

TRƯỜNG THPT CHUYÊN TRẦN ĐẠI NGHĨA  
NĂM HỌC 2023 – 2024

**BÁO CÁO NGHIÊN CỨU**

**MAXIMUM A POSTERIORI**

Học sinh 1: Nguyễn Quý Hạo Nhiên	Lớp: 12CTIN
Học sinh 2: Nguyễn Hoàng Minh Trức	Lớp: 12CTIN
Học sinh 3: Nguyễn Đăng Thiện Hòa	Lớp: 12CTIN
Học sinh 4: Lê Bảo Long	Lớp: 12CTIN

Thành phố Hồ Chí Minh, tháng 10 năm 2023

## MỤC LỤC

GIỚI THIỆU .....	3
NỘI DUNG .....	4
1. Lý thuyết cơ bản .....	4
2. Conjugate Prior .....	5
3. Ví dụ minh họa về tối ưu tham số .....	5
4. Thực nghiệm và chạy mô hình trên máy tính .....	7
4.1. Chia dữ liệu thành dữ liệu huấn luyện và kiểm tra .....	8
4.2. Xây dựng mô hình MAP và dự đoán .....	8
4.3. Kết quả trực quan của dữ liệu và thử nghiệm .....	9
4.4. Ảnh hưởng của việc phân phối trước .....	11
KẾT LUẬN .....	12
5. Ưu điểm và nhược điểm .....	12
5.1. Ưu điểm .....	12
5.2. Nhược điểm .....	12
TÀI LIỆU THAM KHẢO .....	13

## GIỚI THIỆU

Có rất nhiều mô hình học máy (Machine Learning) được xây dựng dựa trên các mô hình thống kê (Statistical Model). Mỗi mô hình thống kê được xây dựng dựa trên các tham số khác nhau (gọi chung là parameter). Ví dụ, phân phối Bernouli dựa trên tham số là biến  $\lambda$ . Với một mô hình thống kê bất kỳ, ký hiệu  $\theta$  là tập hợp tất cả các tham số của mô hình đó. Quá trình học (Learning) của mô hình là quá trình ước lượng (estimate) bộ tham số  $\theta$  sao cho mô hình tìm được khớp với mẫu dữ liệu nhất. Quá trình này được gọi là quá trình ước lượng tham số (Parameter estimation).

Có hai cách ước lượng chính thường được dùng trong các mô hình học máy thống kê. Cách thứ nhất chỉ dựa trên dữ liệu đã biết trong tập huấn luyện, được gọi là Maximum Likelihood Estimation hay MLE. Cách thứ hai không những dựa trên tập huấn luyện mà còn dựa trên những thông tin biết trước của các tham số, được gọi là Maximum A Posteriori Estimation hay MAP Estimation. Bài viết này sẽ tìm hiểu ý tưởng và giải quyết bài toán ước lượng tham số mô hình theo MAP Estimation.

## NỘI DUNG

### 1. Lý thuyết cơ bản

Ta có bài toán: Giả sử tung đồng xu  $N$  lần và nhận được  $n$  mặt ngửa (head) và  $m = N - n$  lần nhận được mặt sấp (tail) với quy ước các giá trị  $x_1, x_2, \dots, x_n$  với  $x_i$  có giá trị 1 khi nhận được mặt ngửa và 0 là mặt sấp. Ước lượng xác suất  $\lambda$  khi tung đồng xu nhận được mặt ngửa.

Khi ta tung đồng xu 5000 lần và nhận được 1000 lần mặt ngửa, ta có thể đánh giá xác suất của mặt ngửa là  $1/5$  và đánh giá này là đáng tin vì số lần thử là lớn. Tuy nhiên, nếu ta chỉ tung đồng xu 5 lần và nhận được 1 lần mặt ngửa thì theo MLE, xác suất của mặt ngửa cũng là  $1/5$ . Khi đó ước lượng này không còn đáng tin vì số lần tung quá ít và nhiều khả năng việc đánh giá đã bị overfitting. Khi tập huấn luyện quá nhỏ (low-training), chúng ta cần phải quan tâm tới một vài giả thiết khác của các tham số.

