J}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i - \mu)$$
$$\frac{\partial J}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i - \mu)^2$$

Từ đó ta có:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$
$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Thay các điểm dữ liệu vào công thức trên, ta tìm được $\mu = 6.2$ và $\sigma = 2.14$.

2.2 Bernoulli Distribution (Phân Phối Bernoulli)

Bài toán: giả sử tung một đồng xu N lần và nhận được n mặt $ngửa$. Tính xác suất có một mặt $ngửa$ khi tung đồng xu đó ở lần tiếp theo.

Lời giải: Một cách trực quan nhất, ta có thể dự đoán được rằng xác suất đó chính là $\lambda = \frac{n}{N}$ Tuy nhiên, là một người muốn biết ngọn ngành vấn đề, bạn có thể chưa cảm thấy thuyết phục, và muốn biết liệu có cơ sở toán học vững chắc hơn chứng minh việc đó không.

⇒ Việc này có thể thực hiện bằng Maximum Likelihood.

Thật vậy, giả sử λ là xác suất để nhận được một mặt ngửa. Đặt $x_1, x_2, \dots, x_N$ là các đầu ra nhận được, trong đó có n giá trị bằng 1 tương ứng với mặt ngửa và $m = N - n$ và giá trị bằng 0 tương ứng với mặt sấp. Ta có thể suy ra ngay:

$$\sum_{i=1}^{N} x_i = n, \approx N - \sum_{i=1}^{N} x_i = N - n = m$$

Có thể nhận thấy việc nhận được mặt $ngửa$ hay $sấp$ khi tung đồng xu tuân theo Bernoulli Distribution:

$$P(x_i|\lambda) = \lambda^{x_i}(1-\lambda)^{1-x_i}$$

Khi đó tham số mô hình λ có thể được đánh giá bằng việc giải bài toán tối ưu:

$$\lambda = \arg\max_{\lambda}[p(x_1, x_2, x_3, \dots, x_N \mid \lambda)] \quad (4)$$

$$= \arg\max_{\lambda}\left[\prod_{i=1}^{N} p(x_i \mid \lambda)\right] \quad (5)$$

$$= \arg\max_{\lambda}\left[\prod_{i=1}^{N} \lambda^{x_i}(1-\lambda)^{1-x_i}\right] \quad (6)$$

$$= \arg\max_{\lambda}\left[\lambda^{\sum_{i=1}^{N} x_i}(1-\lambda)^{N-\sum_{i=1}^{N} x_i}\right] \quad (7)$$

$$= \arg\max_{\lambda}[\lambda^n(1-\lambda)^m] \quad (8)$$

$$= \arg\max_{\lambda}[n\log(\lambda) + m\log(1-\lambda)] \quad (9)$$

Ở trên, giả sử rằng kết quả của mỗi lần tung đồng xu là độc lập với nhau
Từ (8) sang (9) ta đã lấy log của hàm mục tiêu.
Tới đây, bài toán tối ưu (9) có thể được giải bằng cách lấy đạo hàm của hàm mục tiêu bằng 0. Tức λ là nghiệm của phương trình:

$$\frac{n}{\lambda} - \frac{m}{1-\lambda} = 0 \leftrightarrow \frac{n}{\lambda} = \frac{m}{1-\lambda} \leftrightarrow \lambda = \frac{n}{n+m} = \frac{n}{N}$$

⇒ Vậy kết quả trên là có cơ sở.

2.3 Categorical Distribution (Phân Phối Đa Thức)

Bài toán: giả sử tung một viên xúc xắc 6 mặt có xác suất rơi vào các mặt có thể không đều nhau. Giả sử trong N lần tung, số lượng xuất hiện các mặt thứ nhất, thứ hai, …, thứ sáu lần lượt là $n_1, n_2, \dots, n_6$ lần với $\sum_{i=1}^{6} n_i = N$. Tính xác suất rơi vào mỗi mặt ở lần tung tiếp theo.

Lời giải: Bài toán này có vẻ phức tạp hơn bài toán trên một chút, nhưng ta cũng có thể dự đoán được đánh giá tốt nhất của xác suất rơi vào mặt thứ i là $\lambda_i = \frac{n_i}{N}$

Mã hoá mỗi quan sát đầu ra thứ i bởi một vector 6 chiều $x_i \in \{0,1\}^6$ trong đó các phần tử của nó bằng 0 trừ phần tử tương ứng với mặt quan sát được là bằng 1. Ta cũng có thể suy ra:

$$\sum_{i=1}^{N} x_i^j = n_j, \forall j = 1,2,\dots,6$$

trong đó x_i^j là thành phần j của vector x_i.

