

Batch Normalization

OpenAI

Ian Goodfellow

Deep Learning Study Group

San Francisco

September 12, 2016

Batch Normalization

$$\mathbf{Z} = \mathbf{XW}$$

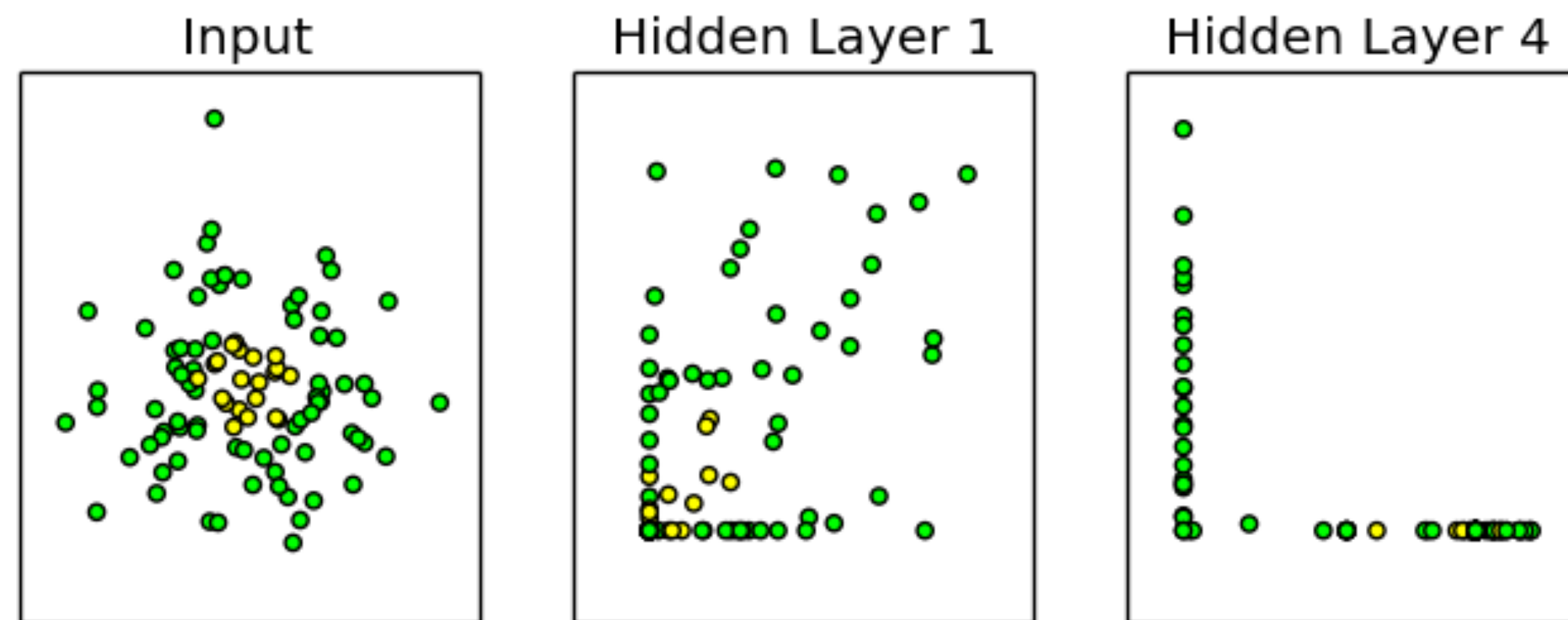
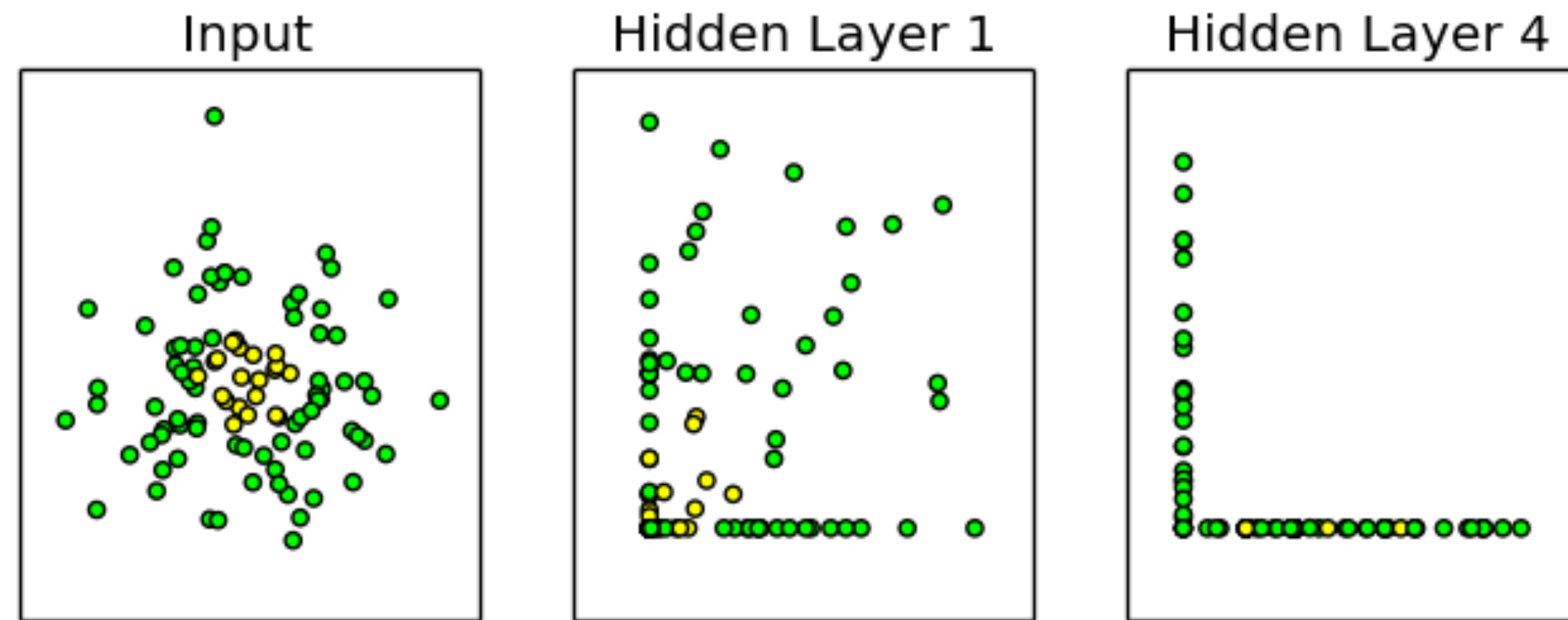
$$\tilde{\mathbf{Z}} = \mathbf{Z} - \frac{1}{m} \sum_{i=1}^m \mathbf{Z}_{i,:}$$

