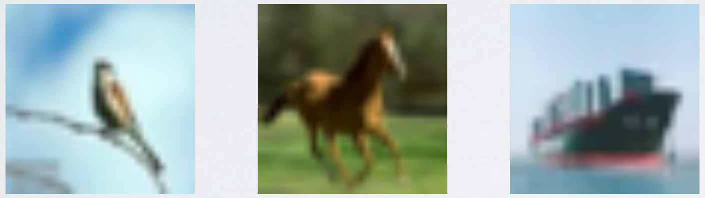
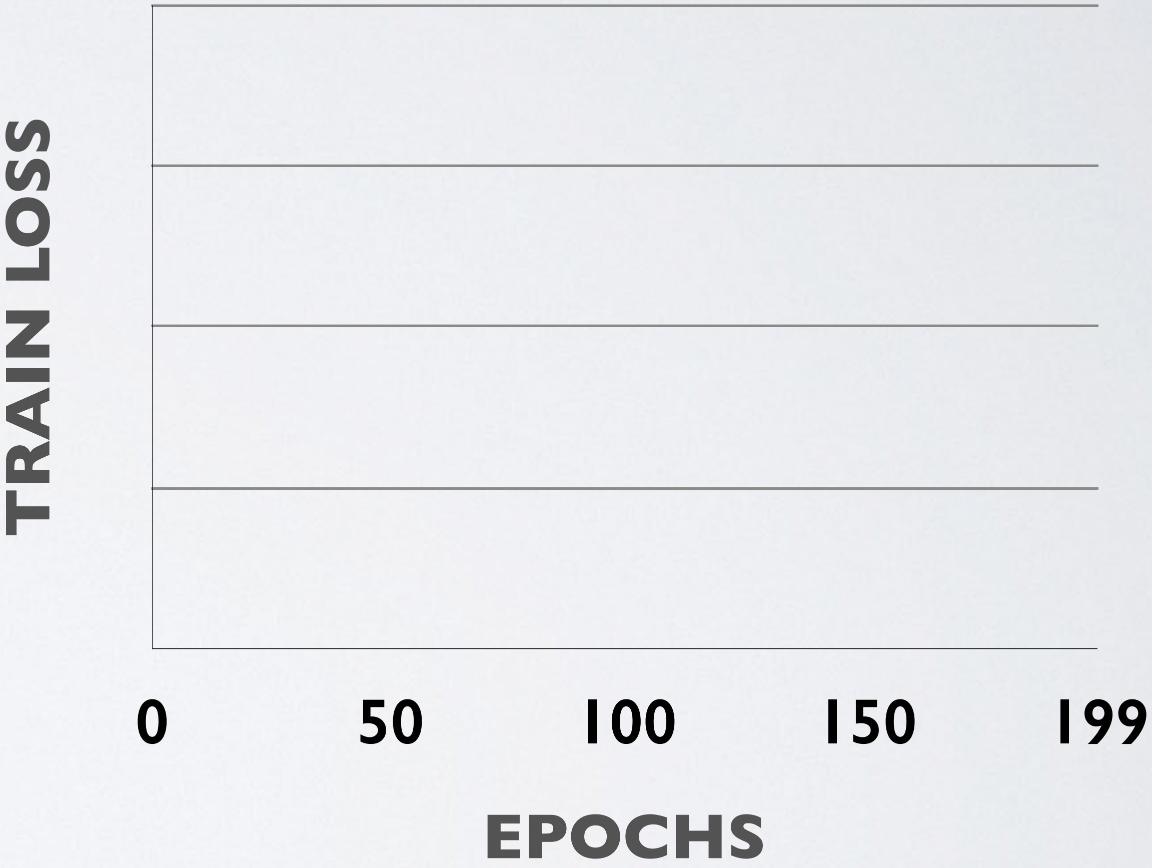
SUBSTITUTING RELUS WITH HERMITE POLYNOMIALS GIVES FASTER CONVERGENCE FOR SSL

