Incorporating Regional Verification of Neural Network Performance into PAC Bounds

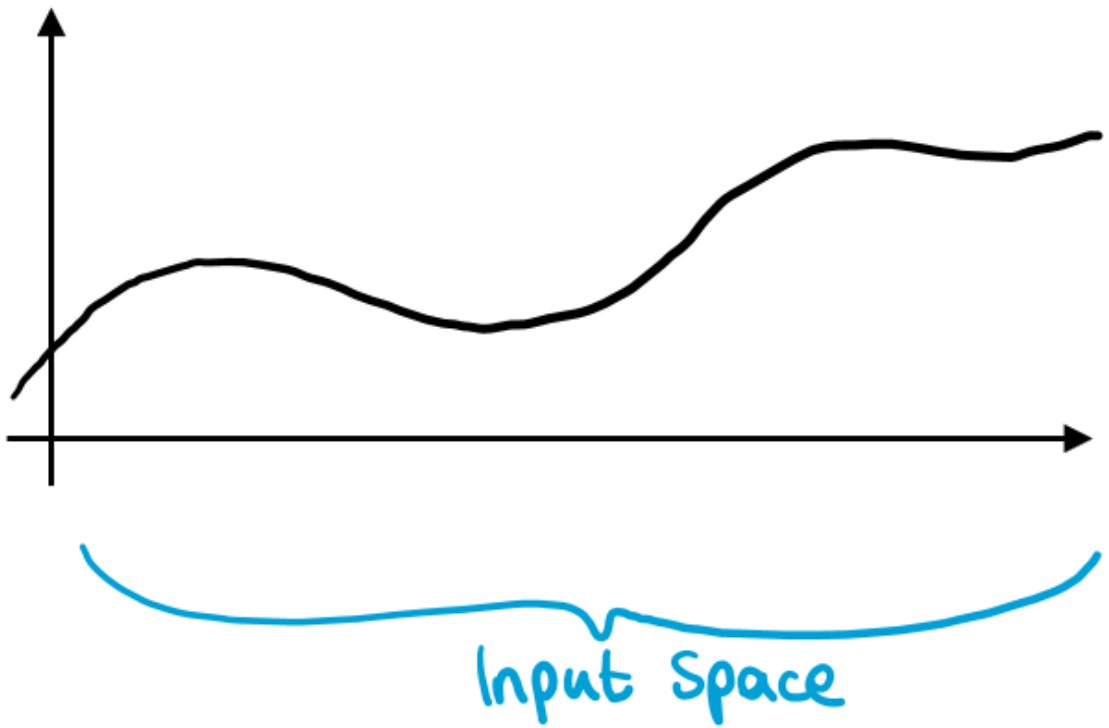
Thomas Walker

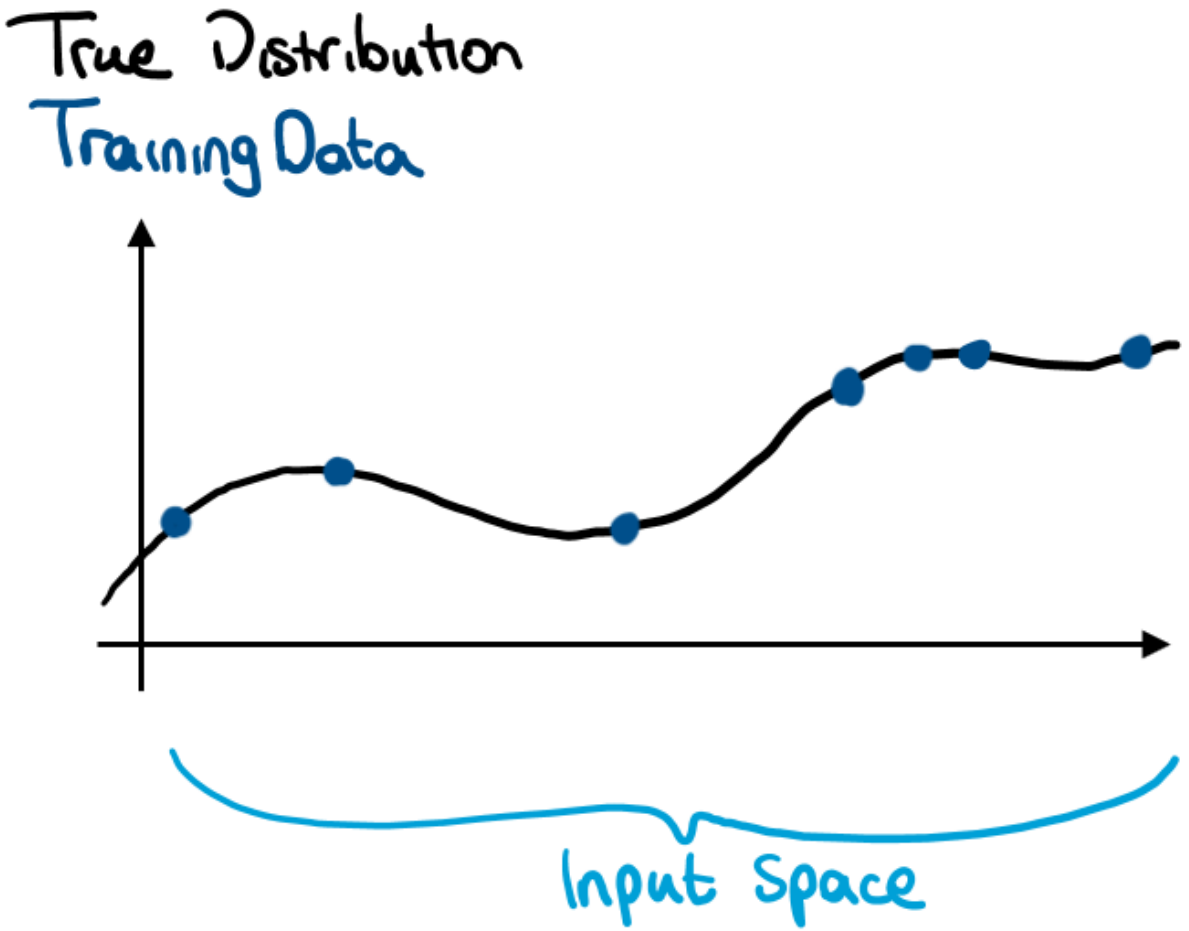
Imperial College London - Verification of Autonomous Systems
Professor Alessio Lomuscio