Maximum A Posteriori (MAP) được ra đời để khắc phục nhược điểm này của MLE. Chúng ta giới thiệu một giả thiết biết trước của tham số  $\theta$ , được gọi là *prior*. Từ giả thiết này, chúng ta có thể suy ra các khoảng giá trị và phân bố của tham số.

Trong MAP, chúng ta đánh giá các tham số như là một xác suất có điều kiện của dữ liệu:

$$\theta = \underset{\theta}{\operatorname{argmax}} \underbrace{p(\theta|x_1, \dots, x_N)}_{\text{posterior}} \quad (1.1)$$

Biểu thức  $p(\theta|x_1, \dots, x_N)$  còn được gọi là xác suất (hậu nghiệm) *posterior* của  $\theta$ . Chính vì vậy mà việc ước lượng  $\theta$  được gọi là Maximum A Posteriori.

Hàm tối ưu trong (1.1) thường khó xác định dạng một cách trực tiếp. Nhưng chúng ta thường biết điều ngược lại, tức nếu biết tham số, ta có thể tính được hàm mật độ xác suất của dữ liệu. Vì vậy, để giải bài toán MAP ta thường dùng quy tắc Bayes. Bài toán thường được biến đổi thành:

$$\theta = \underset{\theta}{\operatorname{argmax}} p(\theta|x_1, \dots, x_N) = \underset{\theta}{\operatorname{argmax}} \left[ \frac{\overbrace{p(x_1, \dots, x_N|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(x_1, \dots, x_N)}_{\text{evidence}}} \right] \quad (1.2)$$

$$= \underset{\theta}{\operatorname{argmax}} [p(x_1, \dots, x_N|\theta)p(\theta)] \quad (1.3)$$

$$= \underset{\theta}{\operatorname{argmax}} [\prod_{i=1}^N p(x_i|\theta)p(\theta)] \quad (1.4)$$

Đẳng thức (1.2) xảy ra theo quy tắc Bayes. Đẳng thức (1.3) xảy ra bởi vì mẫu số của (1.2) không phụ thuộc vào tham số  $\theta$ . Đẳng thức (1.4) xảy ra nếu chúng ta có giả thiết về sự độc lập giữa các  $x_i$ .

Tuy nhiên điều này lại mở ra một vấn đề mới: trong đa số các bài toán, chúng ta biết trước được  $p(x_1, \dots, x_N | \theta)$  nhưng  $p(\theta)$  thì không, vì vậy ta bước tiếp với khái niệm mới được đề cập phía dưới như sau.

## 2. Conjugate Prior

Ta có định nghĩa rằng: nếu phân phối xác suất của prior  $p(\theta)$  và phân phối xác suất  $p(\theta | x_1, \dots, x_n)$  có *cùng dạng (same family)* với nhau thì  $p(\theta)$  là một conjugate prior của  $P(x_1, \dots, x_n | \theta)$ . Để dễ hình dung hơn thế nào là định nghĩa của *cùng dạng (same family)* ta hãy xem xét ví dụ sau với phân phối Bernoulli có hàm mật độ xác suất như sau:

$$p(x|\lambda) = \lambda^x(1 - \lambda)^{1-x} \quad (2.1)$$

Nếu ta chọn prior  $p(\lambda)$  là một phân phối Beta có hàm mật độ xác suất sau:

$$p(\lambda) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda^{\alpha-1}(1 - \lambda)^{\beta-1} \quad (2.2)$$

Nhận xét thấy rằng khi ta bỏ qua thừa số  $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$  đi thì thấy phần còn lại của hàm Beta có công thức tương tự như công thức của phân phối Bernoulli ban đầu. Vì vậy conjugate prior của hàm Bernoulli ở trên chính là hàm Beta, do đó ta có thể suy ra được điều sau với posterior  $p(\lambda|x)$ :

$$p(\lambda|x) \propto p(x|\lambda)p(\lambda) \quad (2.3)$$

$$\Leftrightarrow p(\lambda|x) \propto \lambda^{x+\alpha-1}(1 - \lambda)^{1-x+\beta-1} \quad (2.4)$$