PSEUDO-LABELS VIA THE BANDERSNATCH PHENOMENON

PSEUDO-LABELS VIA THE BANDERSNATCH PHENOMENON

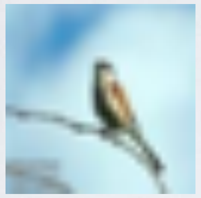
CIFAR10

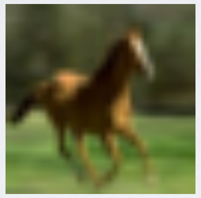


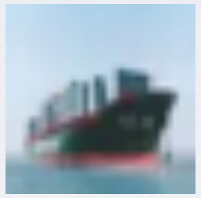


PSEUDO-LABELS VIA THE BANDERSNATCH PHENOMENON

CIFAR10

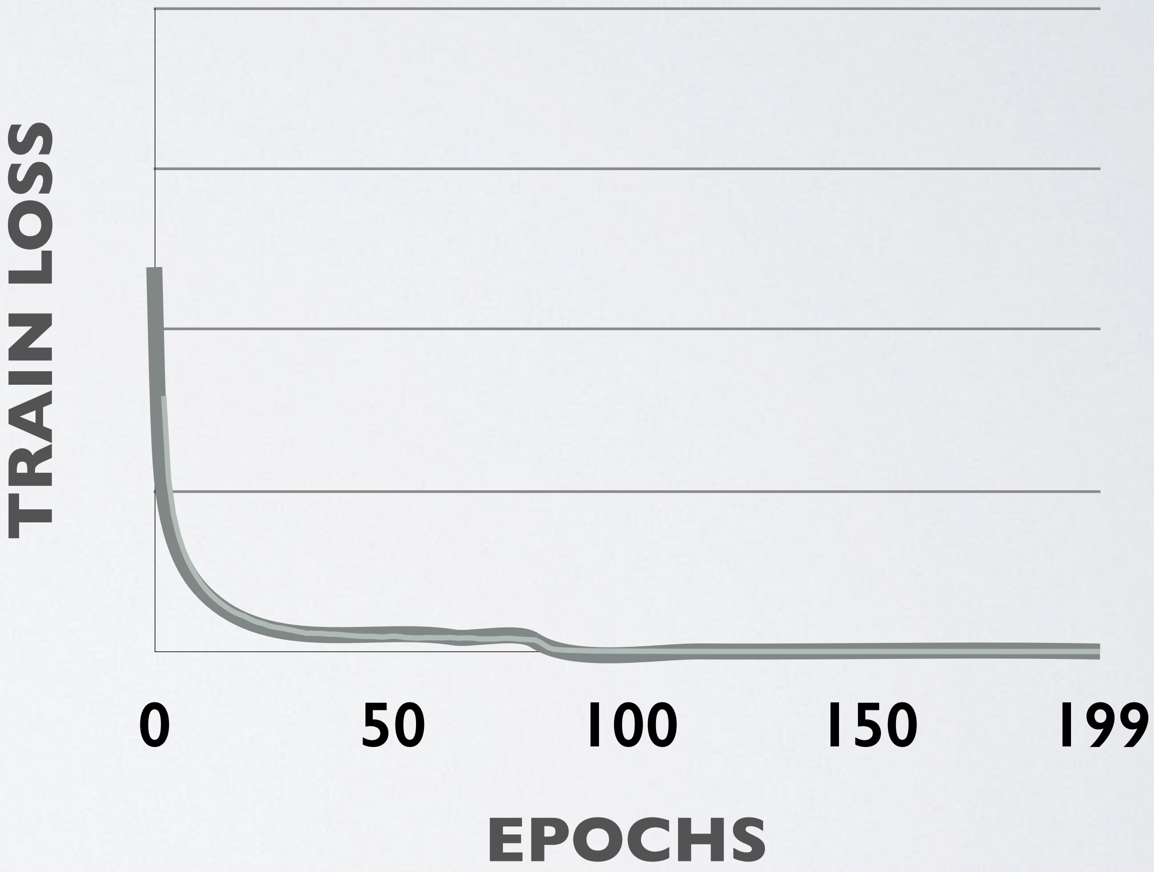




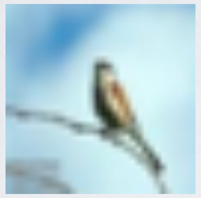
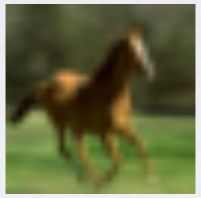
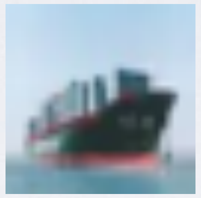


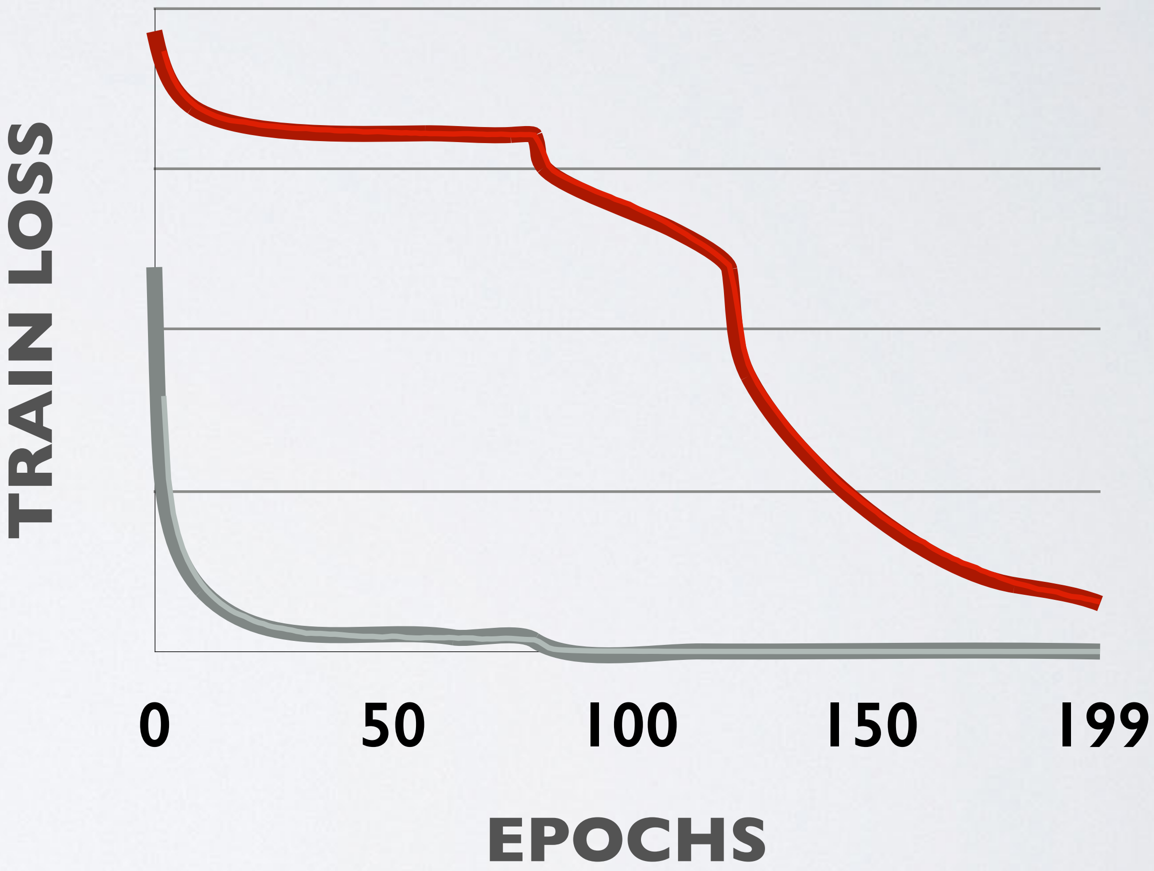
TRUE LABELS

Bird Horse Ship



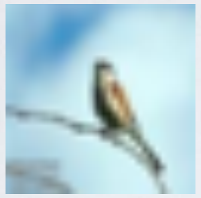
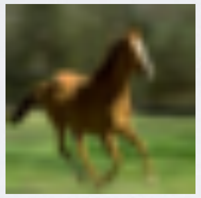
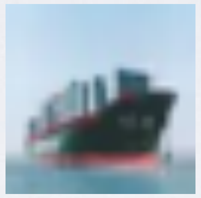
PSEUDO-LABELS VIA THE BANDERSNATCH PHENOMENON

