Types of Data

UNIVERSITY OF
LIVERPOOL

Two general classes of data

- **Nondependency-oriented data:** objects do not have dependencies

ID	Height (cm)	Weight (kg)	Marital status	Employed
1	175	60	Single	Yes
2	168	80	Married	Yes
3	183	85	Single	No
4	178	65	Divorced	Yes
5	194	90	Married	No
6	185	78	Married	Yes

- **Dependency-oriented data:** implicit or explicit dependencies between objects may exist
  - Networks: nodes (objects) are connected by edges (relationships)
  - Successive measurements collected from a sensor

Nondependency-oriented data (multidimensional data)

- The simplest form of data
- A multidimensional data set $\mathscr{D}$ typically contains a set of records $\overline{X_1}, \ldots, \overline{X_n}$
- Each record $\overline{X_i}$ containing a set of d features $(x_i^1, \ldots, x_i^d)$
- This data set can be represented by an $n \times d$ data matrix

Types of data

- **Numerical** or **quantitative** (values have natural ordering)
  - integer values (number of petals in a flower)
  - real values (length of a petal)

- **Categorical** or **unordered discrete-valued**
  - discrete unordered values/categories (colour of a flower petal)

- **Binary data** (two values: 0 and 1)
  - Can be seen as a categorical data (two categories) or a numerical data (0<1)
  - Can be used to represent **Set Data** via characteristic vectors

- **Text data**
  - Document as a **string** (dependency-oriented data type)
  - Document as a **set of words** or **terms** (vector-space representation: frequencies of the words in the document)

Dependency-oriented data

- **Implicit dependencies**
  - Are not explicitly specified but are known to exist
  - **Example**: temperature values collected by a sensor
- **Explicit dependencies**
  - Graphs or network data (edges specify explicit relationships)

Types of data with implicit dependencies

- **Time-series**
  - values that are generated by sequential measurements over time (**time-stamp** or **index value** is a contextual attribute; the measurement is behavioral attribute)

- **Discrete Sequences** and **Strings**
  - The categorical analog of time-series data

- **Spatial data** (every record has a location attribute)
  - **Example**: temperature, pressure are measured at spacial locations

- **Spatiotemporal data** (contain both **spatial** and **temporal** attributes)

Types of data with explicit dependencies

- **Network/Graph data**
  - Objects correspond to nodes of the network
  - Relationships between the objects correspond to the edges of the network
  - Edges may be directed or undirected
  - A set of attributes may be associated with a node
  - A set of attributes may be associated with an edge
  - Examples:
    - Web graph
    - Facebook/Instagram/LinkedIn social networks