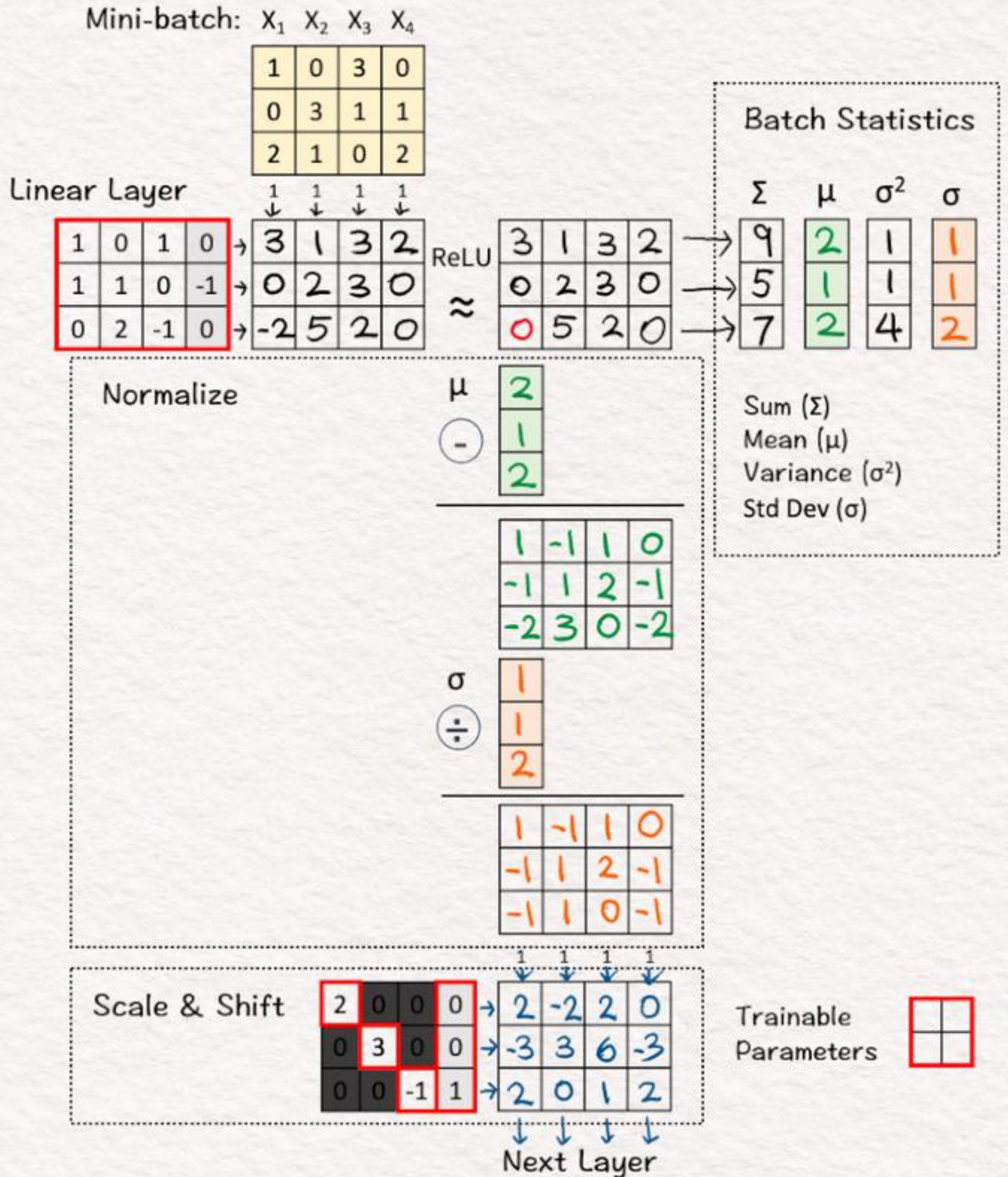


Batch Normalization



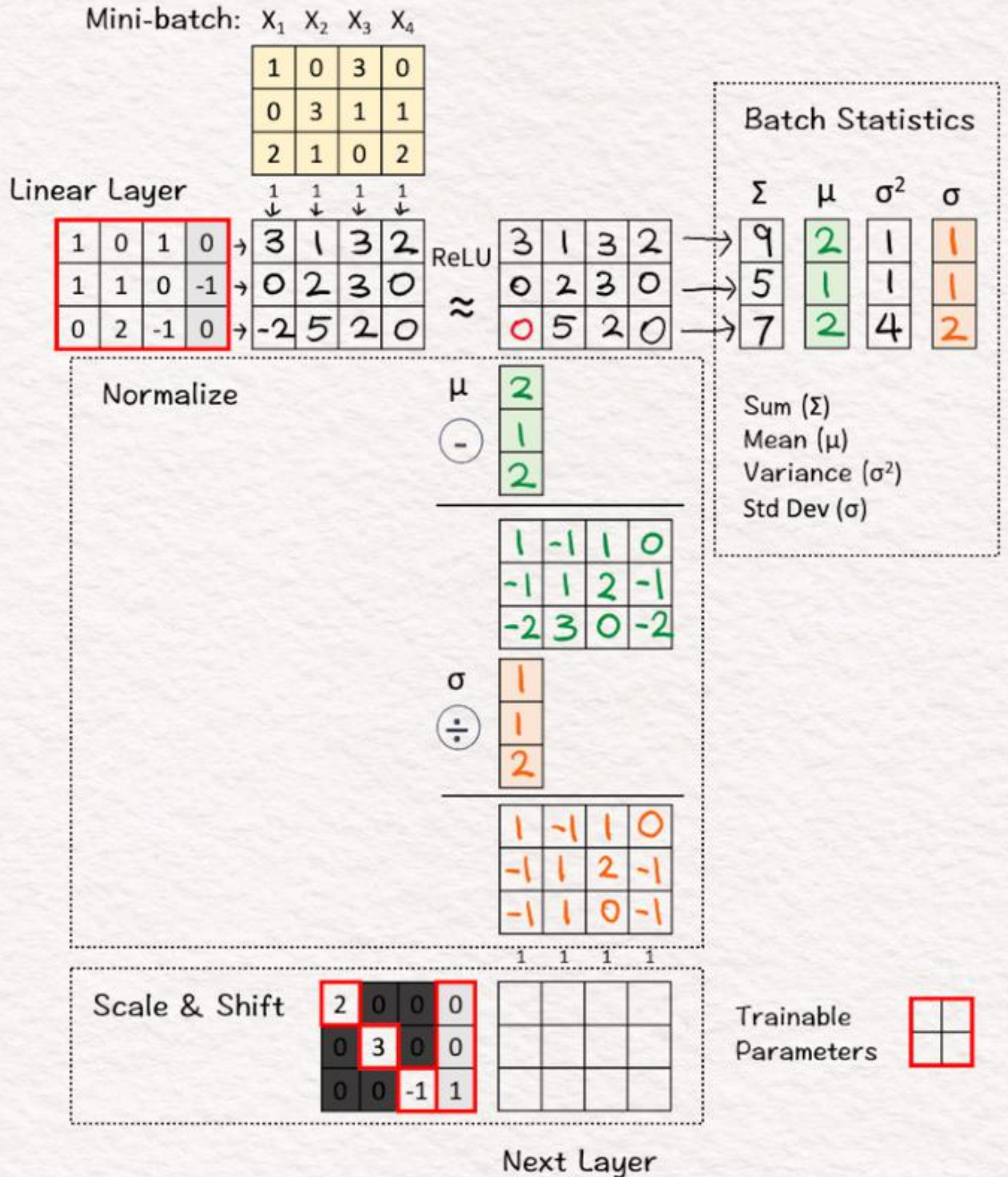
© 2024 Tom Yeh

7

[1] 给定 一个包含4个训练样本的小批量，每个样本有3个特征。

[1] Given A mini-batch of 4 training examples, each has 3 features.

