L02: The Mathematical Building Blocks of Neural Networks

CSci 560 Deep Learning w/ Python (Chollet) Ch. 2

Derek Harter

Department of Computer Science
East Texas A&M University

Summer 2025

L02.1 A First Look at a Neural Network

CSci 560: Deep Learning w/ Python (Chollet) Ch. 2.1

Derek Harter

Department of Computer Science
East Texas A&M University

Summer 2025

L02.2 Data Representations for Neural Networks

CSci 560: Deep Learning w/ Python (Chollet) Ch. 2.2

Derek Harter

Department of Computer Science
East Texas A&M University

Summer 2025

L02.3 The Gears of Neural Networks: Tensor Operations

CSci 560: Deep Learning w/ Python (Chollet) Ch. 2.3

Derek Harter

Department of Computer Science
East Texas A&M University

Summer 2025

Matrix · Matrix Dot-Product

- `x`, `y` are input matrices of shape `(a,b)` and `(b, c)` respectively
- Each row of `x` performs an inner product with corresponding column of `y`
- So resulting element `z[i,j]` is the inner product between row `i` of `x` and column `j` of `y`
- Resulting shape of `z` is `(a,c)`

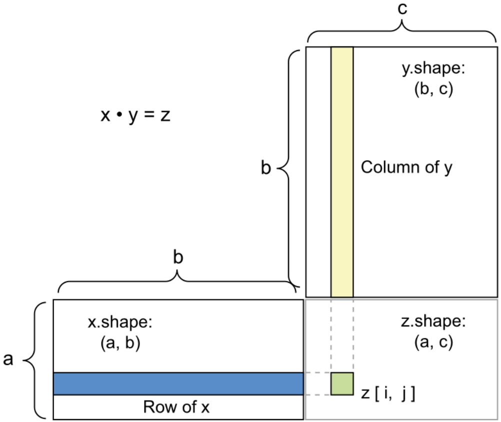

Figure 1: Matrix dot-product box diagram. (Chollet, 2021, pg.43)

Vector Addition is Translation

- Adding a vector to a point will move the point by a fixed amount.
- Applied to a set of points, result is a translation.

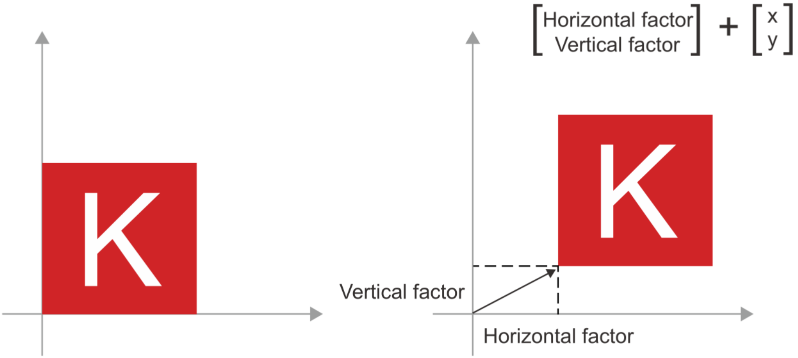

Figure 2: 2D translation as vector addition. (Chollet, 2021, pg.45)

Matrix Multiplication (dot product) is Rotation

- A counterclockwise rotation by angle `theta` achieved by dot product.

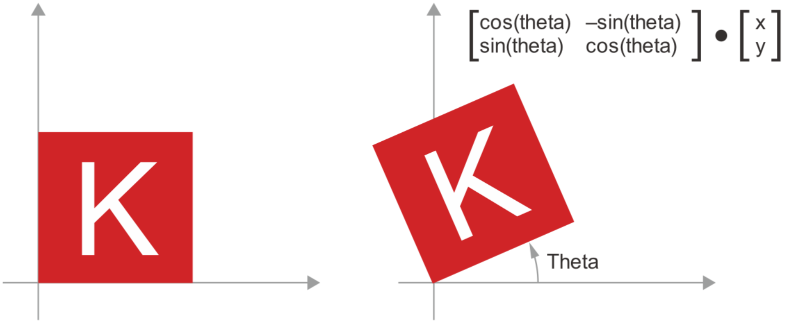

Figure 3: 2D rotation (counterclockwise) as a dot product. (Chollet, 2021, pg.45)

Matrix Multiplication can also Scale

- Vertical and horizontal scaling also achieved by dot product with a suitable matrix.