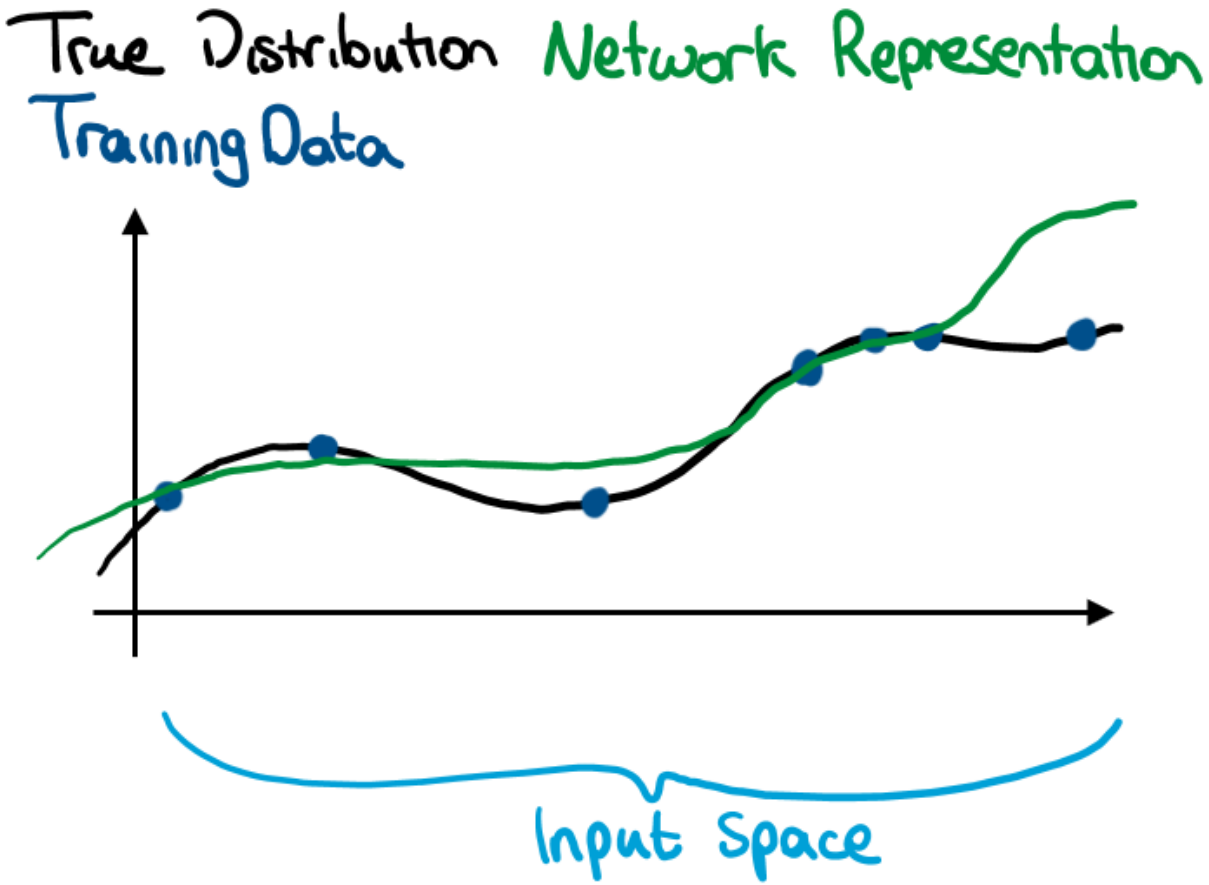
thomas.walker21@imperial.ac.uk

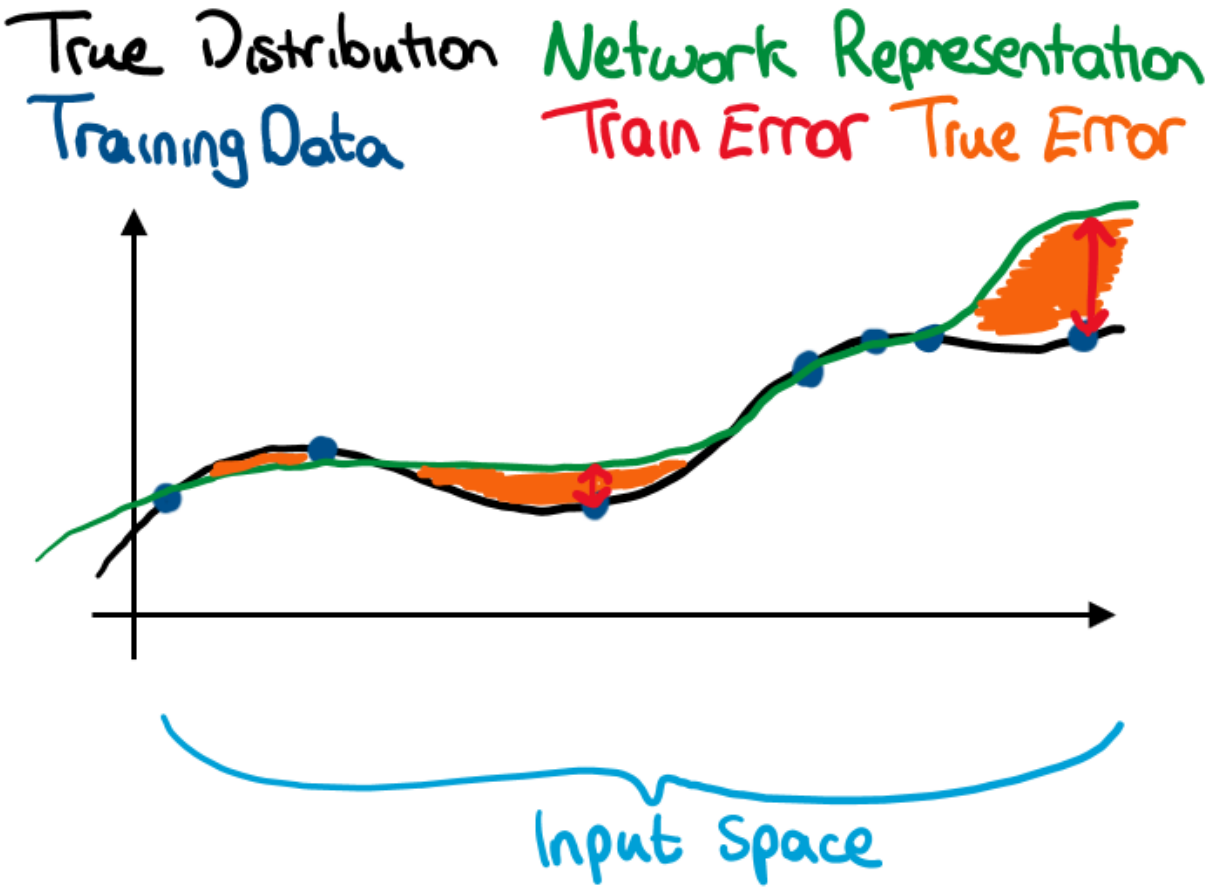
October 25, 2023

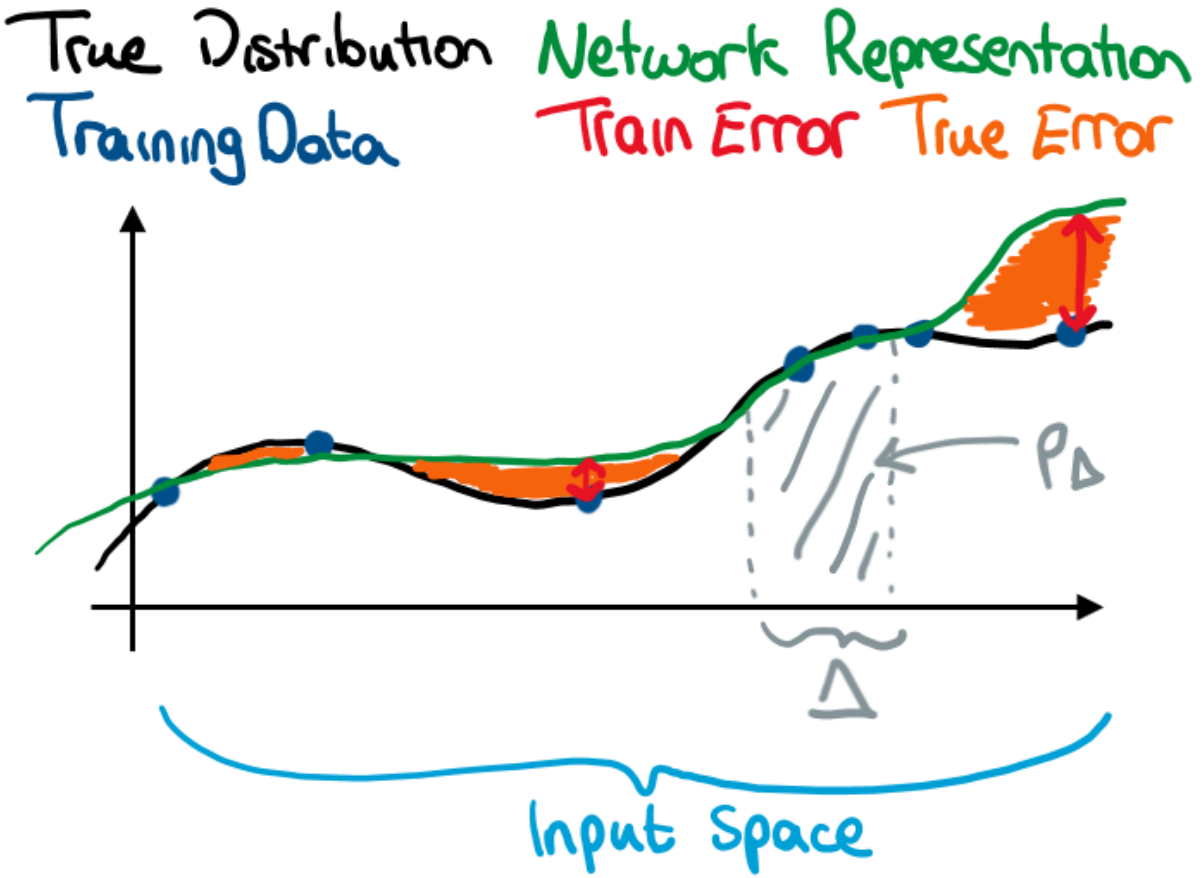
True Distribution




True Distribution
Training Data
Input Space


True Distribution
Network Representation
Training Data
Input Space


True Distribution
Training Data
Network Representation
Train Error
True Error
Input Space


True Distribution
Training Data
Network Representation
Train Error
True Error
P_Δ
Δ
Input Space

Theoretical Result

Theorem

For an $\mathrm{i.i.d}$ *sample of size* m *from the true distribution, given that performance is verified in a region* Δ *of the input space, with probability greater than* $1-\delta$ *we can say that*

$$\textit{True Error} \leq \textit{Train Error} + \textit{Bound}$$

where the bound is given by

$$\sqrt{\frac{\log\left(\frac{(1-p_\Delta)+\sqrt{(1-p_\Delta)^2+4\delta^{\frac{1}{m}}p_\Delta}}{2\delta^{\frac{1}{m}}}\right)}{2}} \leq \sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2m}}.$$