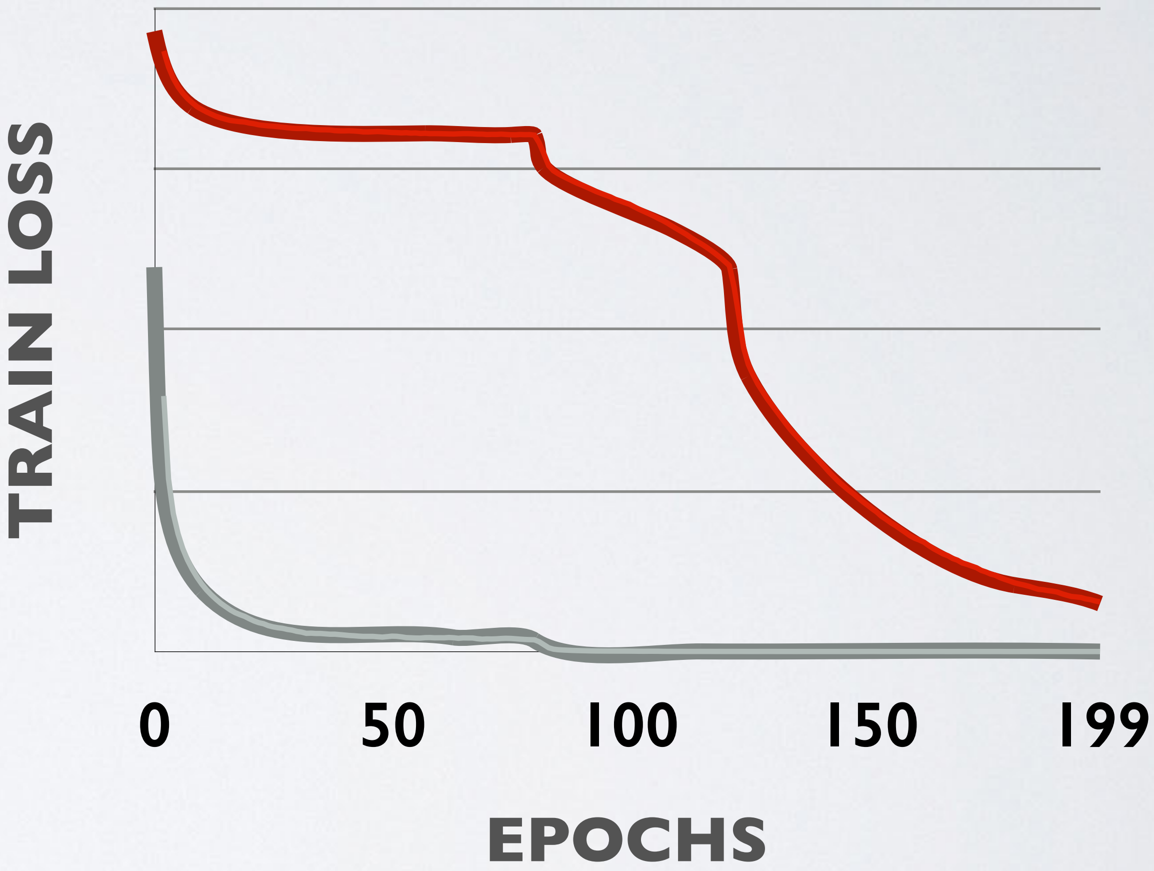
CIFAR10			
TRUE LABELS	Bird	Horse	Ship
RANDOM LABELS	Cat	Truck	Dog



PSEUDO-LABELS VIA THE BANDERSNATCH PHENOMENON

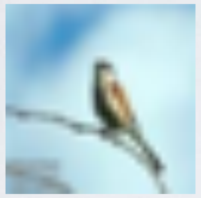
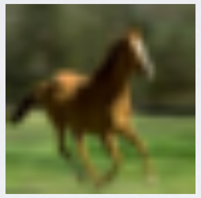
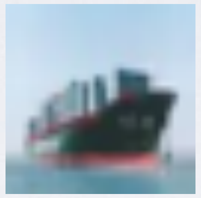
Good Labels → Least Training Time

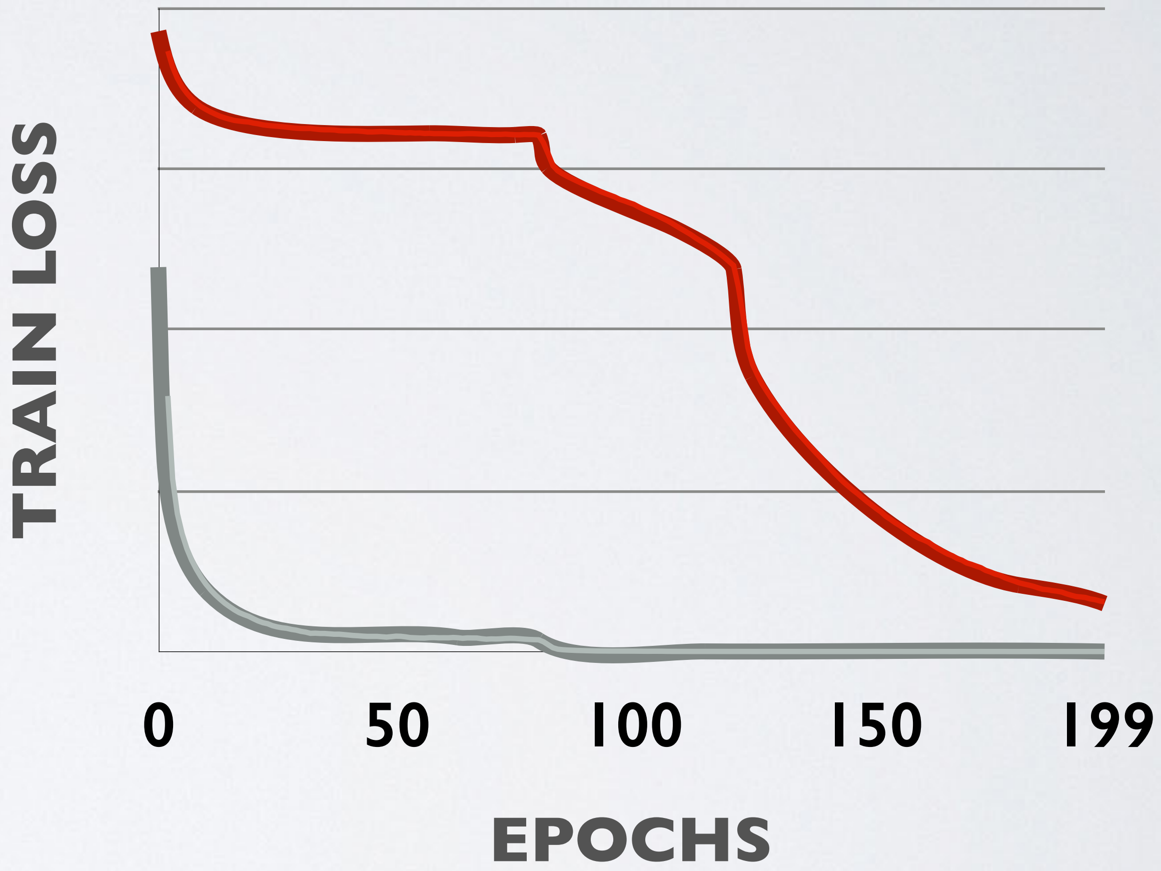
CIFAR10			
TRUE LABELS	Bird	Horse	Ship
RANDOM LABELS	Cat	Truck	Dog



