1. **Theories and Related Method**
   1. **Covariance Matrices**

ความแปรปรวนร่วมเกี่ยว(Covariance) เป็นการวัดปริมาณการเปลี่ยนแปลงของสองตัวแปรว่าจะมีการเปลี่ยนแปลงตามกันมาน้อยเท่าใด ในการคำนวณความแปรปรวนของของตัวแปรสามารถคำนวณได้จาก

เมทริกความแปรปรวนร่วมเกี่ยว(Covariance Matrix) เป็นเมทริกที่คำนวณ Covariance ที่มีมากกว่า 2 ตัวแปรให้อยู่ในรูปของเมทริก โดย ซึ่งคำนวณได้ในรูปของเมทริก

โดยผลลัพธ์ที่ได้จะอยู่ในรูปของเมทริกที่รวม Variance ของแต่ละคู่ตัวแปร ดังเมทริกต่อไปนี้

* 1. **Multivariate normal distribution**

การแจกแจงปรกติ(Normal distribution) เป็นการแจกแจงความน่าจะเป็นของค่าของตัวแปรสุ่มที่เป็นค่าแบบต่อเนื่อง โดยที่ค่าของตัวแปรสุ่มมีแนวโน้มที่จะมีค่าอยู่ใกล้ ๆ กับค่า ๆ หนึ่ง กราฟแสดงค่าฟังก์ชันความหนาแน่น (probability density function) จะเป็นรูปคล้ายระฆังคว่ำ หรือเรียกว่า Gaussian function

การแจกแจงแบบปกติหลายตัวแปร(Multivariate normal distribution) เป็นการแจกแจงความน่าจะเป็นของค่าหลายตัวแปร โดยมีสมการดังนี้

* 1. **Bayes’ Theorem**

ทฤษฎีของเบย์(Bayes’ theorem) เป็นทฤษฎีที่เป็นส่วนขยายของความน่าจะเป็นแบบมีเงื่อนไข หรืออาจกล่าวว่าเป็นการหาความน่าจะเป็นของส่วนย่อยจากเหตุการณ์ที่สนใจหนึ่งที่เกิดขึ้นแล้ว โดยมีสมการดังนี้

กฏของเบย์(Bayes’ rule) ที่ใช้ในการทดลองนี้จะเปรียบเทียบว่า sample ที่เข้ามานั้นมีโอกาสอยู่ในคลาสไหนมากที่สุด กล่าวคือจะเปรียบเทียบ และ หาก มีค่ามากกว่า จะถือว่า อยู่ใน ในทางกลับกันหาก มีค่าน้อยกว่า จะถือว่า จะอยู่ใน

* 1. **Cross-validation Test**

การวัดประสิทธิภาพด้วยวิธี Cross-validation นี้จะทําการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) เช่น 5-fold cross-validation คือ ทําการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจํานวนข้อมูลเท่ากัน หรือ 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจํานวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทําวนไปเช่นนี้จนครบจํานวนที่แบ่งไว้ ซึ่งในการทดลองนี้จะใช้ 10-fold cross-validation

1. **Algorithm (Pseudocode)**

อันดับแรกจะทำการ shuffle ข้อมูลทั้งหมด จากนั้นจะทำการทดลองโดยจะทำการแบ่ง 10-fold cross-validation จากข้อมูล 200 records กล่าวคือ ข้อมูลสำหรับหา Covariance, Mean, P(w) ของคลาสใดๆ มี 180 และสำหรับทดสอบตามกฏของเบย์ 20 records โดยจะนำข้อมูลสำหรับทดสอบไปหาค่าของการแจกแจงแบบปกติหลายตัวแปร ของแต่ละคลาส จากนั้นจะทำการเปรียบเทียบตามกฏของเบย์และวัดผลลัพธ์ ดัง Pseudocode นี้

1 Shuffle Data  
2 For **Cross validation k=1 to 10**:

3 Find **Covariance , mean**  and P each population

, , , , and

4 Find **Multivariate normal distribution each sample**

and

5 Compare with **Bayes’ Rule**

6 Evaluate by **Confusion Matrix** and **Error**

7 End For

1. **Analyze data**

ข้อมูลจะประกอบไปด้วย 4 features มีทั้งหมด 200 records ซึ่งได้ถูกแบ่งกลุ่ม(class) ไว้ทั้ง หมด 2 class ได้แก่ class 1 และ class 2 ดังตารางนี้

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Feature 1** | **Feature 2** | **Feature 3** | **Feature 4** | **Class** |
| 1 | 5.91 | 3.09 | 3.74 | 1.17 | 1 |
| … | … | … | … | … | … |
| 200 | 6.29 | 3.22 | 5.77 | 1.59 | 2 |

1. **ผลการทดลอง**

|  |  |  |
| --- | --- | --- |
| **K = 1** | | |
| **Evaluation Information** | | |
| **Data set** | **Test - I** | **Test - II** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Confusion Matrix** |  |  |
| **Error estimates** |  |  |
| **Test vector** | | |
|  |  | |
|  |  | |
|  |  | |

1. **วิเคราะห์ผลการทดลอง**