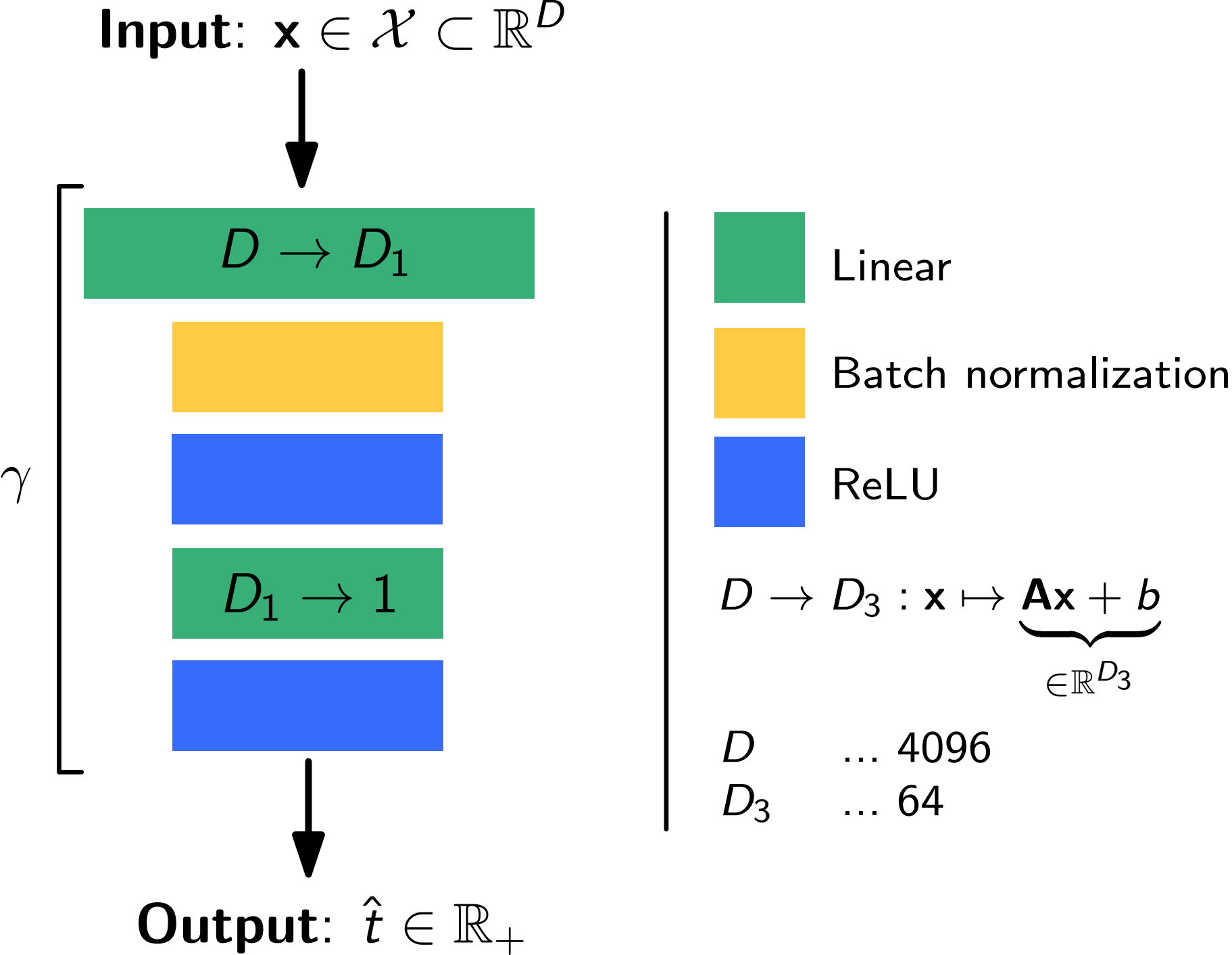
Input: $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D$
$D \to D_1$
γ
$D_1 \to 1$
Output: $\hat{t} \in \mathbb{R}_+$
Linear
Batch normalization
ReLU
$D \to D_3 : \mathbf{x} \mapsto \underbrace{\mathbf{A}\mathbf{x} + b}_{\in \mathbb{R}^{D_3}}$
D ... 4096
D_3 ... 64