PSEUDO-LABELS VIA THE BANDERSNATCH PHENOMENON

Good Labels → Least Training Time

CIFAR10			
TRUE LABELS	Bird	Horse	Ship
RANDOM LABELS	Cat	Truck	Dog



SaaS: Find labels with least training time

NON-SMOOTH RELUS TO POLYNOMIAL ACTIVATIONS

NON-SMOOTH RELUS TO POLYNOMIAL ACTIVATIONS

- Ge et al showed that we can avoid spurious local minima using Hermites

NON-SMOOTH RELUS TO POLYNOMIAL ACTIVATIONS

- Ge et al showed that we can avoid spurious local minima using Hermites
- Other orthogonal polynomial bases are possible

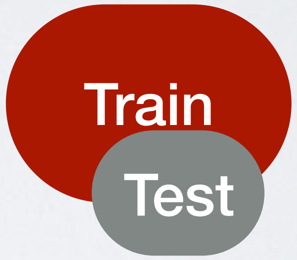
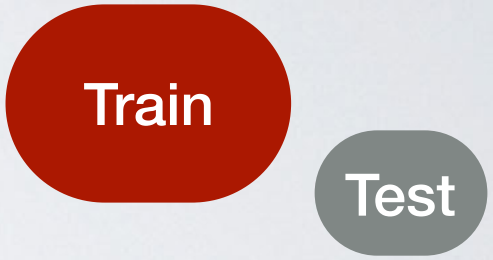
NON-SMOOTH RELUS TO POLYNOMIAL ACTIVATIONS

- Ge et al showed that we can avoid spurious local minima using Hermites

- Other orthogonal polynomial bases are possible

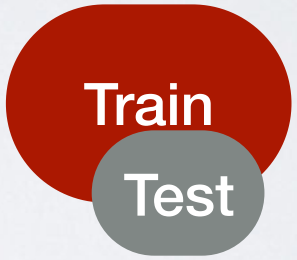
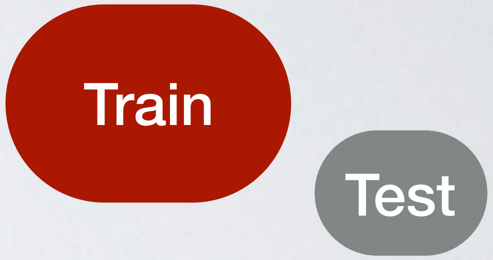
- Idea: use the lower order terms in the Hermite polynomial series **Expansion of ReLU** as **Activation Functions**

HERMITES CAN DO WHAT? THEORETICAL RESULT

	 Train Test	 Train Test
ReLU	High Confidence	High Confidence 🙁
Hermite	High Confidence	Low Confidence 😀

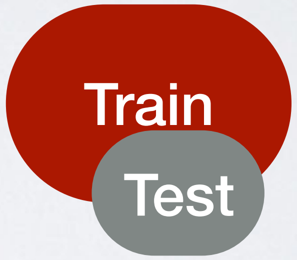
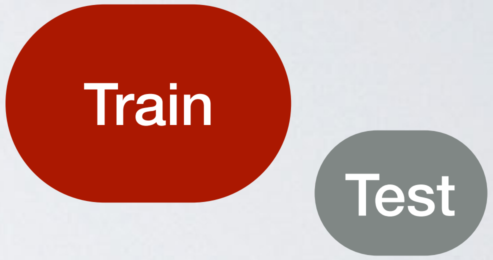
HERMITES CAN DO WHAT? THEORETICAL RESULT

- When the test data is somewhat different from the training data, the predictions made by Hermite networks are **approximately (uniformly) random**.

	 Train Test	 Train Test
ReLU	High Confidence	High Confidence 🙁
Hermite	High Confidence	Low Confidence 😀

HERMITES CAN DO WHAT? THEORETICAL RESULT

- When the test data is somewhat different from the training data, the predictions made by Hermite networks are **approximately (uniformly) random**.
- This enables the network to make **lower confidence prediction** on the testing data dissimilar to the training data which can later be **labelled manually.**

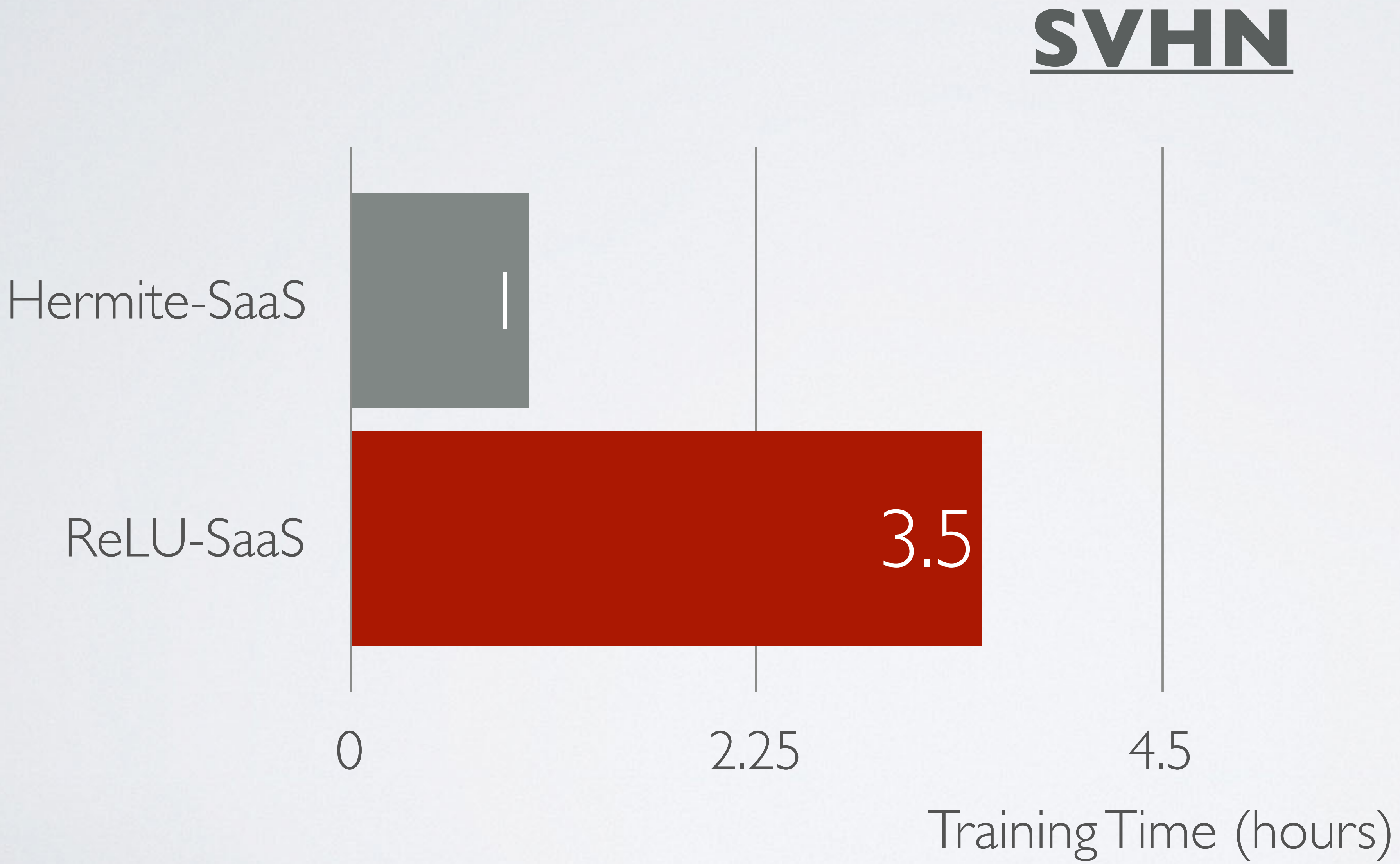
	Train Test	Train Test
ReLU	High Confidence	High Confidence 🙁
Hermite	High Confidence	Low Confidence 😀

