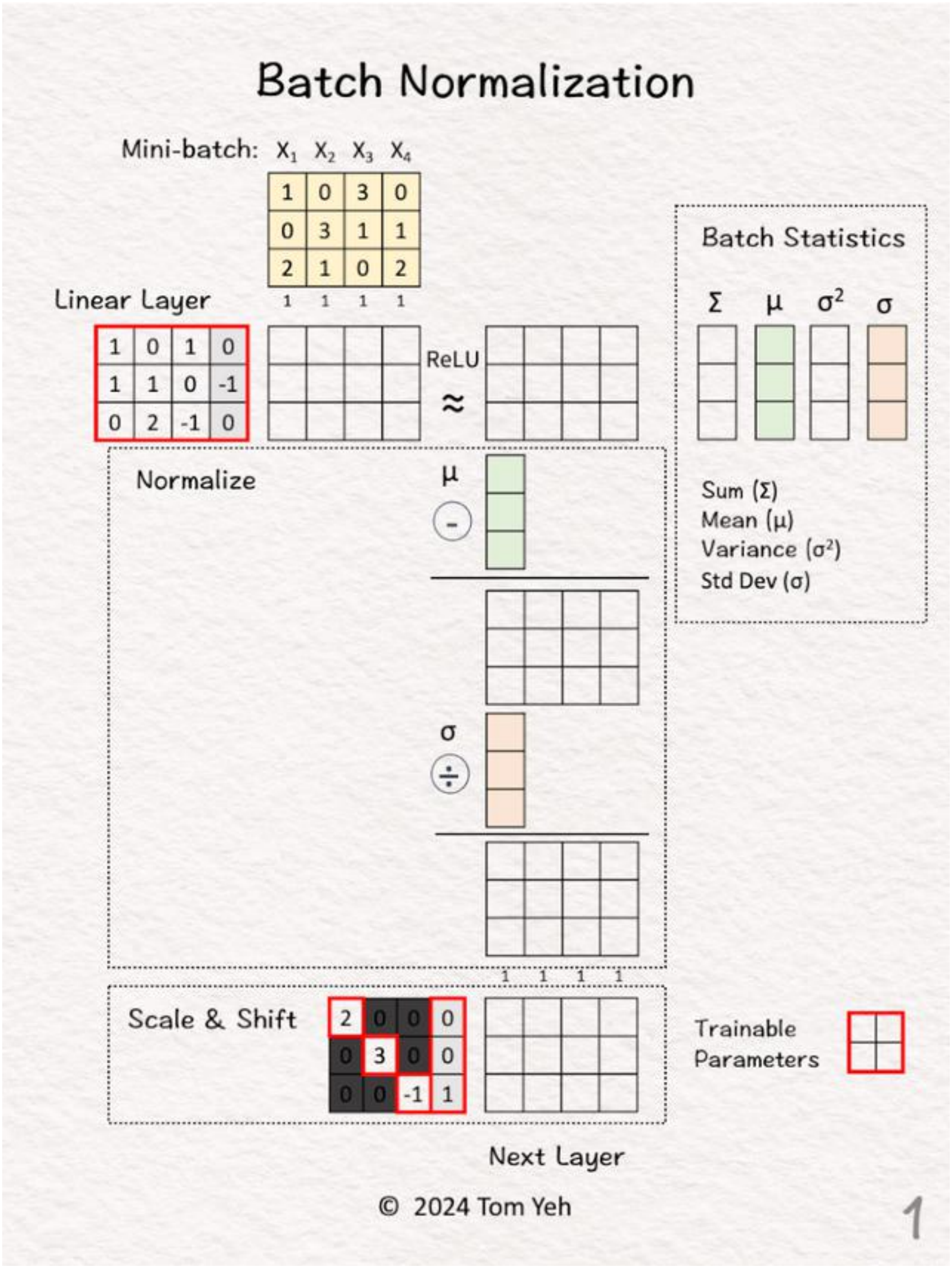