Batch Normalization



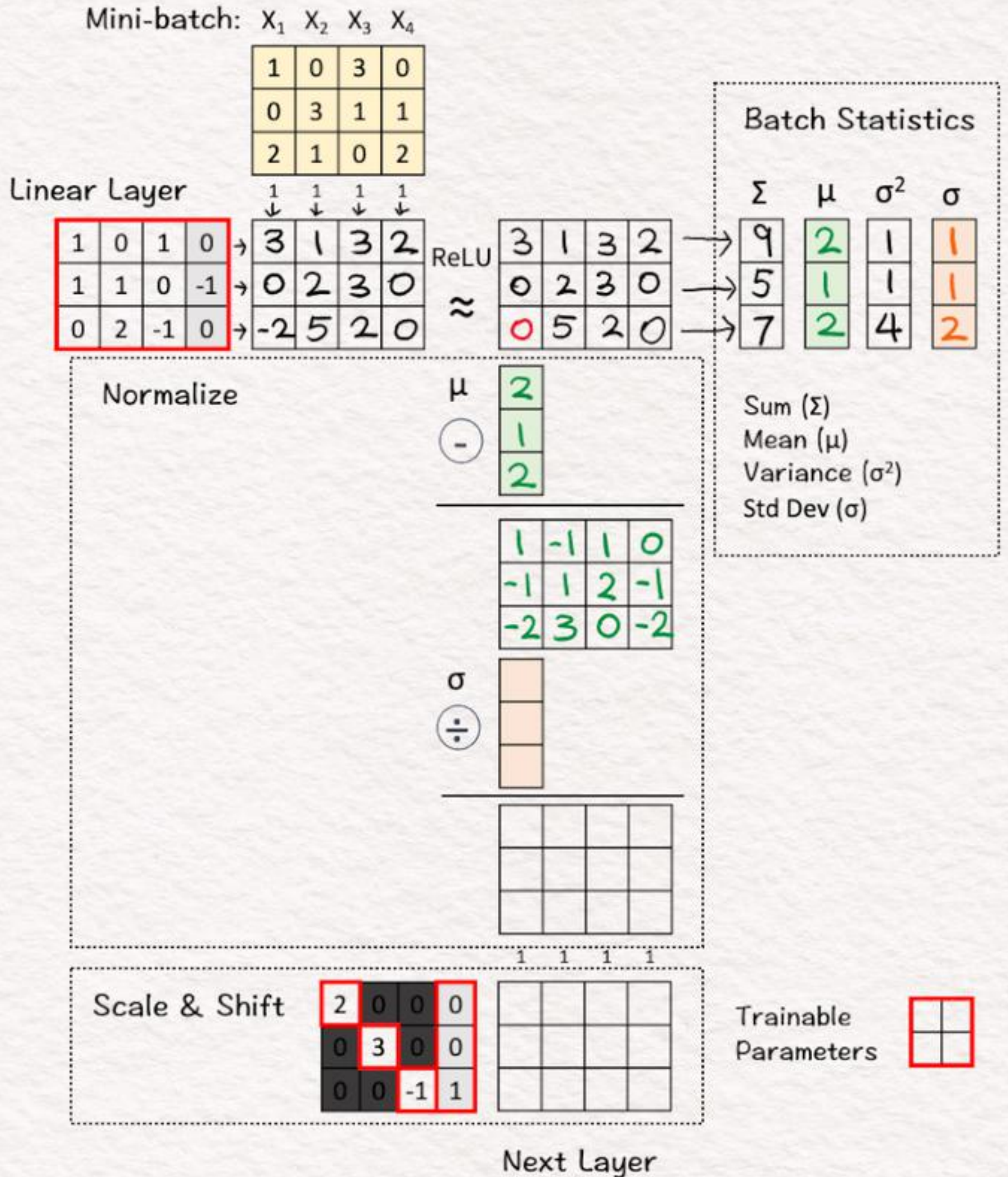
© 2024 Tom Yeh

6

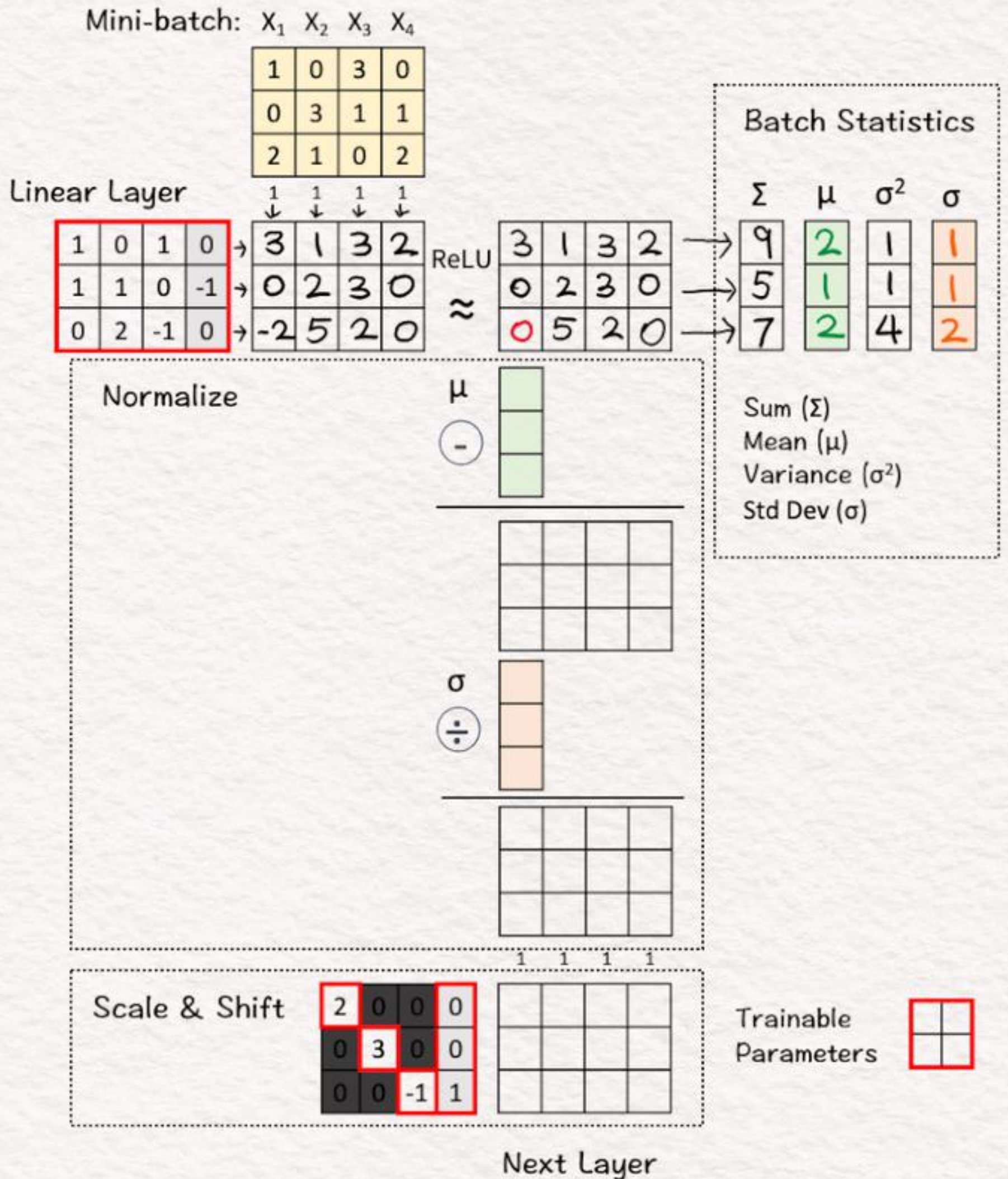
[2] 线性层 通过权重和偏置进行乘法操作以获得新的特征。

[2] Linear Layer ■ Multiply with the weights and biases to obtain new features.