## 3. Ví dụ minh họa về tối ưu tham số

Ta quay trở lại ví dụ về đồng xu ở trên; ta tung đồng xu  $N$  lần và ghi nhận được  $n$  lần mặt ngửa (head) và  $m = N - n$  lần mặt sấp (tail). Nếu ta sử dụng MAP để ước lượng  $\lambda$  là xác suất nhận được mặt ngửa với prior được chọn là một phân phối  $Beta[\alpha, \beta]$  thì bài toán tối ưu sẽ trở thành:

$$\lambda = \underset{\lambda}{argmax} [p(x_1, \dots, x_N | \lambda)p(\lambda)] \quad (2.5)$$

$$= \underset{\lambda}{argmax} [(\prod_{i=1}^N \lambda^{x_i}(1 - \lambda)^{1-x_i}) \lambda^{\alpha-1}(1 - \lambda)^{\beta-1}] \quad (2.6)$$

Áp dụng quy tắc tích của các lũy thừa cùng cơ số  $\lambda$  ta được đẳng thức sau:

$$\lambda = \underset{\lambda}{argmax} [\lambda^{\sum_{i=1}^N x_i + \alpha - 1} (1 - \lambda)^{N - \sum_{i=1}^N x_i + \beta - 1}] \quad (2.7)$$

Do  $\sum_{i=1}^N x_i = n$  (ban đầu ta đã quy ước  $x_i = 1$  là các mặt ngựa) nên  $N - \sum_{i=1}^N x_i = m$  và ta có được đẳng thức sau:

$$\lambda = \underset{\lambda}{argmax} [\lambda^{n+\alpha-1} (1-\lambda)^{m+\beta-1}] \quad (2.8)$$

Đến đây, chúng ta có thể đưa hàm mục tiêu về logarit của nó bởi chúng ta có 2 tính chất sau:

- Log của một tích bằng tổng các log.
- Log là hàm đồng biến; khi giá trị bên trong Log tăng liên tục thì Log cũng tăng theo.

Vì vậy ta có biểu thức sau:

$$\lambda = \underset{\lambda}{argmax} [(n + \alpha - 1) \log \lambda + (m + \beta - 1) \log(1 - \lambda)] \quad (2.9)$$

Bằng cách lấy đạo hàm của hàm mục tiêu và cho nó bằng không, ta tìm được  $\lambda$  theo các siêu tham số (*hyperparameters*)  $\alpha$  và  $\beta$  như sau:

$$\frac{n+\alpha-1}{\lambda \ln 10} + \frac{-(m+\beta-1)}{(1-\lambda) \ln 10} = 0 \quad (2.10)$$

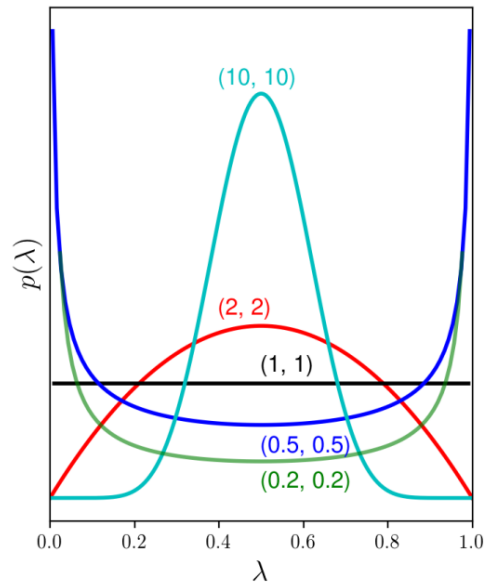
$$\Leftrightarrow \frac{(n+\alpha-1)}{\lambda} = \frac{m+\beta-1}{1-\lambda} \quad (2.11)$$

$$\Leftrightarrow (1-\lambda)(n+\alpha-1) = \lambda(m+\beta-1) \quad (2.12)$$

$$\Leftrightarrow n + \alpha - 1 = \lambda(m + n + \alpha + \beta - 2) \quad (2.13)$$

$$\Leftrightarrow \lambda = \frac{n+\alpha-1}{N+\alpha+\beta-2} \quad (2.14)$$

Ta thử nghiệm một số các cặp giá trị  $(\alpha, \beta)$  để tìm ra  $\lambda$  và vẽ hàm mật độ xác suất của  $p(\lambda)$ , nhận xét thấy khi các 2 siêu tham số  $\alpha = \beta$  và đạt giá trị lớn thì xác suất  $\lambda$  thu được càng gần 0.5 tức xác suất để tung được mặt ngựa (head) là  $\frac{1}{2}$  đúng với thực tế.