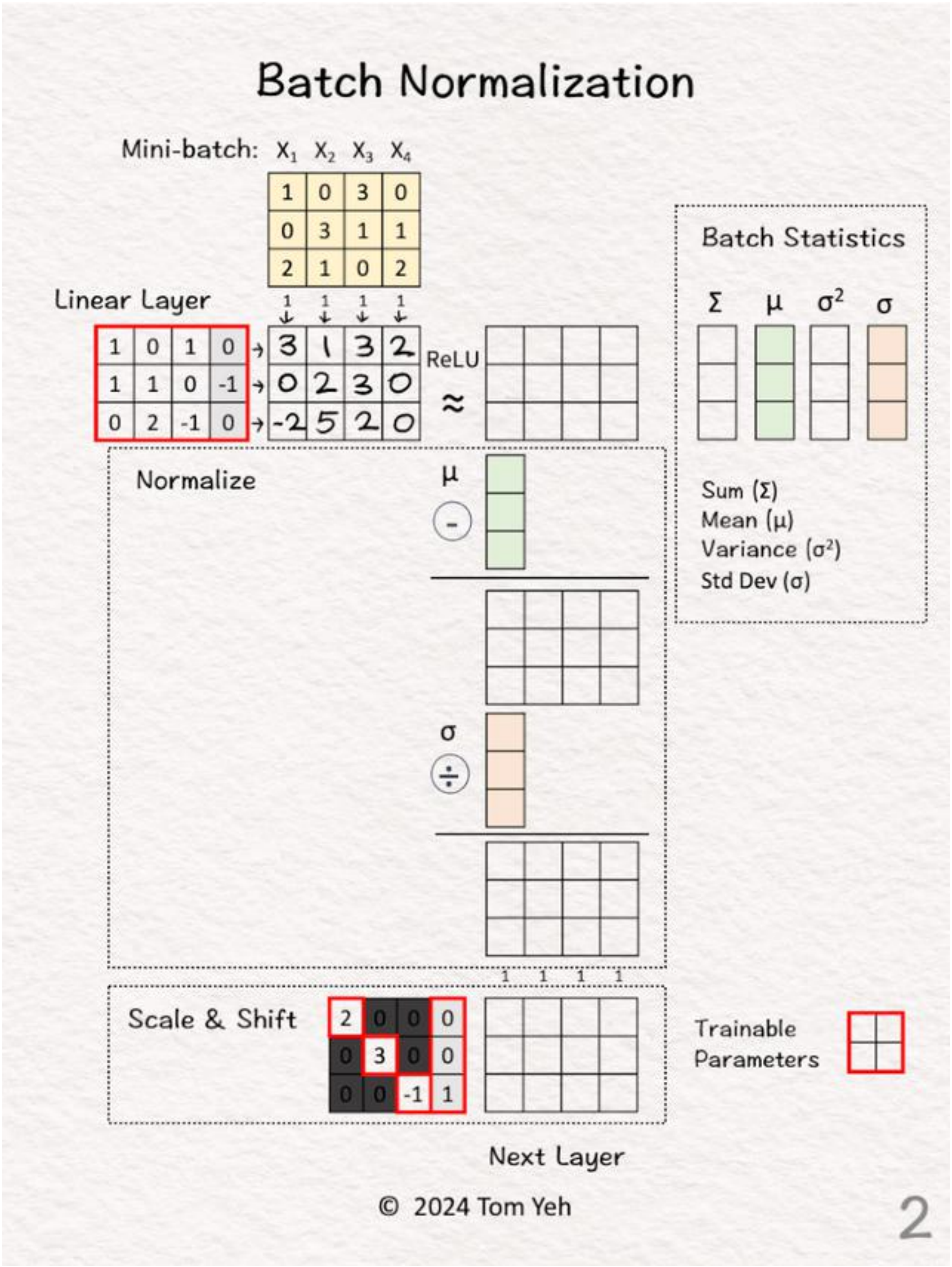


