

Input: $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D$



$\mathbb{R}^D \rightarrow \mathbb{R}^{D_1}$



$\mathbb{R}^{D_1} \rightarrow \mathbb{R}$



Output: $\hat{t} \in \mathbb{R}_+$

γ



Linear



Batch normalization



ReLU

$$\mathbb{R}^D \rightarrow \mathbb{R}^{D_1} : \mathbf{x} \mapsto \underbrace{\mathbf{A}\mathbf{x} + b}_{\in \mathbb{R}^{D_1}}$$

$D \quad \dots \quad 4096$

$D_1 \quad \dots \quad 64$