Hình 1. Hàm mật độ xác suất  $p(\lambda)$  của theo 1 vài giá trị  $\alpha$  và  $\beta$

#### 4. Thực nghiệm và chạy mô hình trên máy tính

Giả sử, ta có bộ dữ liệu là kết quả của một phân tích hóa học trên các loại rượu được trồng trong cùng một vùng ở Ý xuất phát từ ba loại cây trồng khác nhau. Bộ dữ liệu cung cấp 178 chai rượu trong tệp **WINE.CSV**. Mỗi chai sẽ thuộc 1 trong 3 loại (Type) và có 13 đặc trưng khác nhau và đã được thông qua quá trình tiền huấn luyện bao gồm:

- |                      |                                  |
|----------------------|----------------------------------|
| 1. Alcohol           | 8. Non Flavonoid phenols         |
| 2. Malic acid        | 9. Proanthocyanins               |
| 3. Ash               | 10. Color intensity              |
| 4. Alcalinity of ash | 11. Hue                          |
| 5. Magnesium         | 12. OD280/OD315 of diluted wines |
| 6. Total phenols     | 13. Proline                      |
| 7. Flavanoids        |                                  |

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Type	Alcohol	Malic Acid	Ash	Alcalinity of ash	Magnesium	Total phenols	Flavanoids	Non Flavanoids phenols	Proanthocyanins	Color intensity	Hue	OD280/OD315 of diluted wines	Proline
2	1	14.23	1.71	2.43	15.6	127	2.8	3.06	0.28	2.29	5.64	1.04	3.92	1065
3	1	13.2	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.4	1050
4	1	13.16	2.36	2.67	18.6	101	2.8	3.24	0.3	2.81	5.68	1.03	3.17	1185
5	1	14.37	1.95	2.5	16.8	113	3.85	3.49	0.24	2.18	7.8	0.86	3.45	1480
6	1	13.24	2.59	2.87	21	118	2.8	2.69	0.39	1.82	4.32	1.04	2.93	735
7	1	14.2	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	6.75	1.05	2.85	1450
8	1	14.39	1.87	2.45	14.6	96	2.5	2.52	0.3	1.98	5.25	1.02	3.58	1290
9	1	14.06	2.15	2.61	17.6	121	2.6	2.51	0.31	1.25	5.05	1.06	3.58	1295
10	1	14.83	1.64	2.17	14	97	2.8	2.98	0.29	1.98	5.2	1.08	2.85	1045
11	1	13.86	1.35	2.27	16	98	2.98	3.15	0.22	1.85	7.22	1.01	3.55	1045
12	1	14.1	2.16	2.3	18	105	2.95	3.32	0.22	2.38	5.75	1.25	3.17	1510
13	1	14.12	1.48	2.32	16.8	95	2.2	2.43	0.26	1.57	5	1.17	2.82	1280
14	1	13.75	1.73	2.41	16	89	2.6	2.76	0.29	1.81	5.6	1.15	2.9	1320
15	1	14.75	1.73	2.39	11.4	91	3.1	3.69	0.43	2.81	5.4	1.25	2.73	1150
16	1	14.38	1.87	2.38	12	102	3.3	3.64	0.29	2.96	7.5	1.2	3	1547
17	1	13.63	1.81	2.7	17.2	112	2.85	2.91	0.3	1.46	7.3	1.28	2.88	1310
18	1	14.3	1.92	2.72	20	120	2.8	3.14	0.33	1.97	6.2	1.07	2.65	1280
19	1	13.83	1.57	2.62	20	115	2.95	3.4	0.4	1.72	6.6	1.13	2.57	1130
20	1	14.19	1.59	2.48	16.5	108	3.3	3.93	0.32	1.86	8.7	1.23	2.82	1680
21	1	13.64	3.1	2.56	15.2	116	2.7	3.03	0.17	1.66	5.1	0.96	3.36	845
22	1	14.06	1.63	2.28	16	126	3	3.17	0.24	2.1	5.65	1.09	3.71	780
23	1	12.93	3.8	2.65	18.6	102	2.41	2.41	0.25	1.98	4.5	1.03	3.52	770
24	1	13.71	1.86	2.36	16.6	101	2.61	2.88	0.27	1.69	3.8	1.11	4	1035
25	1	12.85	1.6	2.52	17.8	95	2.48	2.37	0.26	1.46	3.93	1.09	3.63	1015
26	1	13.5	1.81	2.61	20	96	2.53	2.61	0.28	1.66	3.52	1.12	3.82	845
27	1	13.05	2.05	3.22	25	124	2.63	2.68	0.47	1.92	3.58	1.13	3.2	830
28	1	13.39	1.77	2.62	16.1	93	2.85	2.94	0.34	1.45	4.8	0.92	3.22	1195
29	1	13.3	1.72	2.14	17	94	2.4	2.19	0.27	1.35	3.95	1.02	2.77	1285
30	1	13.87	1.9	2.8	19.4	107	2.95	2.97	0.37	1.76	4.5	1.25	3.4	915
31	1	14.02	1.68	2.21	16	96	2.65	2.33	0.26	1.98	4.7	1.04	3.59	1035
32	1	13.73	1.5	2.7	22.5	101	3	3.25	0.29	2.38	5.7	1.19	2.71	1285

