# 資料科學入門
# numpy套件

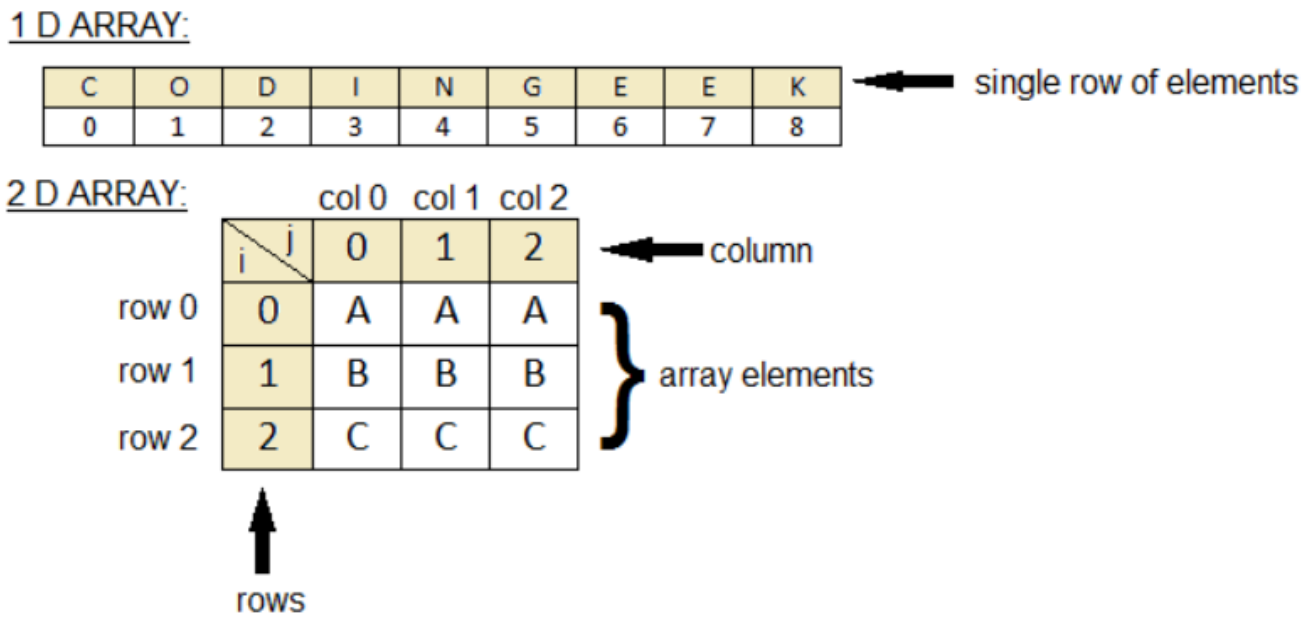
林彥廷

# 什麼是 numpy？

- numpy 是 Python 的一個重要模組
- numpy主要用於資料處理，能快速操作多重維度的陣列
- Python 處理龐大資料時，其原生 list 效能表現並不理想（但可以動態存異質資料）
- numpy 具備平行處理的能力，可以將操作動作一次套用在大型陣列上
- Python 其餘重量級的資料科學相關套件（例如：Pandas、SciPy、Scikit-learn 等）都幾乎是奠基在 numpy 的基礎上

# numpy 基礎操作

- numpy 陣列



# numpy 基礎操作

- numpy 陣列
  - numpy 的重點在於陣列的操作，其所有功能特色都建築在同質且多重維度的 ndarray（N-dimensional array）上
  - ndarray 的關鍵屬性是維度（ndim）、形狀（shape）和數值類型（dtype）

```python
# 引入 numpy 模組
import numpy as np
np1 = np.array([1, 2, 3])
np2 = np.array([3, 4, 5])
# 陣列相加
print(np1 + np2) # [4 6 8]
# 顯示相關資訊
print(np1.ndim, np1.shape, np1.dtype) # 1 (3,) int64 => 一維陣列, 三個元素, 資料型別
```

# numpy 基礎操作

- numpy 陣列
  - 改變陣列維度：使用reshape()

```
np3 = np.array([1, 2, 3, 4, 5, 6])
np3 = np3.reshape([2, 3])
print(np3.ndim, np3.shape, np3.dtype) # 2 (2, 3) int64
```

# numpy 基礎操作

- numpy 陣列
  - 改變陣列型別（bool、int、float、string）
  - bool 可以包含 True、False，int 可以包含 int16、int32、int64。其中數字是指 bits。float 可以包含 16、32、64 表示小數點後幾位。string 可以是 string、unicode。nan 則表示遺失值。

```
np3 = np.array([1, 2, 3, 4, 5, 6])
np3 = np3.reshape([2, 3])
print(np3.ndim, np3.shape, np3.dtype) # 2 (2, 3) int64
np3 = np3.astype('int32')
np3.dtype
# dtype('int32')
```

# numpy 基礎操作

- numpy 陣列
  - 建立填滿 0 或 1 的陣列

```
np1 = np.zeros([2, 3]) # array([[ 0.,  0.,  0.], [ 0.,  0.,  0.]])
np2 = np.ones([2, 3]) # array([[ 1.,  1.,  1.], [ 1.,  1.,  1.]])
```

# numpy 基礎操作

- numpy 陣列
  - 使用布林遮罩來取值

```
np3 = np.array([1, 2, 3, 4, 5, 6])
print(np3 > 3) # [False False False  True  True  True]
print(np3[np3 > 3]) # [4 5 6]
```

# numpy 基礎操作

- numpy 陣列
  - 元素加總

```
np3 = np3.reshape([2, 3])
print(np3.sum(axis=1)) # 將 axis=1 列(橫向)加總 [6 15]

np3 = np3.reshape([2, 3])
print(np3.sum(axis=0)) # 將 axis=0 行(縱向)加總 [5 7 9]
```

ytlin@mail.nptu.edu.tw