Deep Learning

A course about theory & practice

UNIVERSITÀ DI PAVIA

Tensor Broadcasting

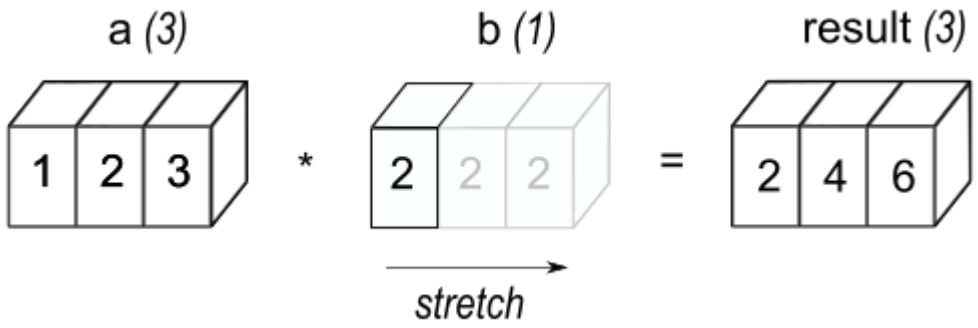
Marco Piastra

Broadcasting in general

PyTorch, TensorFlow, Jax and all adopt the general broadcasting rules of NumPy

- When a tensor is broadcast, its entries are **conceptually copied**

  *Broadcasting is a performance optimization, thus **no actual copying occurs***



[image from https://numpy.org/doc/stable/user/basics.broadcasting.html]

The general broadcasting rule

- When operating on two arrays, NumPy compares their shapes element-wise
  It starts with the **trailing** dimensions, and works its way backward

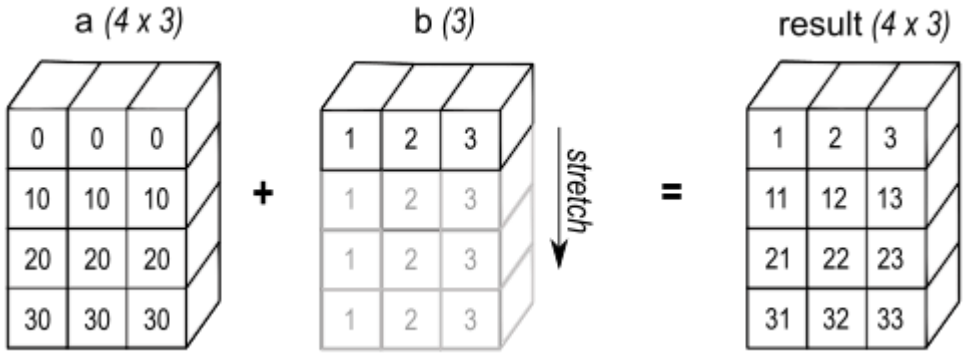
Two dimensions are **compatible** when

1. they are *equal*, or
2. one of them is 1

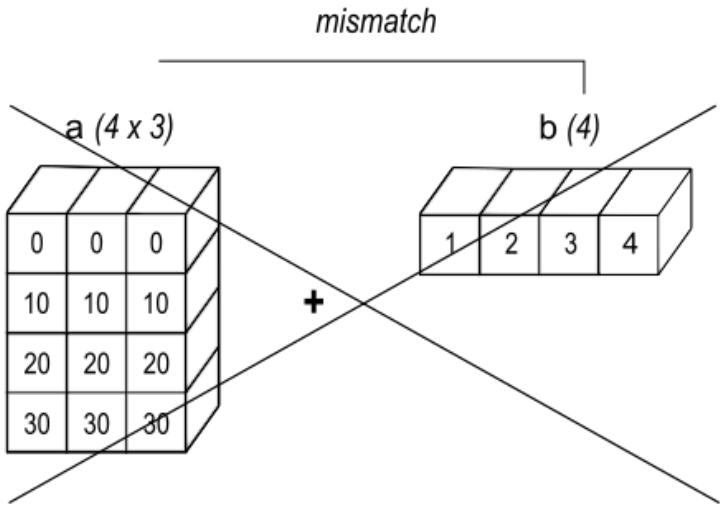
- The size of the resulting array is the **maximum size** along each dimension of the input arrays

```
a      (2d array):   5 x 4
b      (1d array):       1
result (2d array):   5 x 4

a      (3d array):  15 x 3 x 1
b      (2d array):      3 x 5
result (3d array):  15 x 3 x 5

a      (4d array):   8 x 1 x 6 x 5
b      (3d array):       7 x 1 x 5
result (4d array):   8 x 7 x 6 x 5
```

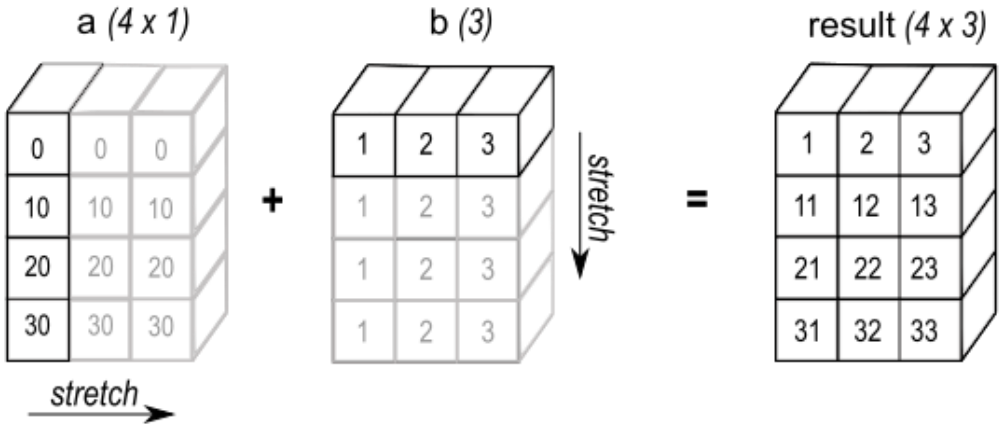
Broadcasting: a few examples



a (2d array): 4 x 3
b (1d array): 3
result (3d array): 4 x 3



a (2d array): 4 x 3
b (1d array): 4
result: error!



a (1d array): 4 x 1
b (1d array): 3
result (3d array): 4 x 3

[images from https://numpy.org/doc/stable/user/basics.broadcasting.html]