Batch Normalization



Batch Normalization

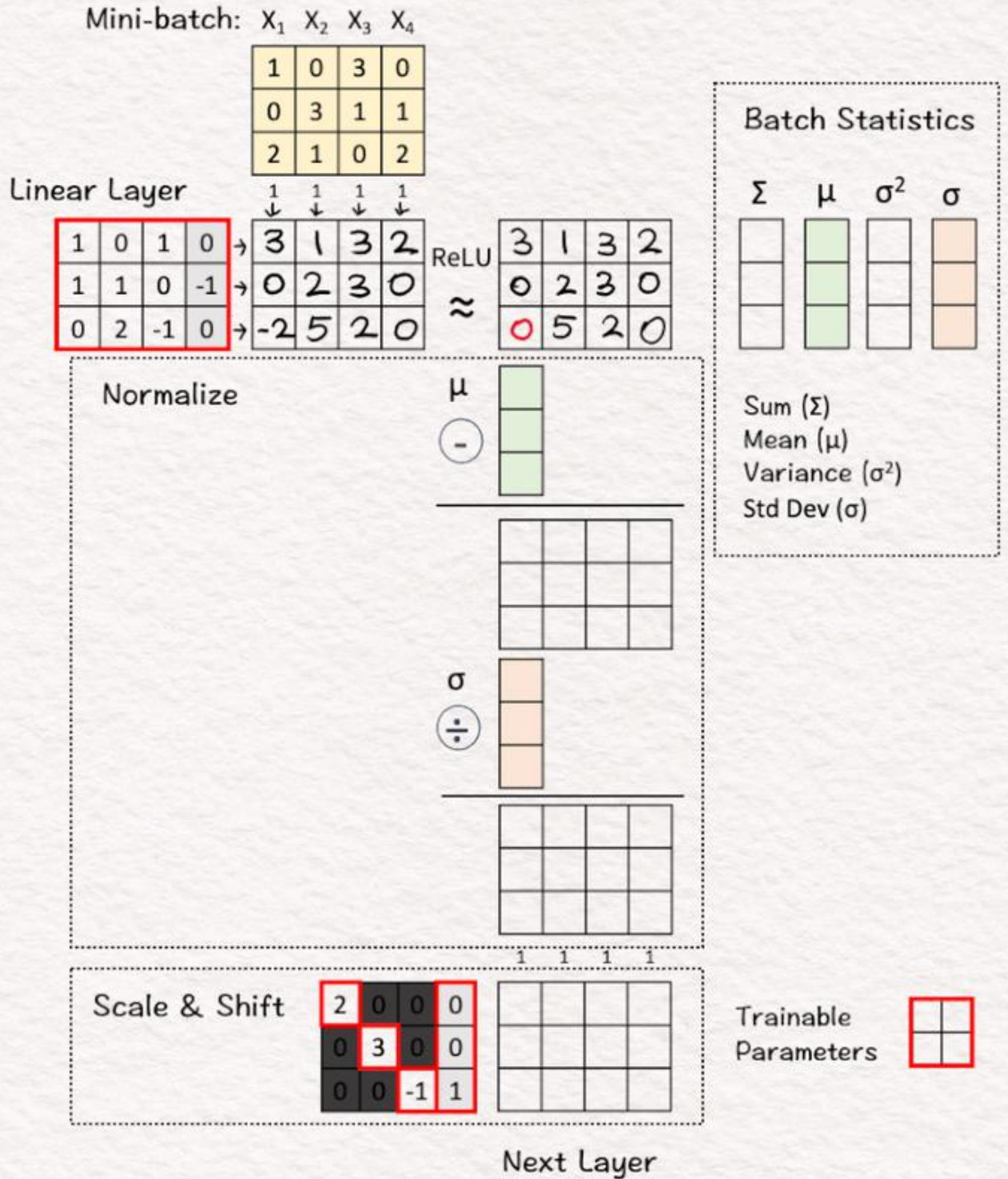


© 2024 Tom Yeh

4

[4] 批量统计 计算这个小批量中四个样本的总和、均值、方差和标准差。注意，这些统计量是针对每一行（即每个特征维度）计算的，and standard dev