Hình 2. Bảng dữ liệu tổng quan về rượu

#### 4.1. Chia dữ liệu thành dữ liệu huấn luyện và kiểm tra

Theo yêu cầu, cần chọn ngẫu nhiên 18 chai từ mỗi loại rượu làm dữ liệu thử nghiệm (tổng cộng 54 trường hợp), phần còn lại sẽ được dùng làm dữ liệu huấn luyện. Sau đó, ta lưu hai tập dữ liệu này vào hai tệp CSV riêng biệt, được đặt tên là TRAIN.CSV và TEST.CSV.

#### 4.2. Xây dựng mô hình MAP và dự đoán

Trong phần này, mô hình phân phối được chúng tôi chọn là phân phối chuẩn (Gaussian Distribution) có hàm phân phối như sau:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Với  $\mu$  và  $\sigma$  lần lượt là trung bình và độ lệch chuẩn (Standard Deviation) của các dữ liệu của từng đặc trưng.

Ban đầu, chúng tôi tạo lớp “MAP\_Estimation” bên trong dòng code để làm mô hình cho bộ dữ liệu. Sau đó, ta đặt ba biến thành viên:

- **label:** Là một danh sách được sử dụng để ghi lại các loại nhãn trong dữ liệu gồm các loại rượu 1, 2 và 3.
- **priors:** Là một danh sách lưu trữ các giá trị ưu tiên tương ứng với từng nhãn.
- **parameters:** Là danh sách lưu trữ các giá trị  $\mu$  và  $\sigma$  là các đặc trưng trong mỗi nhãn.



```
class MAP_Estimation:
    def __init__(self) -> None:
        self.labels = []
        self.priors = {}
        self.parameters = {}
```

Hình 3. Lớp "MAP\_Estimation"

Trong "MAP\_Estimation", ta dựa trên mô hình "scikit-learn" và triển khai hai hàm, đó là "**fit(X, y)**" và "**predict(X)**".

Hàm '**fit(X, y)**' dùng để cập nhật các tham số cần thiết cho việc ước tính MAP dựa trên dữ liệu huấn luyện, bao gồm các bước sau:

- Ghi lại các nhãn có trong mảng nhãn "y" và lưu vào 'self.labels'.
- Tính toán tỷ lệ (xác suất) của từng nhãn trong dữ liệu huấn luyện và lưu trữ nó trong 'self.priors'.
- Tính giá trị trung bình  $\mu$  và độ lệch chuẩn  $\sigma$  của từng các đặc trưng X của mỗi nhãn. Các giá trị này được tính toán dựa trên dữ liệu huấn luyện.

Hàm '**predict(X)**' được sử dụng để dự đoán nhãn của dữ liệu đầu vào 'X' dựa trên ước tính MAP. Quá trình dự đoán được thực hiện như sau:

- Lặp qua các nhãn từ 'self.labels' viết tắt là 'C'
- Sử dụng công thức của MAP để tính xác suất "posterior" của từng nhãn dựa trên dữ liệu đầu vào 'X'.