COMPUTATIONAL BENEFITS

Hermite-SaaS **TRAINS FASTER** than ReLU-SaaS

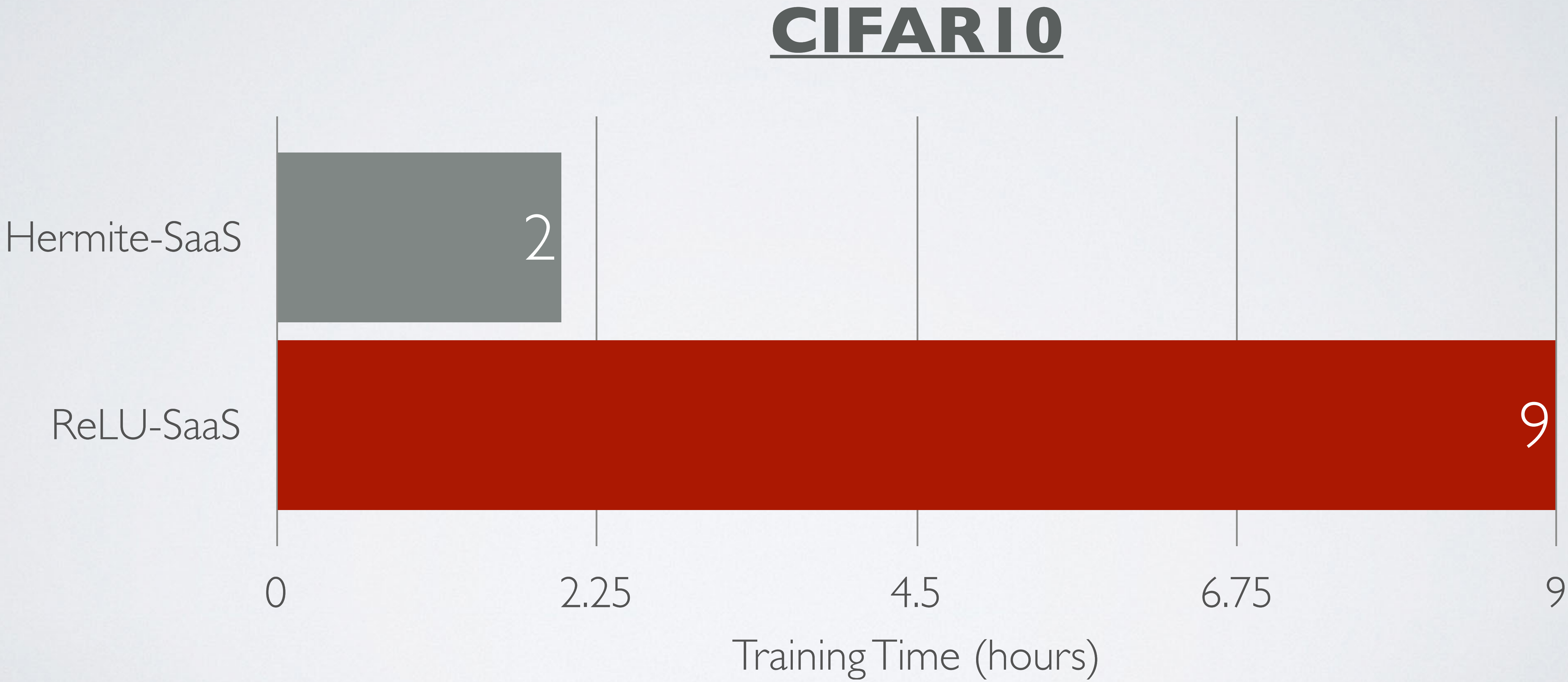
On AWS
p3.2xlarge

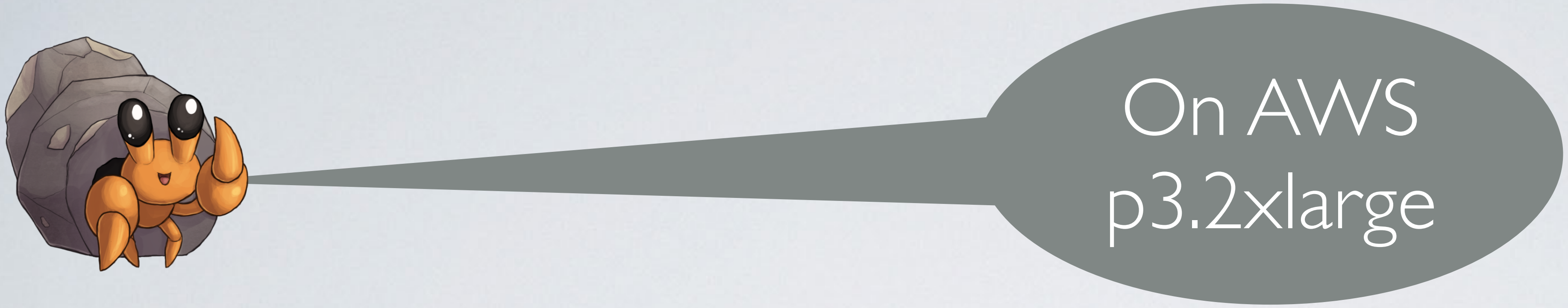
Hermite-SaaS **TRAINS FASTER** than ReLU-SaaS