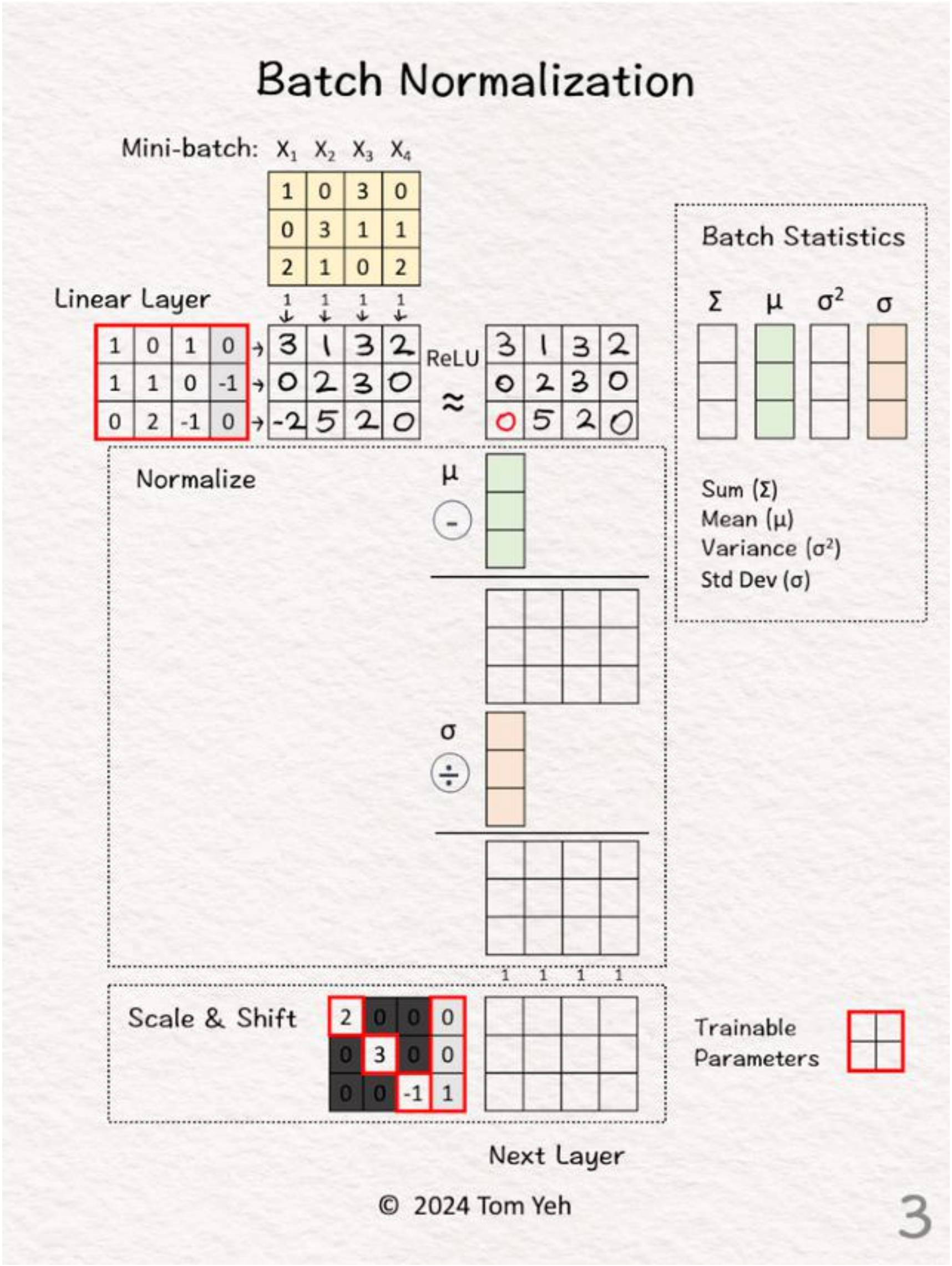

[1] Given

A mini-batch of 4 training examples, each has 3 features.



