

Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift

Ravi Raju

12-28-2018

1 Motivation

The problem is that training deep networks is difficult because the distribution of each layer's input is changing due to previous layer parameters changing. This is making models with saturating non-linearities which requires lower learning rates and careful initializations tough to train. The author call this internal covariate shift.

2 Solution

They address this shift with normalizing layer inputs over all the mini batches. This allows use of higher learning rates and less scurtiny over initialization. Really, this is a problem with vanishing gradients since inputs, besides those with small absolute values, will have small/no changes.

3 Results

They apply BatchNorm to ImageNet network where they claim they can match performance with only 7 %, demonstrating how quick the convergence.

4 Discussion/Takeaway

Basically what needs to happen is that the input to the non-linearity wants to be in the linear region since this is the best for optimization.