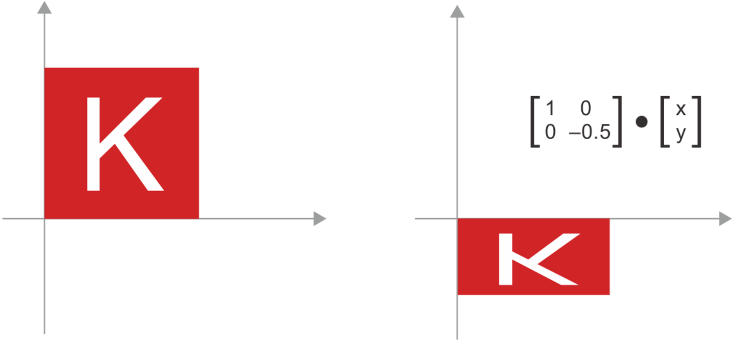

Figure 4: 2D scaling as a dot product. (Chollet, 2021, pg.46)

Affine Transformation: Linear transform and translation

- Translation, Rotation, Scaling are all linear transformations.
- Can combine, for example $y = W \cdot x + b$ is a linear transformation plus a translation.

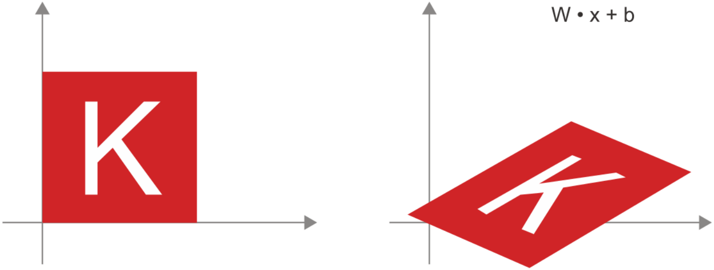

Figure 5: Affine transform in the plane. (Chollet, 2021, pg.46)

`relu` Nonlinear Activation Function

- The previous are just combinations of linear transformations.
- Multilayer NN made entirely of linear transformations can be reduced to a single linear transformation.
- Activation functions produce a nonlinear translation.

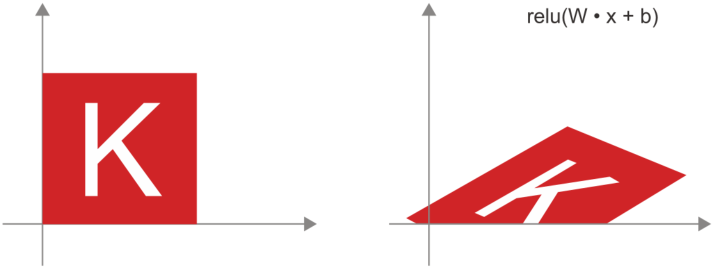

Figure 6: Affine transform followed by (nonlinear) ‘relu‘ activation. (Chollet, 2021, pg.47)

A Geometric Interpretation of Deep Learning

- NN consist entirely of chains of tensor operations.
- Can be interpreted as geometric transformations of the input data into a new space/shape/manifold.
- Can interpret a NN as a very complex geometric transformation in a high-dimensional space, implemented via a series of simple linear + nonlinear transformations.

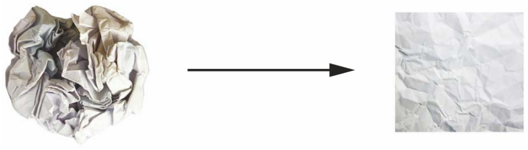

Figure 7: Uncrumpling a complicated manifold of data. (Chollet, 2021, pg.47)

L02.4 The Engine of Neural Networks: Gradient-based Optimization

CSci 560: Deep Learning w/ Python (Chollet) Ch. 2.4

Derek Harter

Department of Computer Science
East Texas A&M University

Summer 2025

L02.5 Looking Back at our First Neural Network Example

CSci 560: Deep Learning w/ Python (Chollet) Ch. 2.5

Derek Harter

Department of Computer Science
East Texas A&M University

Summer 2025

Bibliography

Chollet, F. (2021). *Deep learning with python* (second). Manning.