

Concepts in ICL

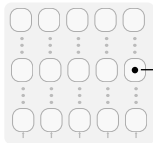
Part-of-Speech tagging

\mathcal{Z} = Noun, Verb, ... Adjective

Bitwise arithmetic

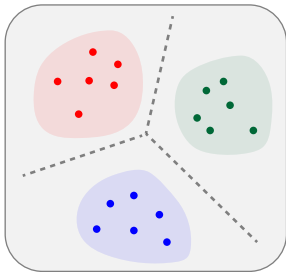
\mathcal{Z} = AND, OR, ..., NAND

Transformers



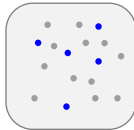
ICL tasks

Concept encoding



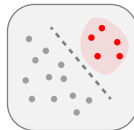
Concept decodability(CD) \rightarrow ICL performance

★ : Not decodable



\rightarrow CD = 0

★ : Decodable



\rightarrow CD = 1