Có thể thấy rằng xác suất rơi vào mỗi mặt tuân theo Categorical Distribution với các tham số $\lambda_j > 0, j = 1,2,\dots,6$ (bỏ qua $\lambda_j = 0$). Ta dùng λ để thể hiện cho cả 6 tham số này. Với các tham số này, xác suất để sự kiện x_i xảy ra là:

$$P(x_i|\lambda) = \prod_{j=1}^{6} \lambda_j^{x_i^j}$$

Khi đó, vẫn với giả sử về sự độc lập giữa các lần tung xúc xắc, đánh giá bộ tham số λ dựa trên Maximum log-likelihood ta có:

$$\begin{aligned}
\lambda &= \arg\max_{\lambda}[P(x_1,\dots,x_N|\lambda)] \\
&= \arg\max_{\lambda}[\textstyle\prod_{i=1}^{N} P(x_i|\lambda)] \quad (10) \\
&= \arg\max_{\lambda}\left[\textstyle\prod_{i=1}^{N}\prod_{j=1}^{6} \lambda_j^{x_i^j}\right] \quad (11) \\
&= \arg\max_{\lambda}\left[\textstyle\prod_{j=1}^{6} \lambda_j^{\sum_{i=1}^{N} x_i^j}\right] \quad (12) \\
&= \arg\max_{\lambda}\left[\textstyle\prod_{j=1}^{6} \lambda_j^{n_j}\right] \quad (13) \\
&= \arg\max_{\lambda}\left[\textstyle\sum_{j=1}^{6} n_j \log(\lambda_j)\right] \quad (14)
\end{aligned}$$

Khác với bài toán (9) một chút, chúng ta không được quên điều kiện $\sum_{j=1}^{6} \lambda_j = 1$. Vậy ta có bài toán tối ưu có ràng buộc:

$$\max_{\lambda} \textstyle\sum_{j=1}^{6} n_j \log(\lambda_j) \text{ subject to: } \sum_{j=1}^{6} \lambda_j = 1 \quad (15)$$

Bài toán tối ưu này có thể được giải bằng phương pháp nhân tử Lagrange.

Lagrangian của bài toán này là:

$$\zeta(\lambda, \mu) = \sum_{j=1}^{6} n_j \log(\lambda_j) + \mu \left(1 - \sum_{j=1}^{6} \lambda_j \right)$$

Nghiệm của bài toán là n_0 của hệ đạo hàm của ζ theo từng biến bằng 0:

$$\frac{\partial \zeta(\lambda,\mu)}{\partial \lambda_j} = \frac{n_j}{\lambda_j} - \mu = 0, \forall j = 1,2,\dots,6 \ (16)$$

$$\frac{\partial \zeta(\lambda,\mu)}{\partial \mu} = 1 - \sum_{j=1}^{6} \lambda_j = 0 \ (17)$$

Từ (16) ta có $\lambda_j = \frac{n_j}{\mu}$. Thay vào (17):

$$\sum_{j=1}^{6} \frac{n_j}{\mu} = 1 \Rightarrow \mu = \sum_{j=1}^{6} n_j = N$$

Từ đó ta có đánh giá:

$$\lambda_j = \frac{n_j}{N}, \ \forall j = 1,2,\dots,6$$

Qua hai ví dụ trên ta thấy Maximum Likelihood cho kết quả hợp lý.

2.4 Univariate Normal Distribution (Phân Phối Chuẩn Đơn Chiều)

Bài toán: Khi thực hiện một phép đo, giả sử rằng rất khó để có thể đo chính xác độ dài của một vật. Thay vào đó, người ta thường đo vật đó nhiều lần rồi suy ra kết quả, với giả thiết rằng các phép đo là độc lập với nhau và kết quả mỗi phép đo là một phân phối chuẩn. Đánh giá chiều dài của vật đó.

