1. How does unsqueeze help us to solve certain broadcasting problems?
2. How can we use indexing to do the same operation as unsqueeze?
3. How do we show the actual contents of the memory used for a tensor?
4. When adding a vector of size 3 to a matrix of size 3×3, are the elements of the vector added to each row or each column of the matrix? (Be sure to check your answer by running this code in a notebook.)
5. Do broadcasting and expand\_as result in increased memory use? Why or why not?
6. Implement matmul using Einstein summation.
7. What does a repeated index letter represent on the lefthand side of einsum?
8. What are the three rules of Einstein summation notation? Why?
9. What are the forward pass and backward pass of a neural network?
10. Why do we need to store some of the activations calculated for intermediate layers in the forward pass?
11. What is the downside of having activations with a standard deviation too far away from 1?
12. How can weight initialization help avoid this problem?