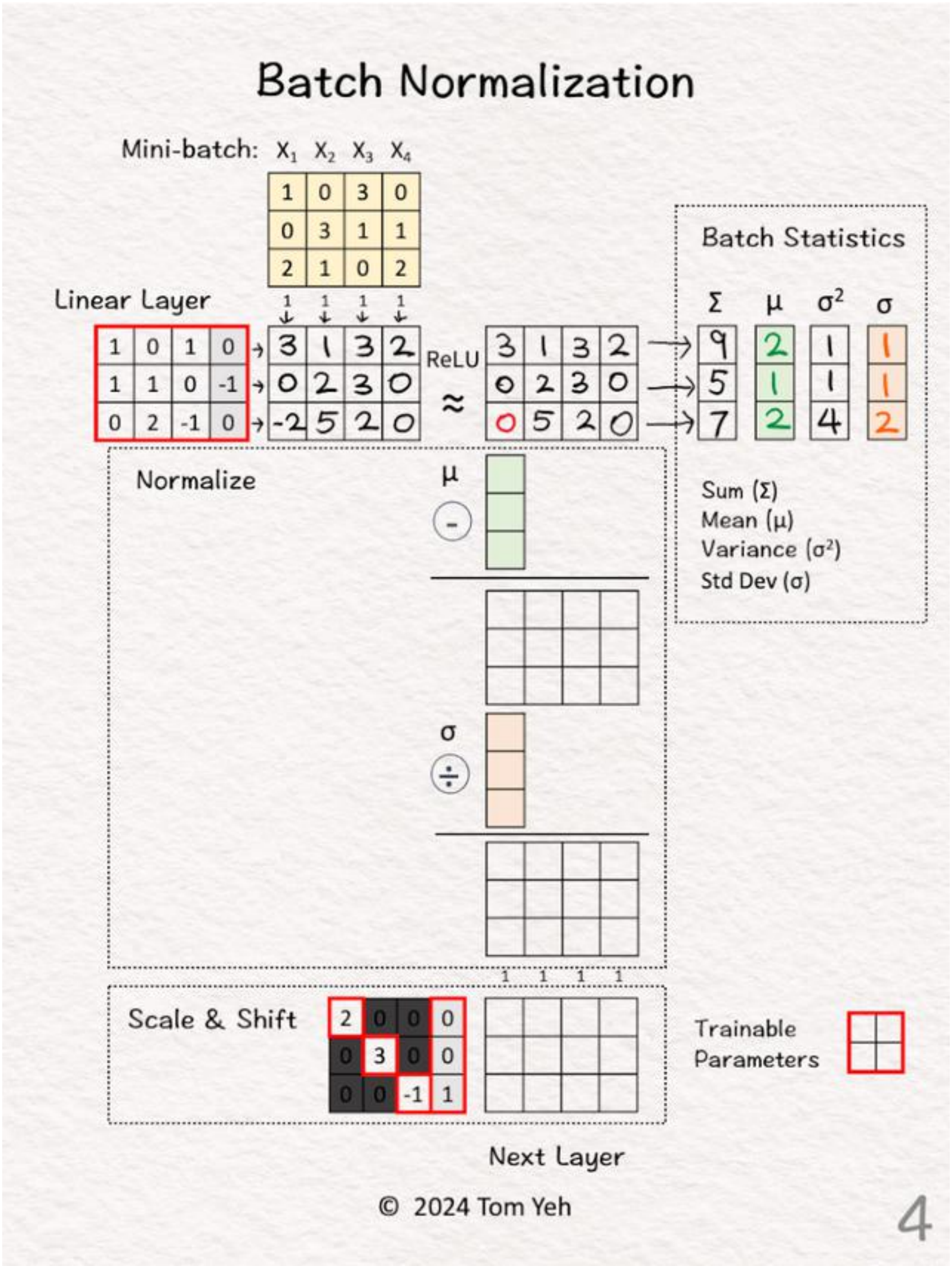

[2] Linear Layer

Multiply with the weights and biases to obtain new features



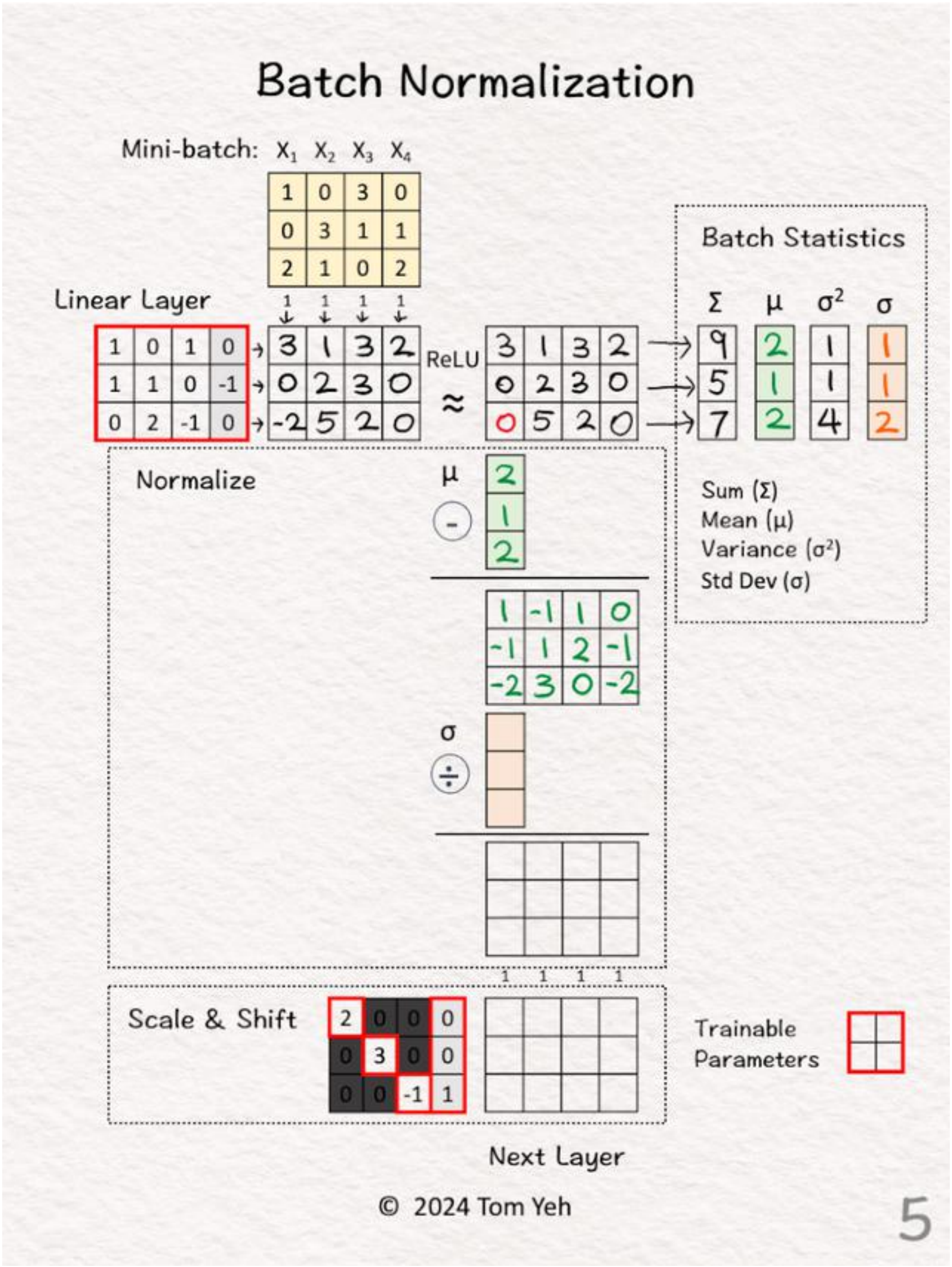

[3] ReLU

Apply the ReLU activation function, which has the effect of suppressing negative values. In this exercise, -2 is set to 0.