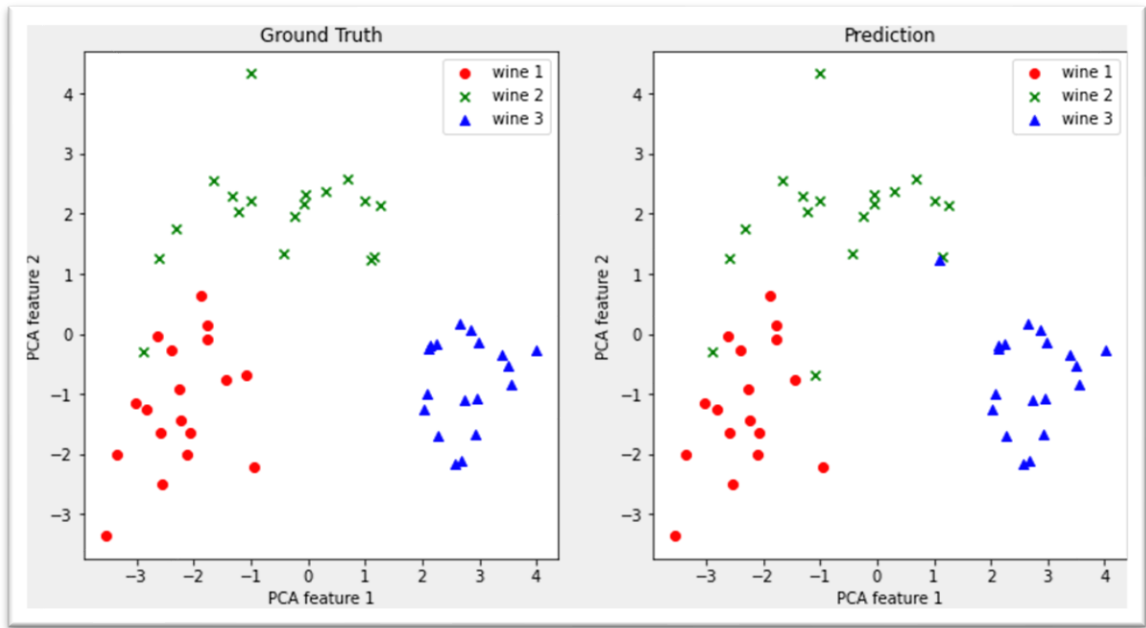
$$\operatorname{argmax}_{c \in C} P(c|X) = \frac{P(X|c)P(c)}{P(X)}$$

- Sau khi tính toán xác suất 'posterior' của từng nhãn, nhãn có xác suất 'posterior' lớn nhất được chọn làm kết quả dự đoán.

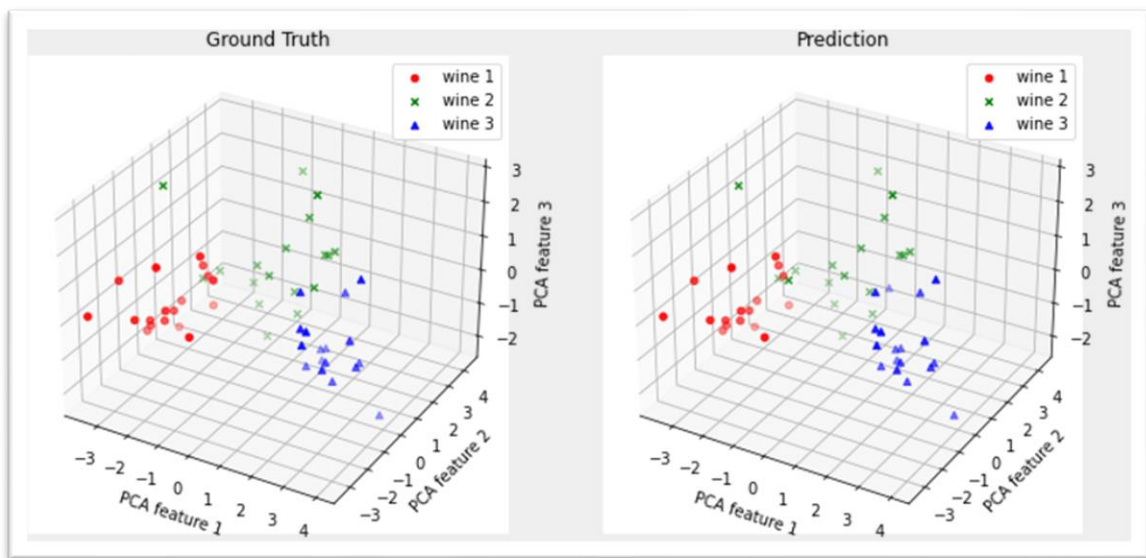
Giả sử, nhãn "1" cho xác suất lớn nhất thì X sẽ thuộc nhãn "1" hay nói cách khác là loại rượu "1".

