# 数据操作-数据预处理（4-6）

1. **N维数组——机器学习&神经网络的主要数据结构**
   1. 例子
      1. 0维：标量（一个类别）
      2. 1维：向量（特征向量）

机器学习中默认特征向量为列向量，行为一个样本，列为一个特征

* + 1. 2维：矩阵（样本-特征矩阵）
    2. 3维：RGB图片（宽[列数]\*高[行数]\*通道）

通道指R,G,B三个通道，RGB色彩空间，就是把一种颜色，用RGB三个分量表达出来。通道在此处与颜色接近，R：Red，G：Grey，B：Blue。
[RGB图像之灰度级和通道的理解\_三维矩阵 r g b-CSDN博客](https://blog.csdn.net/silence2015/article/details/53789748)​

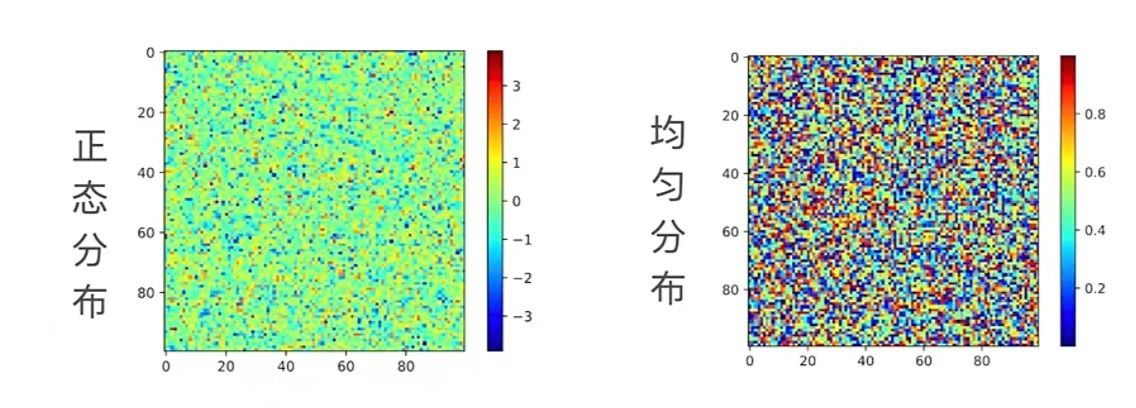
* + 1. 4维：RGB图片批量（批量大小\*宽\*高\*通道）

批量大小即一次处理的图片个数

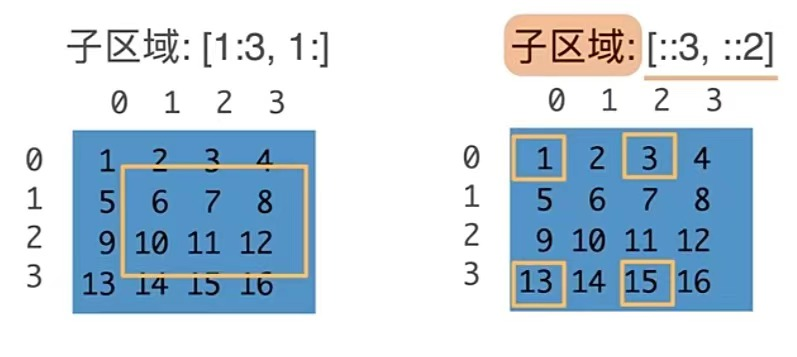
* + 1. 5维：一个视频批量（批量大小\*时间\*宽\*高\*通道）

一个视频的维度取决于其特征提取和表示方式，如基于图像帧的向量表示，基于特征提取的向量表示和基于深度学习的向量表示等

* 1. 创建数组的必要条件



* + 1. 形状（维度，每个维度的长度）
    2. 每个元素的数据类型（eg：32位浮点数）
    3. 每个元素的值
  1. 访问元素
     1. 一个元素：[row\_number,column\_number]
     2. 一行：[row\_number,:]
     3. 一列：[:,column\_number]
     4. 子区域：[row\_range,column\_range]



* + - 1. 连续取：[1:3,1:]表示取2-3行的第一列以后的所有区域。注意此处的区间是开区间，3并不会被取到（左闭右开）（python中从0开始）
      2. 间隔取：[::3,::2]表示从第0行开始到第3行结束，间隔两行取一次；从第0列开始到第2列结束，间隔两列取一次。（:重复了两次所以间隔数是2）

[start:end:step],start默认为起点，end默认为最后一个，步长默认为1