Lời giải: Vì ta biết rằng kết quả phép đo tuân theo phân phối chuẩn nên ta sẽ cố gắng đi xây dựng phân phối chuẩn đó. Chiều dài của vật có thể được coi là giá trị mà hàm mật độ xác suất đạt giá trị cao nhất. Trong phân phối chuẩn, ta biết rằng đó chính là kỳ vọng của phân phối đó. Chú ý rằng kỳ vọng của phân phối và kỳ vọng của dữ liệu quan sát được có thể không bằng nhau, chúng chỉ xấp xỉ bằng nhau khi mà số lượng phép đó là một số rất lớn.

Tuy nhiên, nếu đánh giá kỳ vọng của phân phối như cách làm dưới đây sử dụng Maximum Likelihood, ta sẽ thấy rằng kỳ vọng của dữ liệu chính là đánh giá tốt nhất cho kỳ vọng của phân phối.
Thật vậy, giả sử các kích thước quan sát được là $x_1, x_2, \dots, x_N$. Ta cần đi tìm một phân phối chuẩn, tức một giá trị kỳ vọng μ và phương sai σ^2, sao cho các giá trị $x_1, x_2, \dots, x_N$ là *likely* nhất.

Ta đã biết rằng, hàm mật độ xác suất tại x_icủa một phân phối chuẩn có kỳ vọng μ và phương sai σ^2 là:

$$P(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Vậy, để đánh giá μ và σ, ta sử dụng Maximum likelihood với giả thiết rằng kết quả các phép đo là độc lập:

$$\mu, \sigma = \arg\max_{\mu,\sigma}[\prod_{i=1}^{N} p(x_i|\mu, \sigma^2)] \quad (18)$$

$$= \arg\max_{\mu,\sigma}\left[\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right)\right] \quad (19)$$

$$= \arg\max_{\mu,\sigma}\left[-N\log(\sigma) - \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right] \triangleq \arg\max_{\mu,\sigma} J(\mu, \sigma) \quad (20)$$

Ta đã lấy log của hàm bên trong dấu ngoặc vuông của (19) để được (20), phần hằng số có chứa 2π cũng được bỏ đi vì nó không ảnh hưởng tới kết quả.

Một lần nữa, để tìm μ và σ, ta giải hệ phương trình đạo hàm của $J(\mu, \sigma)$ theo mỗi biến bằng 0:

$$\frac{\partial J}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i - \mu) = 0 \quad (21)$$

$$\frac{\partial J}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (22)$$

Từ đó ta có:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad (23)$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} \quad (24)$$

Đây chính là công thức đánh giá hai giá trị kỳ vọng và phương sai của dữ liệu mà chúng ta quen dùng.

2.5 Multivariate Normal Distribution (Phân Phối Chuẩn Đa Chiều)

Bài toán: Giả sử tập dữ liệu ta thu được là các giá trị nhiều chiều $\boldsymbol{x_1, \dots, x_N}$. Giả sử thêm rằng dữ liệu này tuân theo phân phối chuẩn nhiều chiều. Hãy đánh giá các tham số, vector kỳ vọng μ và ma trận hiệp phương sai Σ, của phân phối này dựa trên Maximum Likelihood, giả sử rằng các $\boldsymbol{x_1, \dots, x_N}$ là độc lập.

Lời giải: Sử dụng công thức:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \Sigma = \frac{1}{N} \sum_{i=1}^{N} (x - \mu)(x - \mu)^T$$

3 Ứng Dụng Vào Machine Learning

MLE là nền tảng của Machine Learning, một trong số đó là Logistic Regression (Hồi quy Logistic). Kỹ thuật này có khả năng tối đa hoá để phân loại dữ liệu. Hãy xem Logistic Regression sử dụng MLE như thế nào. Các quy trình MLE cụ thể có ưu điểm là chúng có thể khai thác các đặc tính của bài toán ước lượng để mang lại hiệu quả tốt hơn và độ ổn định về mặt số học. Những phương pháp này thường có thể tính toán khoảng tin cậy xác định. Tham số "bộ giải" của Logistic Regression được sử dụng để chọn các chiến lược giải khác nhau nhằm phân loại nhằm xây dựng MLE tốt hơn.