Batch Normalization

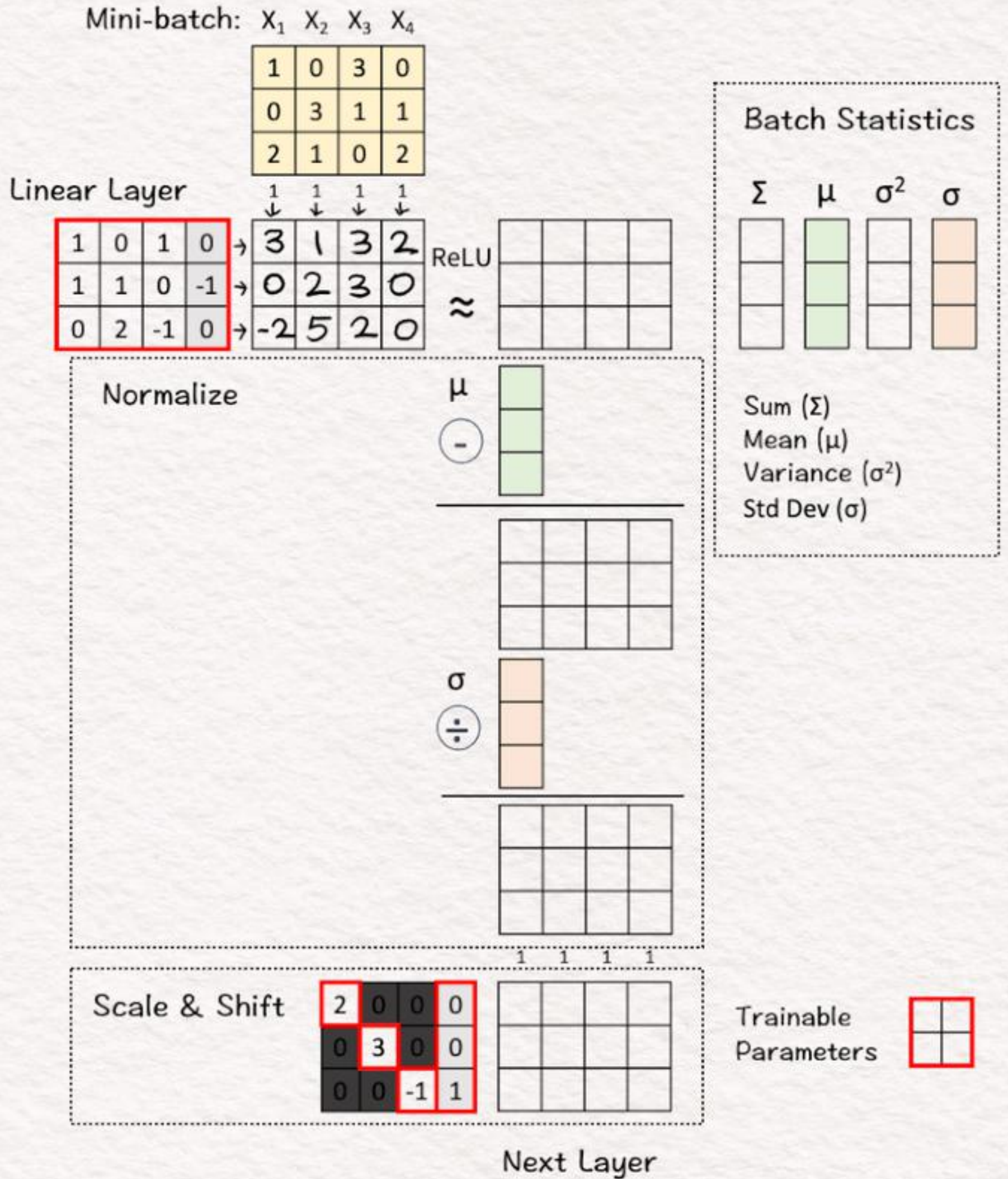


© 2024 Tom Yeh

3

[5] 均值归零 从每个训练样本的激活值中减去均值（绿色）目的是使每个维度中的4个激活值的平均值为零 Subtract the mean (green) from the activation values

