Data

ขนาดของข้อมูล

- 1D จะมีข้อมูลความยาวหรือความกว้างอย่างใดอย่างหนึ่ง
- 2D จะมีขนาดของ กว้าง X ยาว
- 3D จะเอาข้อมูลของ 2D มาวางซ้อนกัน
- 4D จะเอาข้อมูลของ 3D มาวางซ้อนกัน

คุณสมบัติของ matrix
- แนวตั้ง : colum จะบอกถึงคุณสมบัติของข้อมูล
- แนวนอน : แถว จะบอกถึงข้อมูลแต่ละชุด
ประเภทของชุดข้อมูล : การบันทึกข้อมูล
- Relational records(เชิงสัมพันธ์)
- Data matrix e.g. numerice matrix crosstabs (เมทริกซ์ข้อมูล เช่น เมทริกซ์ตัวเลข)
- Transaction data (ข้อมูลธุรกรรม)
- Document data : Term-frequency vector (matrix) of text documents เมทริกซ์ของเอกสารข้อความ

ประเภทของชุดข้อมูล : กราฟและเครือข่าย
- Transportation (เครือข่ายขนส่ง)
- World Wide Web
- Moleclar Structures (โครงสร้างโมเลกุล)
- Social or information networks (เครือข่ายสังคมหรือข้อมูล)

ประเภทของชุดข้อมูล : Ordered Data
- Video data : sequence of images (ลำดับภาพ)
- Temporal data time-series (อนุกรมเวลา)
- SEquential Data transation SeQuences (ลำดับการทำธุรกรรม)
- Genetic sequence data (ข้อมูลลำดับพันธุกรรม)

ประเภทของชุดข้อมูล : ข้อมูลเชิงพื้นที่ ภาพ และมัลติมีเดีย
- Spatial data : maps (แผนที่)
- Image gata (การนำภาพมาซ้อนกันหลายๆรูป)

ลักษณะสำคัญของข้อมูล

- Dimensionality มิติของข้อมูล

- Sparstity การเก็บเฉพาะคู่อันดับ ไม่เก็บตัวเลข

- Resolution ความละเอียดในการเก็บข้อมูล

- Distridution การกระจายตัวของข้อมูล

ประเภทของข้อมูล

1. Nominal

ข้อมูลประเภท nominal คือข้อมูลที่ไม่ใช่ตัวเลข หรือ เชิงปริมาณ (quantitative) ดังนั้นข้อมูลประเภทนี้จึงไม่สามารถนำมาคำนวณ หรือ เปรียบเทียบในทางคณิตศาสตร์ได้ เราอาจจะเรียกข้อมูลชนิดนี้ว่าเป็น “ป้าย” หรือ “ฉลาก” (label) ที่เอาไว้กำกับชื่อของสิ่งใดๆ วิธีจำให้ง่ายก็คือ nominal ก็คือ name มันคือป้ายชื่อดีๆนี่เอง ตัวอย่างของข้อมูลประเภทนี้ เช่น เพศชาย เพศหญิง สีดำ สีเขียว ประเทศไทย หรือ รองเท้า เป็นต้น ในบางครั้งเราอาจจะได้ยินคำว่า binominal หรือ dichotomous ซึ่งแปลว่ามีข้อมูลได้แค่ 2 ค่า เช่น ชาย/หญิง ใช่/ไม่ใช่ เป็นต้น

2. Ordinal

Ordinal มาจากคำว่า “order” ซึ่งแปลว่าข้อมูลนั้นสามารถนำมาเรียงลำดับ หรือ ให้ค่าความสำคัญเป็นลำดับได้ แต่ค่าความสำคัญนั้นไม่สามารถตีค่าออกมาเป็นตัวเลขได้ ตัวอย่างที่เห็นชัดเช่นการที่เราแบ่งระดับของ “ความสุข” ออกเป็นลำดับขึ้น (scale) เช่น Very Unhappy, Unhappy, OK, Happy, และ Very Happy เราทราบว่า Happy ต้องมีความสุขมากกว่า OK แต่เราไม่สามารถบอกออกมาเป็นค่าตัวเลขได้ว่า ค่าความสุขของ Happy กับค่าความสุขของ OK นั้นห่างกันเท่าไหร่ นอกจากนี้เรายังไม่สามารถนำค่าความสุขของช่วงต่างๆมาเปรียบเทียบกันในเชิงตัวเลขได้ เช่น เราไม่สามารถบอกได้ว่าค่าความต่างของ Very Happy และ Happy จะมากกว่า/น้อยกว่า/หรือเท่ากับ ค่าความต่างของ Happy และ OK

3. Interval

Interval ก็คือข้อมูลแบบตัวเลข ซึ่งเราสามารถนำค่ามาจัดเรียงลำดับ และสามารถบอกค่าความต่างของแต่ละตัวเลขได้ เช่น 5 มีค่ามากกว่า 3 และ 5 ห่างจาก 3 อยู่ 2 หน่วย (5-3 = 2) คำว่า interval แปลว่า ช่วง หรือ ระยะห่างระหว่างจุดสองจุด (space in between) คุณสมบัติที่สำคัญของข้อมูลแบบ interval อีกอันหนึ่งก็คือข้อมูลประเภทนี้ไม่มี “true zero” ซึ่งหมายถึง ค่า 0 ไม่ได้แปลว่าไม่มีข้อมูล แต่หมายถึงข้อมูลมีค่าเท่ากับ 0 เช่น คำว่าอุณภูมิเท่ากับ 0 องศา ไม่ได้แปลว่าไม่มีค่าของอุณภูมิ แต่แปลว่า อุณภูมิมีเท่ากับ 0 ดังนั้น ถึงแม้ว่าเราจะสามารถบวกและลบข้อมูลประเภทนี้เพื่อหาความแตกต่างระหว่างค่าต่างๆได้ เราจะไม่สามาณคูณและหารข้อมูลประเภทนี้ได้เนื่องจากไม่มี true zero

4. Ratio

ข้อมูลประเภทนี้คือ ‘สุดยอด’ ของประเภทข้อมูลเนื่องเราสามารถใช้งานข้อมูลประเภทนี้ได้อย่างครบถ้วน ตั้งแต่นำมาจัดเรียงลำดับ หาค่าความต่าง และมีสิ่งที่เรียกว่า “true zero” หรือ “absolute zero” (ศูนย์สัมบูรณ์) อยู่ด้วย ดังนั้น

ข้อมูลประเภทนี้จึงสามารถนำมาคำนวณได้อย่างหลากหลายและถูกนำมาใช้ในสถิติทั้งแบบพรรณา (descriptive) และเพื่อการอนุมาน (inferential) การมี true zero ทำให้เราสามารถทำการคูณและหารข้อมูลประเภทนี้ได้