3.1 Mã Giả (Pseudocode)

```
def gaussian_pdf(x, mu, sigma):
    return (1 / math.sqrt(2 * math.pi * sigma*2)) * math.exp(-(x -
mu)**2 / (2 * sigma**2))

def log_likelihood(data, mu, sigma):
    log_likelihood_val = 0

    for x in data:
        log_likelihood_val += math.log(gaussian_pdf(x, mu, sigma))
    return log_likelihood_val

def maximum_likelihood_estimation(data, initial_mu, initial_sigma,
epsilon, max_iter):
    best_log_likelihood = float('-inf')
    best_mu = initial_mu
    best_sigma = initial_sigma

    for _ in range(max_iter):
        # Tính giá trị log-likehood cho tham số hiện tại
        current_log_likelihood = log_likelihood(data, best_mu,
best_sigma)

        # Cập nhật thông số khi giá trị log_likelihood hiện tại lớn
hơn
        if current_log_likelihood > best_log_likelihood:
            best_log_likelihood = current_log_likelihood

        # Tạo ra ngẫu nhiên giá trị trung bình và độ lệch chuẩn mới
        new_mu = random.gauss(best_mu, epsilon)
        new_sigma = random.gauss(best_sigma, epsilon)

        # Tính toán log-likelihood cho các tham số mới
        new_log_likelihood = log_likelihood(data, new_mu,
new_sigma)

        # Cập nhật thông số khi giá trị log_likelihood mới lớn hơn
        if new_log_likelihood > best_log_likelihood:
            best_log_likelihood = new_log_likelihood
            best_mu = new_mu
            best_sigma = new_sigma
```

```
    return best_mu, best_sigma
```

3.2 Bài minh hoạ

3.2.1 Nhập Thư Viện

```
1.import numpy as np
2.import pandas as pd
3.import seaborn as sns
4.from sklearn.linear_model import LogisticRegression
5.from sklearn.model_selection import train_test_split
6.from sklearn import preprocessing
```

3.2.2 Đọc dữ liệu

```
7.df=pd.read_csv("Social_Network_Ads.csv")
8.df.head()
```

	User ID	Giới tính	Tuổi	Lương ước tính	Đã trả
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

Dữ liệu liên quan đến các quảng cáo trên mạng xã hội có giới tính, độ tuổi và mức lương ước tính của người dùng mạng xã hội đó. Giới tính là một cột phân loại cần được dán nhãn và mã hóa trước khi cung cấp dữ liệu cho máy học.

3.2.3 Mã hoá dữ liệu

```
9.le = preprocessing.LabelEncoder()
10.df['Giới tính']=le.fit_transform(df['Giới tính'])
```

	User ID	Giới tính	Tuổi	Lương ước tính	Đã trả	Giới tính
0	15624510	Male	19	19000	0	1
1	15810944	Male	35	20000	0	1
2	15668575	Female	26	43000	0	0
3	15603246	Female	27	57000	0	0
4	15804002	Male	19	76000	0	1

Các kết quả được mã hóa được lưu trữ trong một tính năng mới gọi là 'giới tính' để bản gốc được giữ nguyên. Bây giờ, chia dữ liệu thành training và kiểm tra để training và xác nhận máy học.

3.2.4 Phân chia dữ liệu

```
11.X=df.drop(['Đã trả','Giới tính'],axis=1)
12.y=df['Giới tính']
13.X_train, X_test, y_train, y_test = train_test_split(X, y,
14.test_size=0.30, random_state=42)
```

Tỷ lệ này được chia thành tỷ lệ 70:30 theo quy tắc tiêu chuẩn.