Batch Normalization



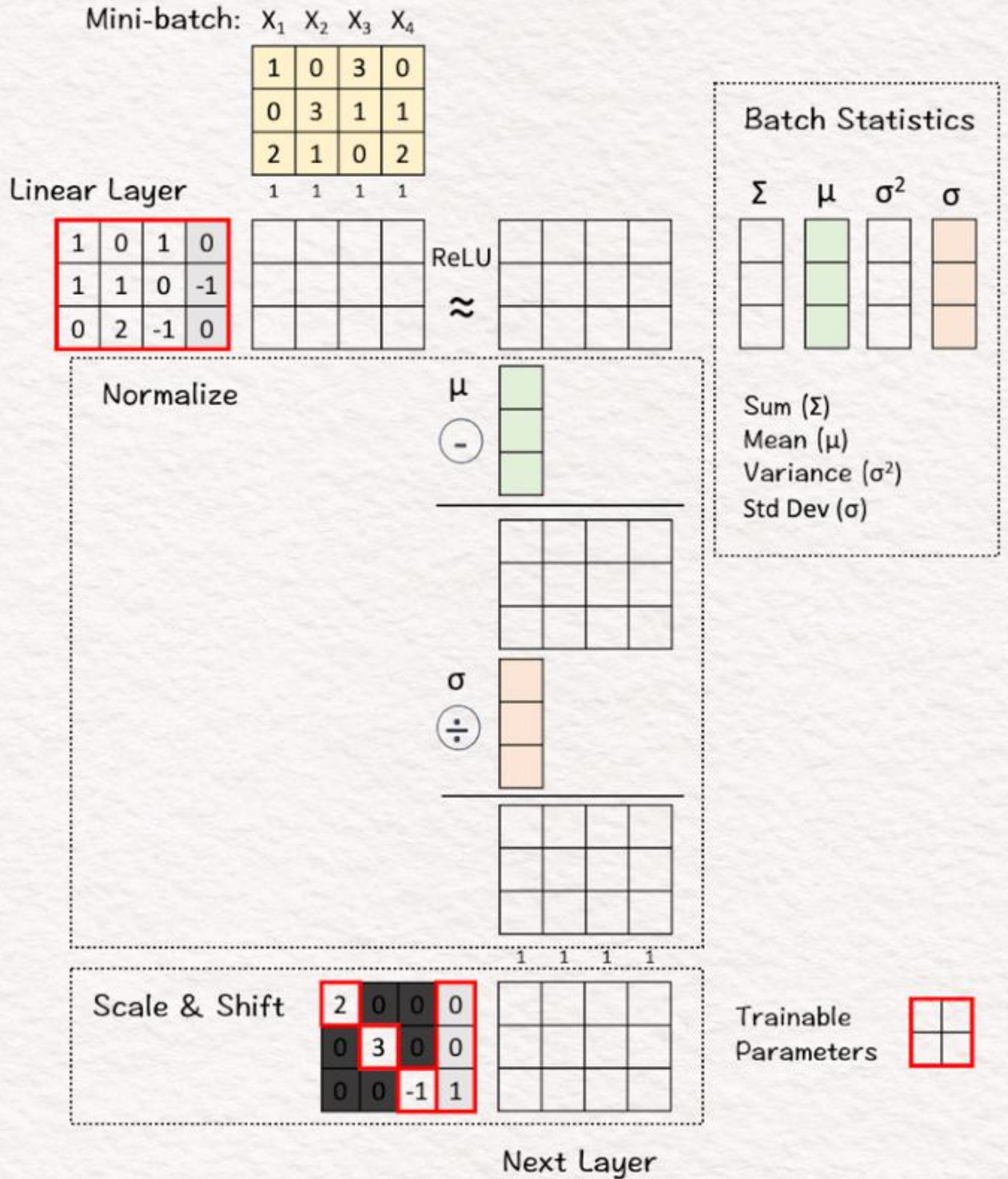
© 2024 Tom Yeh

2

[6] 方差归一 除以标准差（橙色） 目的是使4个激活值的方差等于一。

[6] Scale to Variance = 1 ■ Divide by the standard deviation (orange) ■ Th

Batch Normalization



© 2024 Tom Yeh

1

[7] 缩放与偏移 将来自[6]的标准化特征乘以线性变换矩阵，并将结果传递给下一层。目的是将归一化的特征值缩放和偏移至新的均值和方差。这些新的均值和方差是模型在训练过程中学习到的参数。