### 4.3. Kết quả trực quan của dữ liệu và thử nghiệm

Trong phần này, ta sử dụng kỹ thuật giảm chiều "PCA" (Phân tích thành phần chính) từ thư viện 'scikit-learn' để giảm dữ liệu thử nghiệm từ 13 chiều xuống 2 hoặc 3 chiều. Từ đó, trực quan hóa để nhận thấy các điểm dữ liệu phân phối trong không gian chiều thấp, thông qua "Ground Truth" (thực tế) và "Prediction" (dự đoán).



Hình 5. Phân phối dữ liệu trong không gian 2 chiều



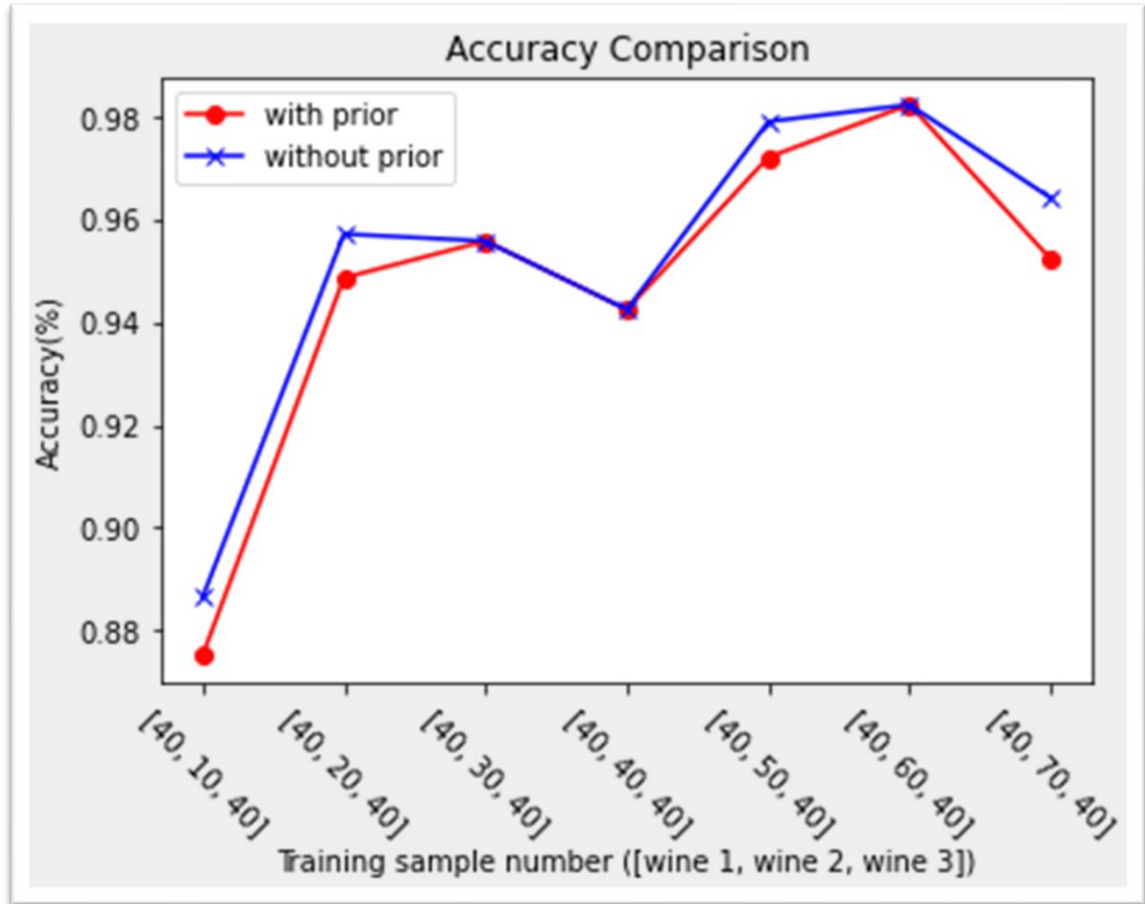
Hình 4. Phân phối dữ liệu trong không gian 3 chiều

