Tensor Layouts

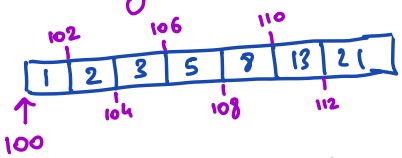
Array/Vector (1D)

Logical view

Shape: [7]

Stride: [1]

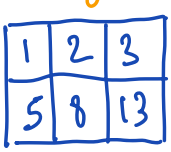
Physical view



In C, we get a pointer to the location in memory where the pointer starts and we use it to access all the rest of the elements

Matrix (2D)

Logical view



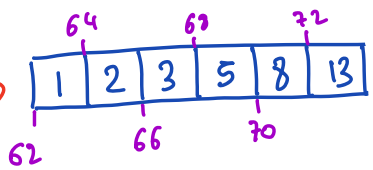
= T

Shape: [2, 3]

Stride: [3, 1]

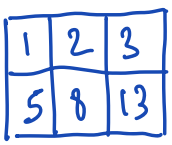
Note: this is known as a row-major layout.

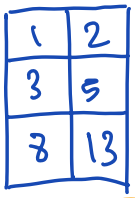
Physical view



To go from one row to the next, we need to skip three elements

Matrix (2D) - Reshape

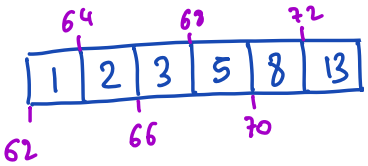




Shape: [2, 3]
Stride: [3, 1]

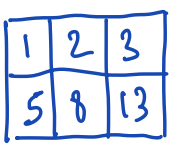
Shape: [3, 2]
Stride: [2, 1]

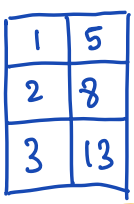
Physical view



Note: physical arrangement does not change!

Matrix (2D) - Transpose

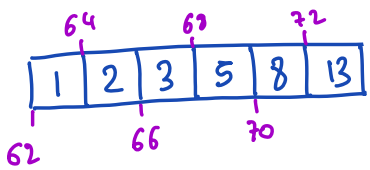




Shape: [2, 3]
Stride: [3, 1]

Shape: [3, 2]
Stride: [1, 3]

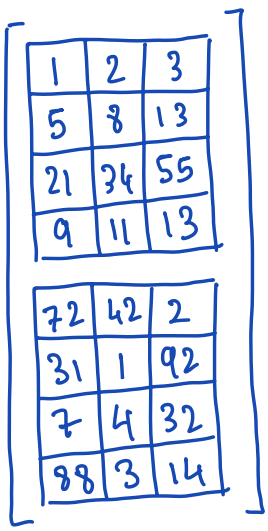
Physical view



To transpose along two dimensions, we need to swap the shape and stride of the two dimensions ... and that's it!

Note: after transposing, the tensor is no longer "contiguous". that's why in PyTorch you can't "view" but you need to "reshape".

Tensor (3D)



Physical view:

Shape: $[2, 4, 3]$

Stride: $[12, 3, 1]$

$$\text{Stride} = \begin{cases} \text{Stride}[i] = \prod_{j=i+1}^{N} \text{Shape}[j] \\ \text{Stride}[N] = 1 \end{cases}$$