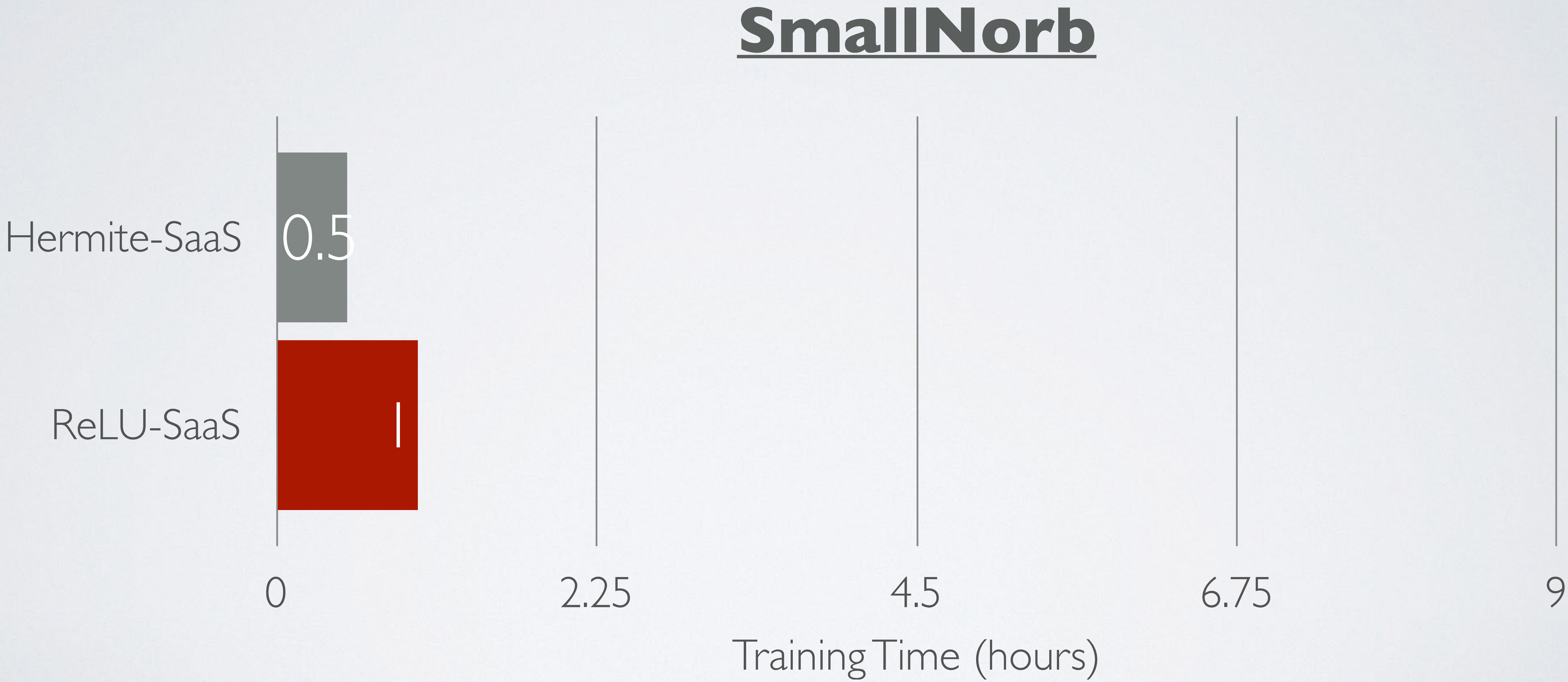
On AWS p3.2xlarge

Hermite-SaaS TRAINS FASTER than ReLU-SaaS





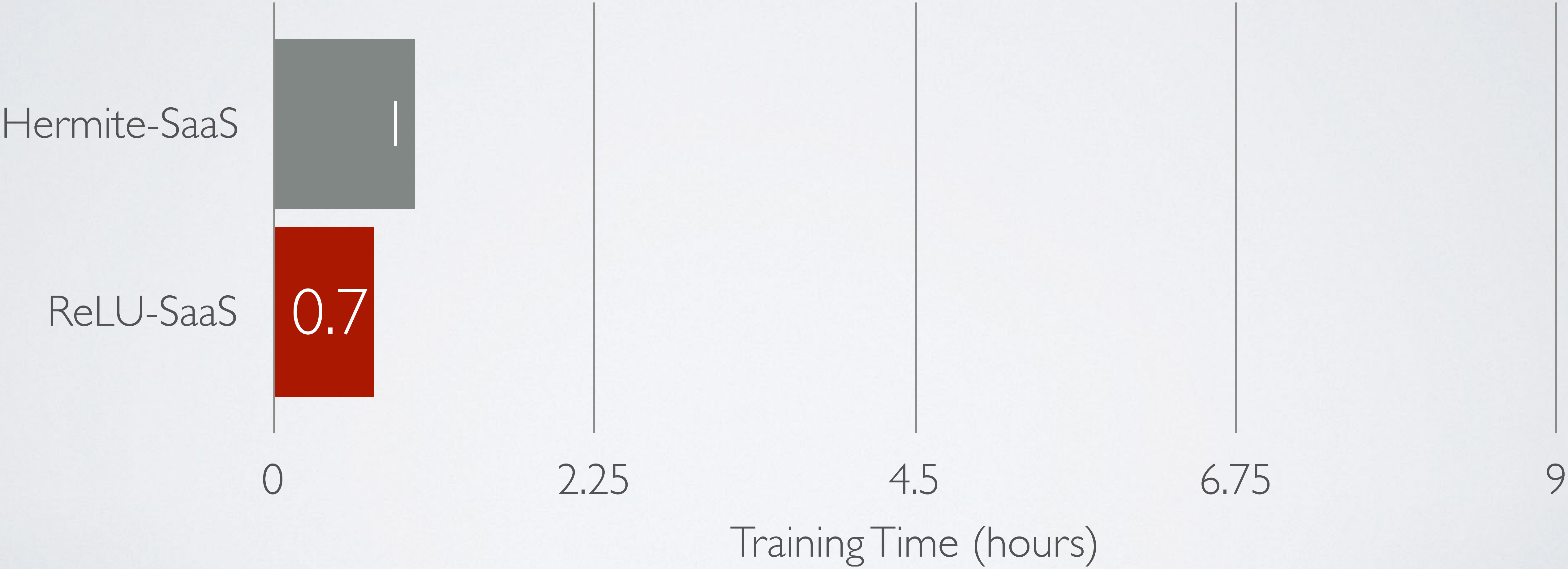
Hermite-SaaS TRAINS FASTER than ReLU-SaaS