$$\hat{\mathbf{Z}} = \frac{\tilde{\mathbf{Z}}}{\sqrt{\epsilon + \frac{1}{m} \sum_{i=1}^m \tilde{\mathbf{Z}}_{i,:}^2}}$$

$$\mathbf{H} = \max\{0, \gamma \hat{\mathbf{Z}} + \beta\}$$

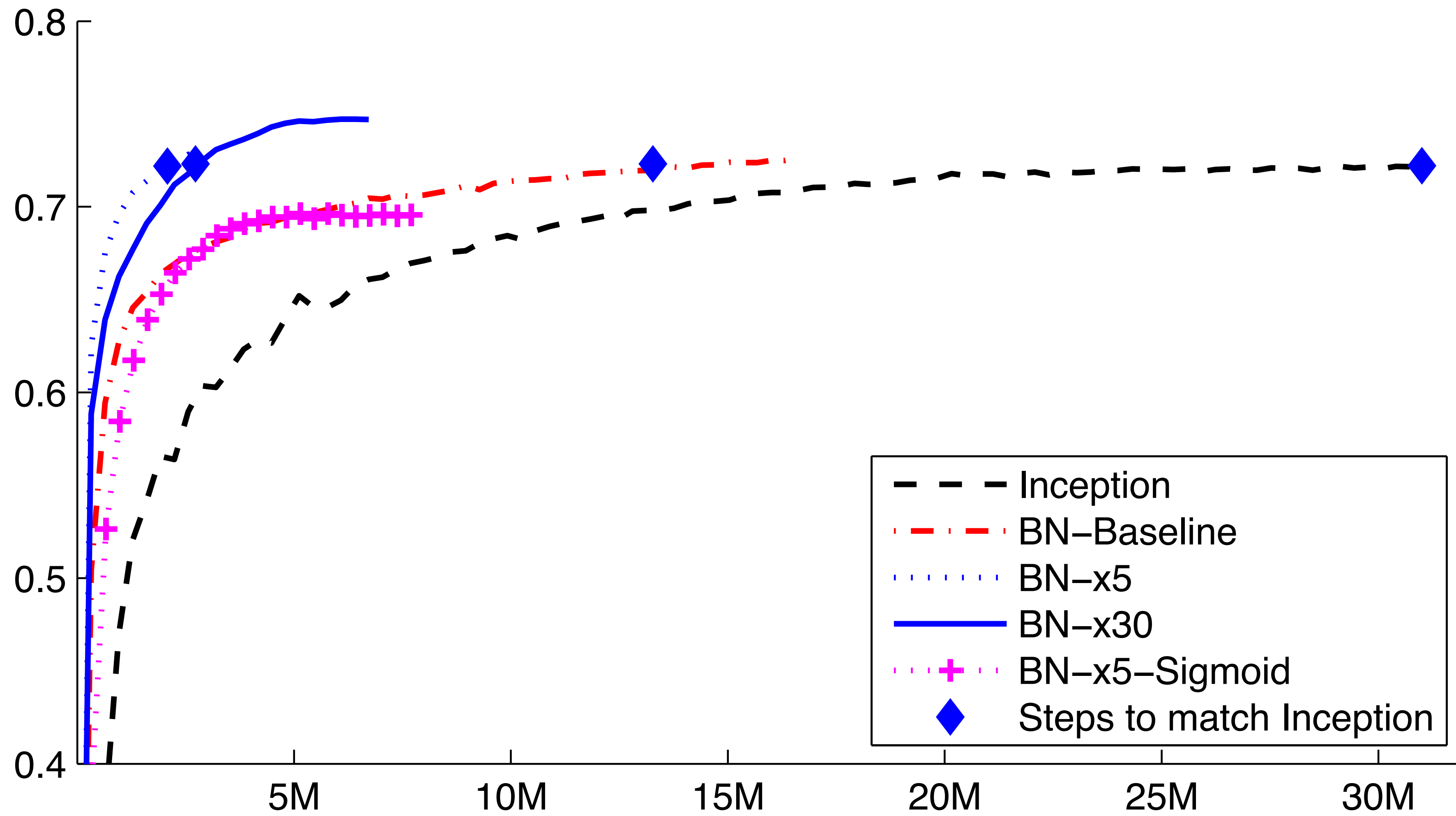
“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Ioffe and Szegedy 2015

Before SGD step



After SGD step

“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Ioffe and Szegedy 2015



“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Ioffe and Szegedy 2015