

Assured Autonomy in CPS

Providing Guarantees in Safety Critical Cyber Physical Systems

Student: Mohit Prashant

Supervisor: Assoc. Prof. Arvind Easwaran

Example of Safety Critical CPS



Image of Autonomous Bus from Straits Times

Density-Based Safety Barriers

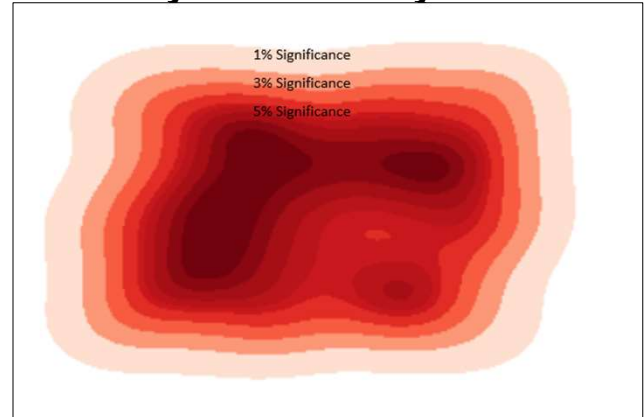


Image of safety barriers at various significance levels constructed in Matplotlib

Project Objectives:

The use of AI in safety critical systems has been able to improve performance of cyber physical systems in various fields like healthcare and avionics. However, it has also raised widespread objection due to potential risks involved from the lack of explainability in blackboxed AI models and the lack of guarantees placed on the learning process.

This project aims to place safety guarantees on variational autoencoders by identifying out-of-distribution data within its latent space and thereby providing an assurance on performance.

What are Safety Guarantees?

Safety guarantees are lower bound estimates on performance that a model may have.

This project considers a two-fold guarantee, firstly, on the error of approximation and, secondly, on the confidence of the guarantee

Estimating an Upper-Bound Error Rate

$$\epsilon = \frac{1}{N} \left(r(1 - e) + \ln \left(\frac{1}{\delta} \right) + \sqrt{\ln^2 \frac{1}{\delta} + 2r(1 - e) \ln \frac{1}{\delta}} \right)$$

Labels in the diagram:

- Error Rate (points to ϵ)
- Sample Size (points to N)
- Confidence (points to δ)
- #Samples in Violation of (points to $r(1 - e)$)
- Significance of Safety Barrier (points to δ)