On AWS
p3.2xlarge

Hermite-SaaS TRAINS FASTER than ReLU-SaaS

MNIST

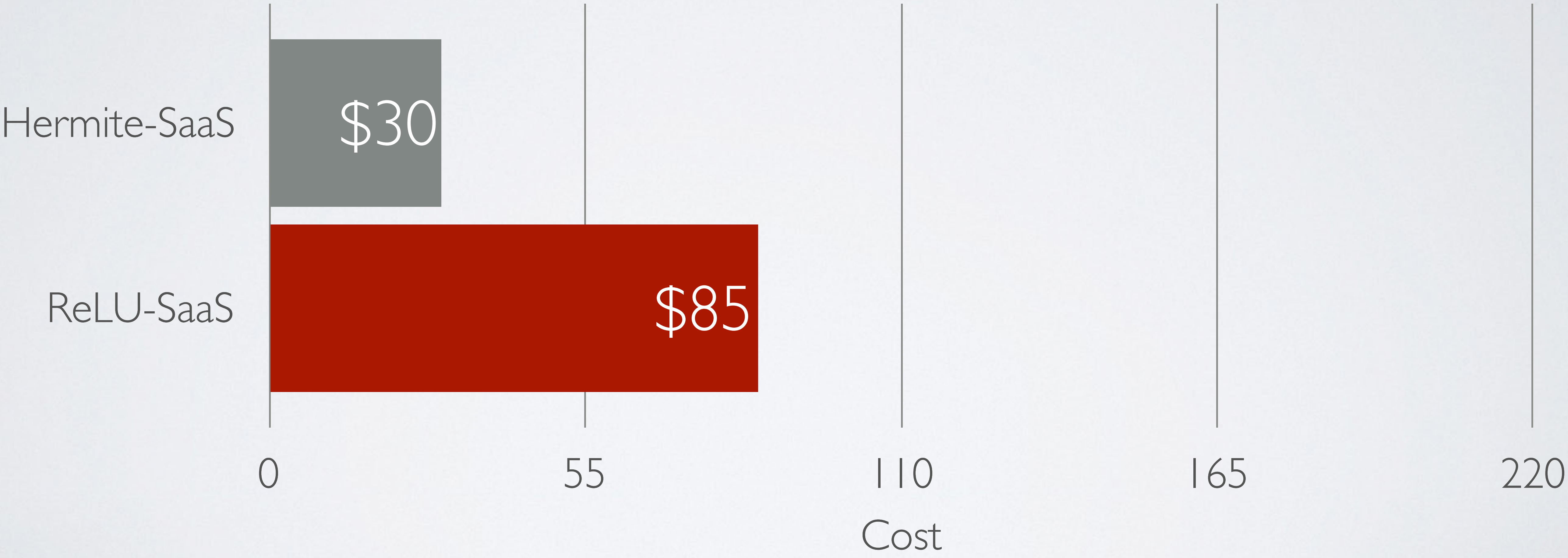


Hermite-SaaS **COSTS LESS** than ReLU-SaaS

On AWS
p3.2xlarge

Hermite-SaaS **COSTS LESS** than ReLU-SaaS

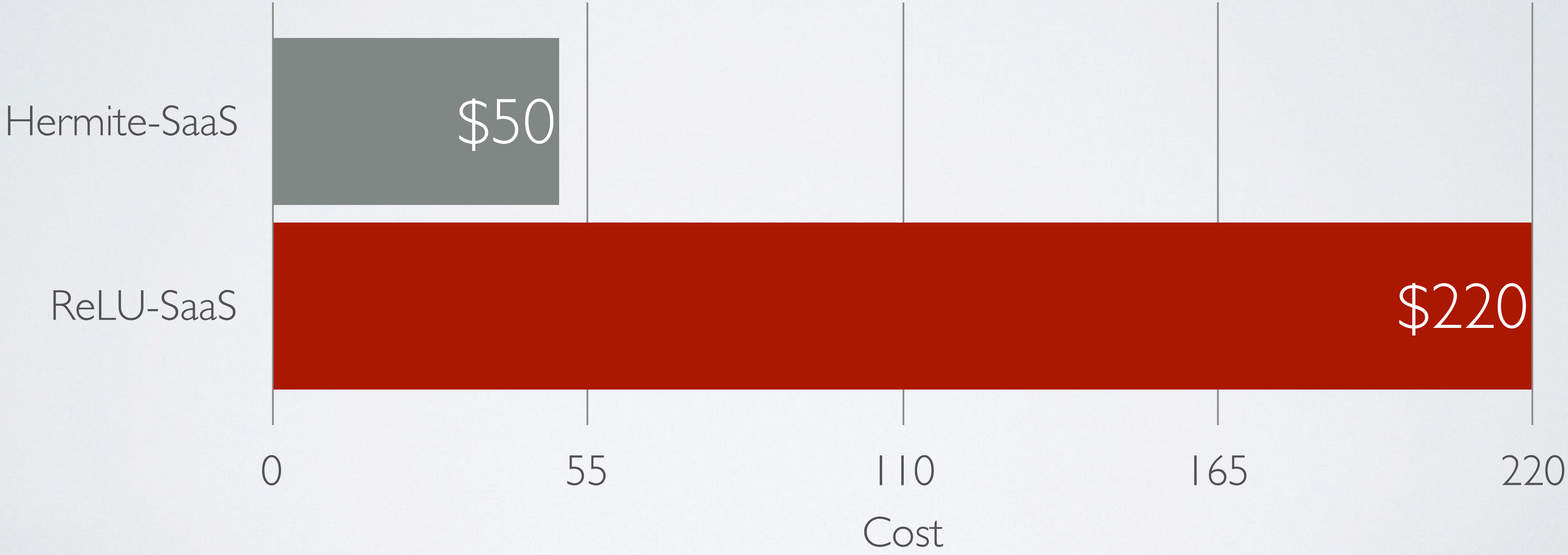
SVHN



On AWS p3.2xlarge

Hermite-SaaS COSTS LESS than ReLU-SaaS

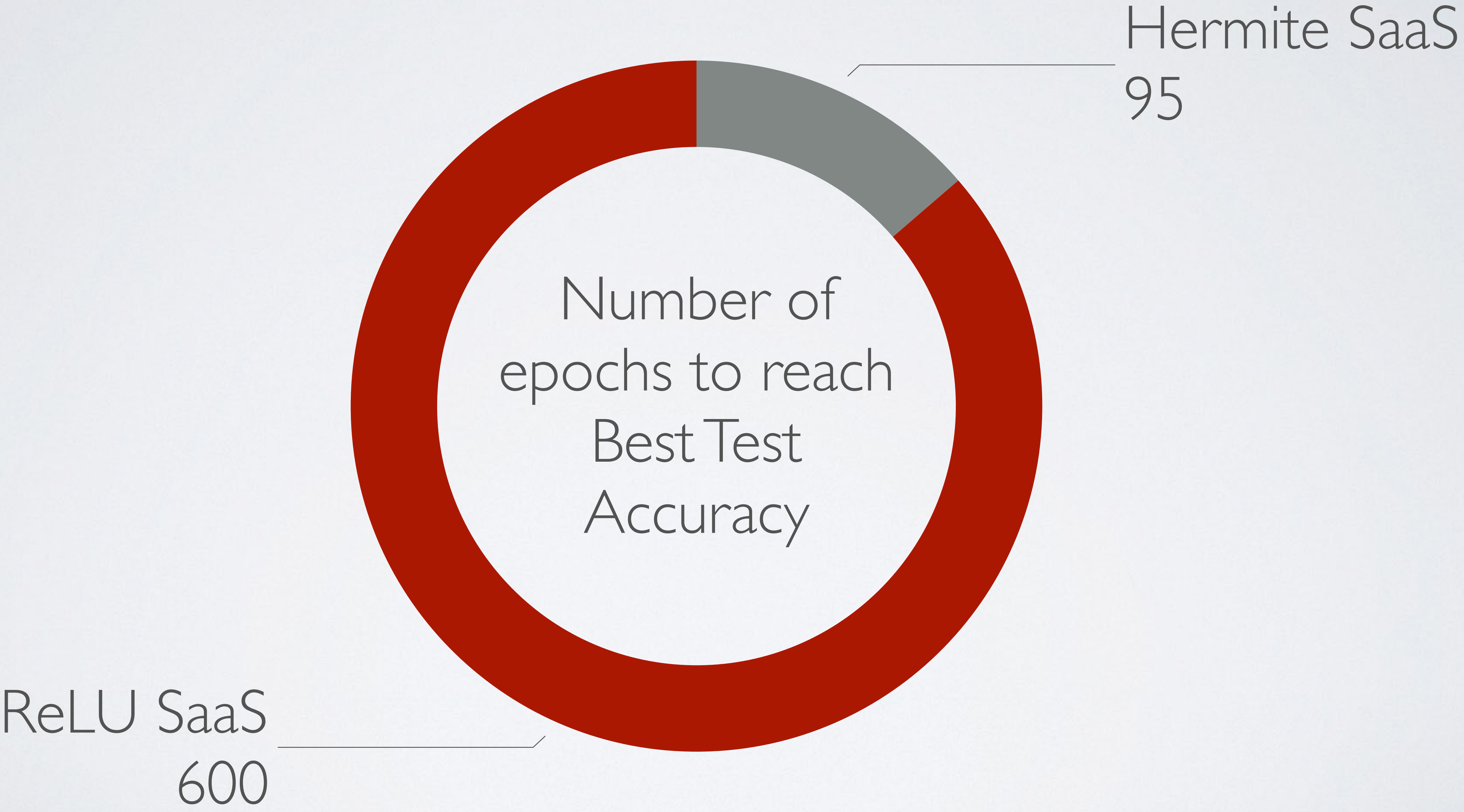