3.2.5 Đưa dữ liệu vào máy học

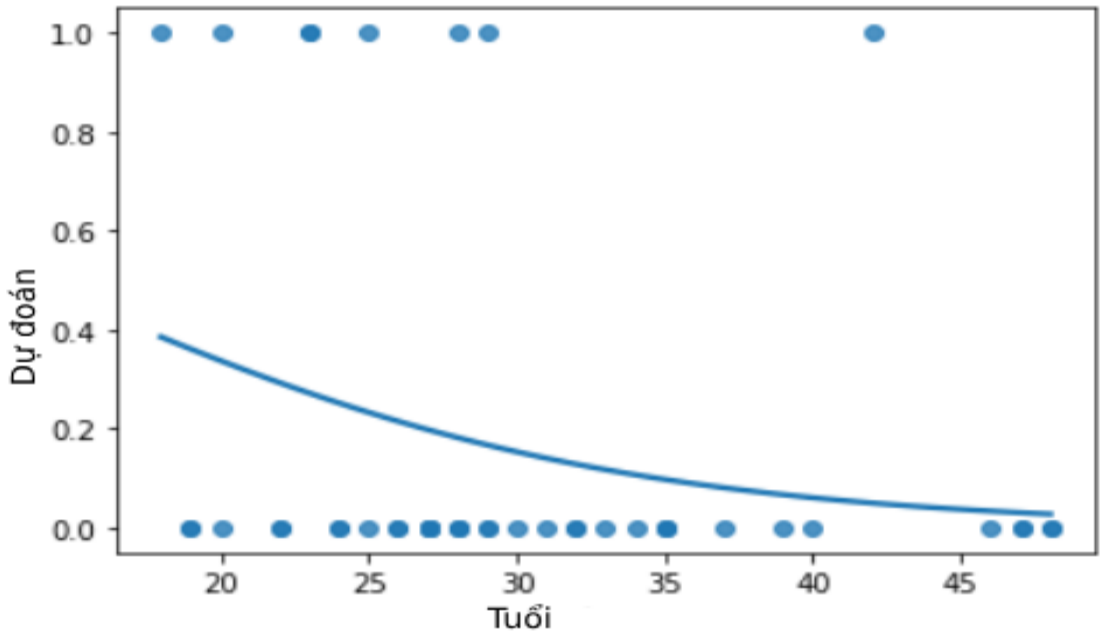
```
15.lr=LogisticRegression(max_iter=100,solver='lbfgs')
16.lr.fit(X_train,y_train)
17.lr_pred=lr.predict(X_test)
18.df_pred=pd.merge(X_test,pd.DataFrame(lr_pred,columns=['predicted
19.']),left_index=True,right_index=True)
```

	USER ID	Tuổi	Lương ước tính	Giới tính	Dự đoán
33	15776733	28	44000	0	0
93	15699284	29	28000	0	0
84	15798659	30	62000	0	0
94	15786993	29	83000	0	0
9	15727311	35	65000	0	0

Các kết quả dự đoán sẽ được thêm vào tập dữ liệu thử nghiệm với tính năng 'dự đoán'.

3.2.6 Vẽ đường máy học

```
20.sns.regplot(x="Tuổi", y='Dự đoán',data=df_pred ,logistic=True,
21.ci=None)
```



Trong biểu đồ trên nằm giữa độ tuổi đặc trưng và dự đoán, đường máy học được hình thành bằng cách sử dụng nguyên tắc ước lượng hợp lí cực đại giúp mô hình Logistic Regression phân loại kết quả. Vì vậy, thuật toán nền sẽ chọn xác suất được tính theo độ tuổi quan sát “1” và sử dụng giá trị này để tính khả năng quan sát “0”. Điều này sẽ áp dụng cho tất cả các điểm dữ liệu và cuối cùng, nó sẽ nhân tất cả các khả năng xảy ra của dữ liệu được đưa ra trong đường. Quá trình này sẽ được tiếp tục cho đến khi không tìm thấy khả năng cực đại hoặc không tìm thấy đường phù hợp nhất.

4 Tài Liệu Tham Khảo

[1]Sách Machine Learning cơ bản - Nguyễn Hữu Tiệp – Lấy các công thức đã chứng minh sẵn và cách chứng minh

*[2]*https://www.youtube.com/watch?v=XepXtl9YKwc *– Tìm hiểu cách xây dựng thuật toán*

*[3]*https://www.youtube.com/watch?v=00krscK7iBA *– nghiên cứu sơ lược*

*[4]*https://www.youtube.com/watch?v=Dn6b9fCIUpM *– nghiên cứu sơ lược*

*[5]*https://vi.wikipedia.org/wiki/Hợp_lý_cực_đại *– nghiên cứu sơ lược về Maximum Likelihood*

*[6]*https://phamdinhkhanh.github.io/deepai-book/ch_ml/NaiveBayes.html

*[7]*https://www.geeksforgeeks.org/ml-expectation-maximization-algorithm/ *- Sử dụng hình ảnh Step-by-Step, nghiên cứu thuật toán*

*[8]*https://viblo.asia/p/so-luoc-ve-maximum-likelihood-estimation-1Je5EvrYKnL *- Tìm hiểu về khái niệm*