Batch Normalization Mini-batch: X₁ X₂ X₃ X₄ 3 3 1 1 **Batch Statistics** 2 0 2 Linear Layer σ^2 σ ReLU 0 Normalize Sum (S) Mean (µ) Variance (o2) Std Dev (o) \odot Scale & Shift Trainable Parameters Next Layer © 2024 Tom Yeh

[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。

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

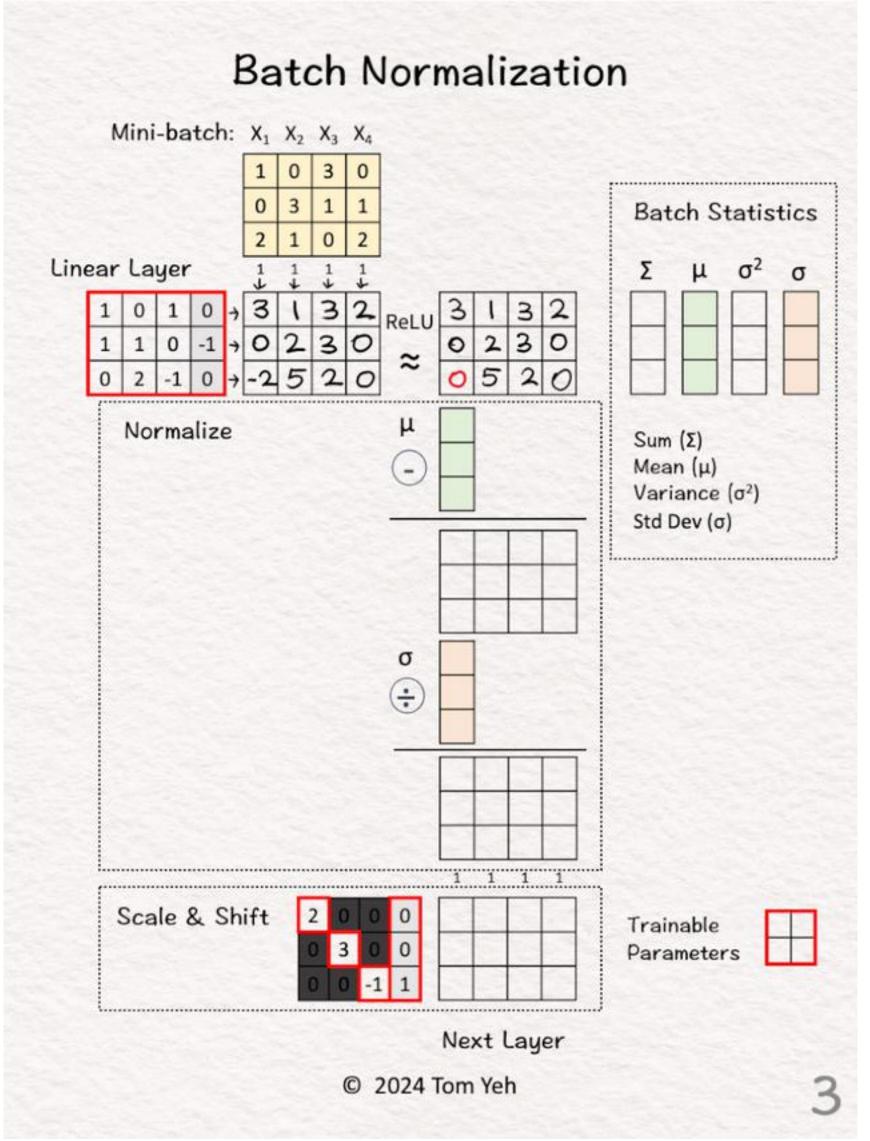
Batch Normalization Mini-batch: X₁ X₂ X₃ X₄ 3 1 **Batch Statistics** 1 2 ReLU σ^2 Linear Layer 3 0 Normalize Sum (S) Mean (µ) Variance (o2) Std Dev (o) σ \odot Scale & Shift Trainable Parameters Next Layer © 2024 Tom Yeh

[2] 线性层

[2] Linear Layer

通过权重和偏置进行乘法操作以获得新的特征。

Multiply with the weights and biases to obtain new features



[3] ReLU

Apply the ReLU activation function, which has the 应用ReLU激活函数,该函数的效果是抑制负值。在此练习中 effe被设为的pressing negative values. In this exercise, -2 is set to 0.

Batch Normalization Mini-batch: X₁ X₂ X₃ X₄ 3 1 **Batch Statistics** σ^2 Linear Layer 2 ReLU 3 Normalize Sum (S) Mean (µ) Variance (σ²) Std Dev (o) σ \odot Scale & Shift Trainable Parameters Next Layer © 2024 Tom Yeh

[4] 批量统计

[4] Batch Statistics

计算这个小批量中四个样本的总和、均值、方差和标准差。

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

注意[,]这些统计量是针对每一行(即每个特征维度)计算的 •Note that these statistics are computed for each row (i.e., each feature dimension).

Batch Normalization Mini-batch: X₁ X₂ X₃ X₄ 3 1 **Batch Statistics** Linear Layer σ^2 2 ReLU 3 Normalize Sum (S) Mean (µ) Variance (σ²) Std Dev (o) σ \odot Scale & Shift Trainable Parameters Next Layer © 2024 Tom Yeh

[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

[5] Shift to Mean = 0

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

The intended effect is for the 4 activation values in each dimension to average to zero

Batch Normalization Mini-batch: X₁ X₂ X₃ X₄ 3 1 **Batch Statistics** Linear Layer σ^2 2 ReLU 3 Normalize Sum (S) Mean (µ) Variance (σ²) Std Dev (o) σ \odot Scale & Shift Trainable Parameters Next Layer © 2024 Tom Yeh

[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)

The intended effect is for the 4 activation values to have variance equal to one.

Batch Normalization Mini-batch: X₁ X₂ X₃ X₄ 3 3 1 1 **Batch Statistics** Linear Layer σ^2 2 ReLU 3 Normalize Sum (S) Mean (µ) Variance (σ2) Std Dev (o) σ \odot Scale & Shift Trainable Parameters Next Layer © 2024 Tom Yeh

[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

目的是将标准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网络将要学习的

对角线上的元素和最后一列是网络将要学习的可训练参数。

[7] Scale & Shift

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

The intended effect is to scale and shift the normalized feature values to a new mean and variance, which are to be learned by the network

The elements in the diagonal and the last column are trainable parameters the network will learn.