CIFAR10



Hermite-SaaS **More Resilient to Noise** than ReLU-SaaS

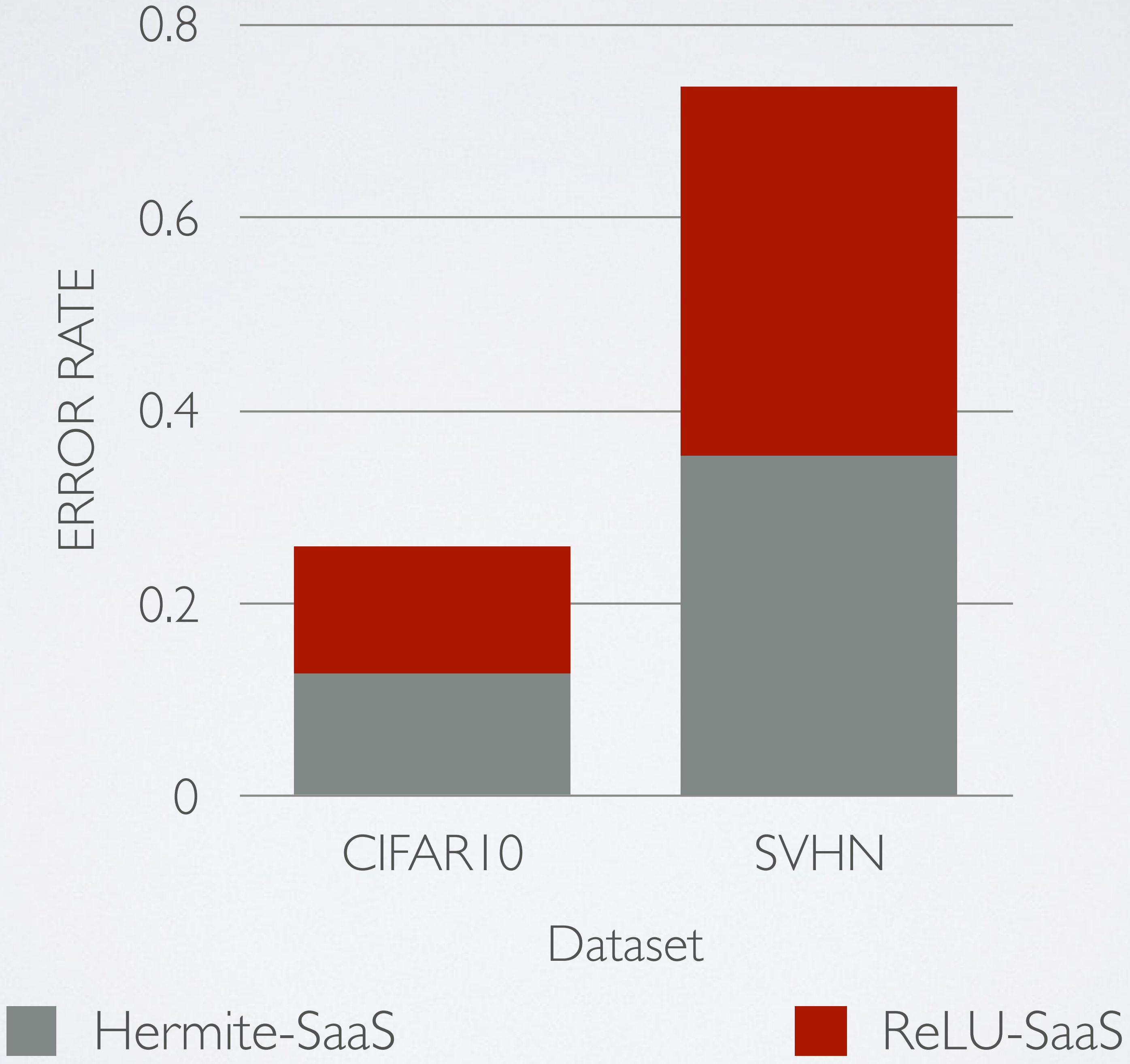
Adding 30% uniform noise to labelled data

Hermite-SaaS More Resilient to Noise than ReLU-SaaS



Hermite-SaaS **GENERALIZES BETTER** than ReLU-SaaS

Hermite-SaaS **GENERALIZES BETTER** than ReLU-SaaS



Please come to our poster!