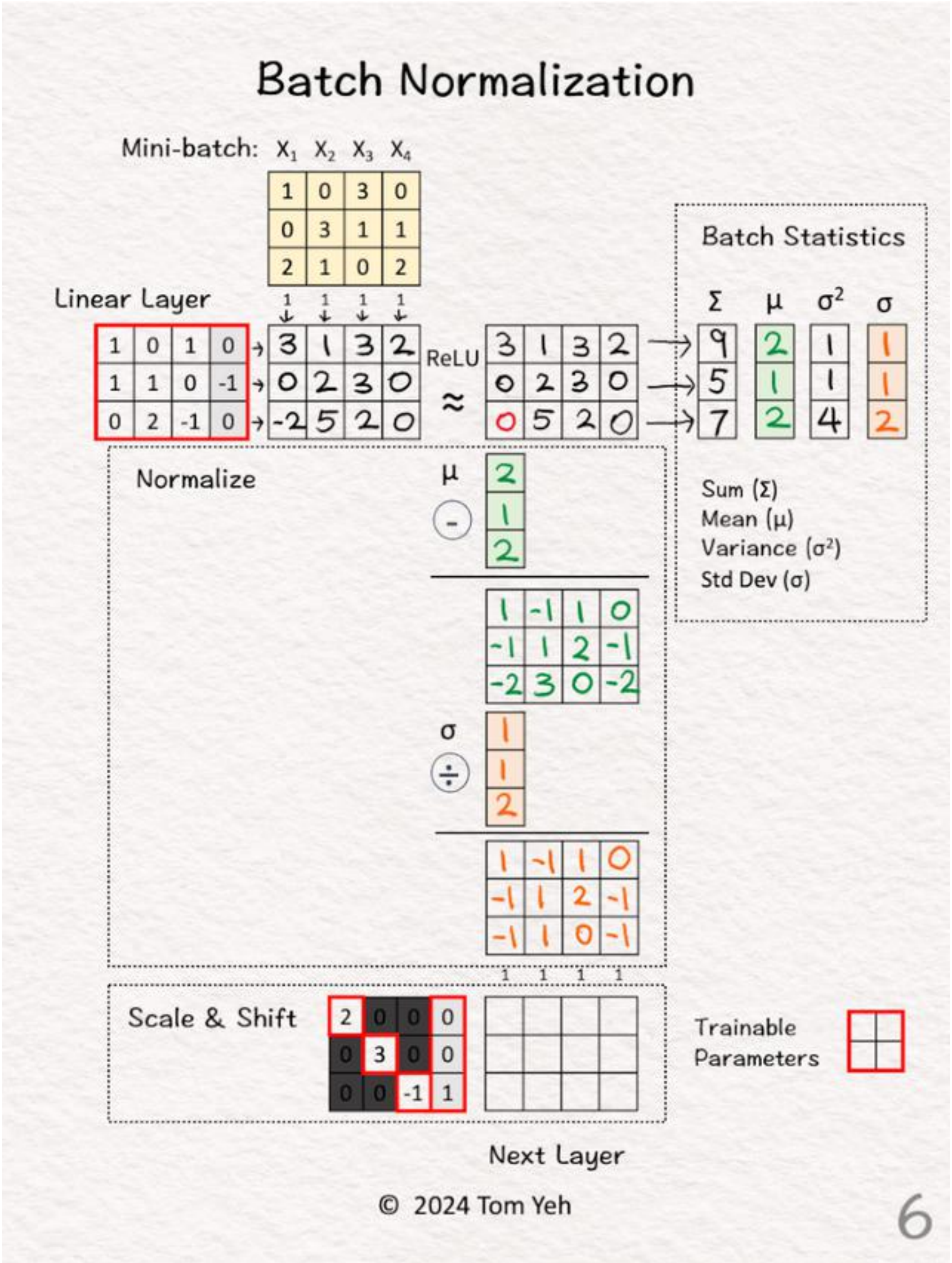

[4] Batch Statistics

Compute the sum, mean, variance, and standard deviation across the four examples in this min-batch.