Dựa trên biểu đồ kết quả, các điểm dữ liệu vẫn có xu hướng tự nhóm lại với nhau theo các danh mục hoặc lớp khác nhau. Các điểm dữ liệu cùng một nhãn có xu hướng gần nhau hơn trong không gian giảm chiều, và điều này hỗ trợ tính hiệu quả của mô hình MAP trong việc phân loại dữ liệu.

Quan sát kết quả giữa "Ground Truth" (thực tế) và "Prediction" (dự đoán), ta cũng nhận thấy rằng độ chính xác của mô hình không đạt 100%, mô hình không thể phân loại hoàn toàn đúng.

#### 4.4. Ảnh hưởng của việc phân phối trước

Để kiểm tra xem phân phối trước có ảnh hưởng đến phần sau hay không, ta viết một lớp mới có tên “**MAP\_Estimation\_Without\_Prior**” dựa trên “**MAP\_Estimation**” và thay đổi phương pháp tính toán xác suất “posterior” bằng cách chỉ sử dụng khả năng (likelihood), tăng số lượng dữ liệu huấn luyện cho rượu 1 và rượu 3 lên 40, sau đó tăng số lượng dữ liệu huấn luyện cho rượu 2 từ 10 lên 70.



Hình 6. So sánh độ chính xác

Từ biểu đồ trên, ta nhận thấy rằng, độ chính xác của mô hình nếu không có mức độ ưu tiên (without prior) sẽ bằng hoặc cao hơn việc có mức độ ưu tiên (with prior).

## KẾT LUẬN

### 5. Ưu điểm và nhược điểm

#### 5.1. Ưu điểm

Hiệu quả cho dữ liệu hạn chế: MAP hoạt động tốt khi **dữ liệu hạn chế**. Khi dữ liệu huấn luyện ít, thông tin tiên nghiệm có thể giúp cải thiện độ chính xác của dự đoán.

**Ứng dụng trong học máy:** MAP thường được sử dụng trong học máy để ước tính tham số của các mô hình thống kê, chẳng hạn như ước tính tham số **trong mô hình Gaussian Naive Bayes và Bayesian Linear Regression**.

#### 5.2. Nhược điểm

Phụ thuộc vào giả định: MAP đòi hỏi các giả định về xác suất tiên nghiệm và hàm likelihood. Nếu các **giả định này không chính xác, kết quả dự đoán có thể bị sai lệch**.

Yêu cầu kiến thức trước đó: Để sử dụng MAP hiệu quả, cần có kiến thức trước đó về dữ liệu và kiến thức về xác suất. Việc xác định xác suất tiên nghiệm đòi hỏi sự hiểu biết về tình huống cụ thể.

Khó khăn trong việc ước tính tham số: Trong một số trường hợp, **ước tính các xác suất tiên nghiệm và hàm likelihood có thể khó khăn và đòi hỏi tính toán phức tạp**.

**Dễ bị overfitting:** Nếu có ít dữ liệu quan sát, MAP có thể dẫn đến overfitting (quá khớp) nếu không kiểm soát được xác suất tiên nghiệm.

## TÀI LIỆU THAM KHẢO

- [1] Pattern Reconition and Machine Learning. Christopher Bishop.
- [2] *Conjugate Prior Explained*. (2020, 01 08). Được truy lục từ <https://towardsdatascience.com/conjugate-prior-explained-75957dc80bfb>
- [3] jun94. (2020, 03 26). *Maximum Likelihood(ML) and Maximum A Posteriori(MAP) Estimation*. Được truy lục từ <https://medium.com/jun94-devpblog/ml-1-maximum-likelihood-ml-and-maximum-a-posteriori-map-estimation-4f9927897daf>
- [4] Tiệp, V. H. (2018). Machine Learning Cơ Bản. Vũ Hữu Tiệp.