

ขนาดของข้อมูล

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

คุณสมบัติของ matrix

- แนวตั้ง : column 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)

มาตรา (scale) ของข้อมูลแบ่งออกเป็น 4 ระดับ คือ

- มาตรานามบัญญัติ (nominal scale) เป็นมาตรการวัดที่จำแนกข้อมูลออกเป็นกลุ่ม โดยข้อมูลที่อยู่ในกลุ่มเดียวกันจะมีลักษณะเหมือนกัน และไม่สามารถเปรียบเทียบความแตกต่างระหว่างกลุ่มได้ว่ามีค่ามากน้อยเพียงใด ตัวอย่างเช่น เพศ อาชีพ วิชา เป็นต้น การกำหนดตัวเลขให้กับข้อมูลแต่ละกลุ่มนั้นสามารถทำได้ อาทิเช่น 1=เพศชาย 2=เพศหญิง เป็นต้น
- มาตราเชิงอันดับ (ordinal scale) เป็นมาตรการวัดที่จำแนกข้อมูลออกเป็นกลุ่ม โดยข้อมูลที่อยู่ในกลุ่มเดียวกันจะมีลักษณะเหมือนกัน และสามารถเปรียบเทียบความแตกต่างระหว่างกลุ่มได้ว่ามีค่ามากน้อยเพียงใด ทำให้สามารถจัดเรียงลำดับข้อมูลแต่ละกลุ่มได้ ตัวอย่างเช่น ส่วนสูง ระดับการศึกษา เป็นต้น การกำหนดตัวเลขให้กับข้อมูลแต่ละกลุ่มสามารถทำได้ อาทิเช่น 1=ช่วงอายุแรกเกิด – 10 ปี 2=ช่วงอายุ 11 – 20 ปี 3=ช่วงอายุ 21-30 ปี 4=ช่วงอายุ ตั้งแต่31 ปีขึ้นไป เป็นต้น จะเห็นได้ว่าสามารถจัดเรียงลำดับกลุ่มได้ทั้งจากน้อยไปมากได้แก่ 1 2 3 4 หรือจากมากไปน้อยได้แก่ 4 3 2 1
- มาตราอันตรภาค (interval scale) เป็นมาตรการวัดของข้อมูลเชิงปริมาณ นั่นคือสามารถวัดค่าของข้อมูลเป็นตัวเลขได้ และสามารถบอกปริมาณของข้อมูลได้ว่ามีค่ามากน้อยเพียงใด ข้อมูลระดับมาตราอันตรภาคเป็นข้อมูลที่ไม่มีศูนย์แท้ (absolute zero) นั่นคือค่าของข้อมูลที่มีค่าเป็น 0 ไม่มีความหมายว่าข้อมูลประเภทนั้นไม่มีค่าเลย ตัวอย่างเช่น น้ำอุณหภูมิ 0 องศาเซลเซียส ไม่ได้หมายความว่าน้ำไม่มีอุณหภูมิเลย เป็นต้น
- มาตราอัตราส่วน (ratio scale) เป็นมาตรการวัดของข้อมูลเชิงปริมาณเหมือนมาตราอันตรภาค แต่ข้อมูลระดับมาตราอันตรภาคเป็นข้อมูลที่มีศูนย์แท้ (absolute zero) นั่นคือค่าของข้อมูลที่มีค่าเป็น 0 จะมีความหมายว่าข้อมูลประเภทนั้นไม่มีค่าเลย ตัวอย่างเช่น นาย ก มีเงินติดตัวอยู่ 0 บาท แสดงว่า ณ ขณะนั้นนาย ก ไม่มีเงินติดตัวเลย เป็นต้น

ลักษณะสำคัญของข้อมูล

- Dimensionality มิติของข้อมูล
- Sparsity การเก็บเฉพาะคู่อันดับ ไม่เก็บตัวเลข
- Resolution ความละเอียดในการเก็บข้อมูล
- Distribution การกระจายตัวของข้อมูล