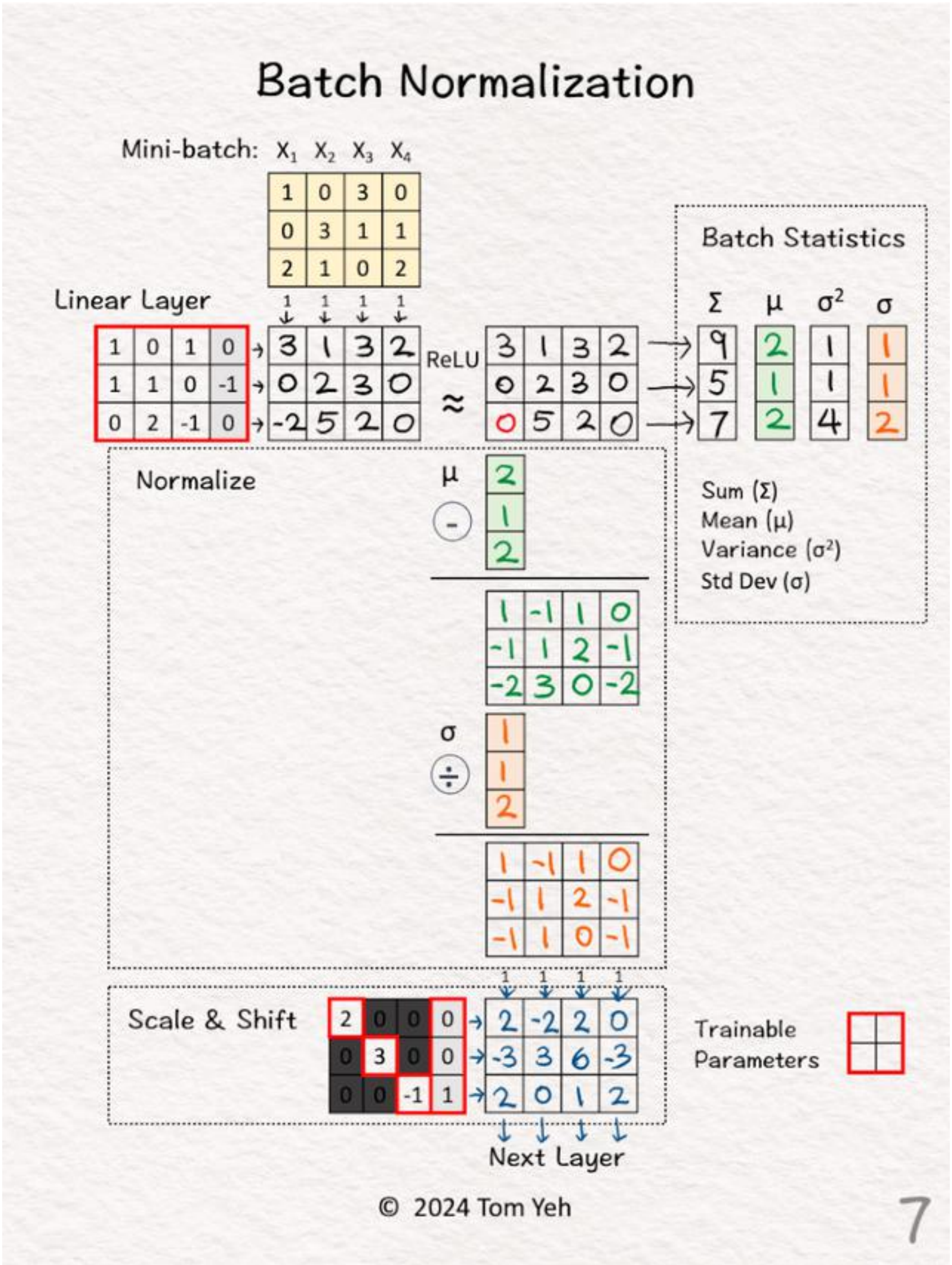

[5] Shift to Mean = 0

Subtract the mean (green) from the activation values for each training example




[6] Scale to Variance = 1

Divide by the standard deviation (orange)




[7] Scale & Shift

Multiply the normalized features from [6] by a linear transformation matrix, and pass the results to the next layer