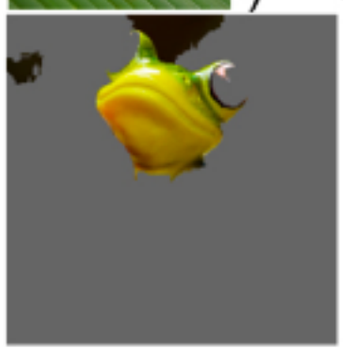
P( ) = 0.54

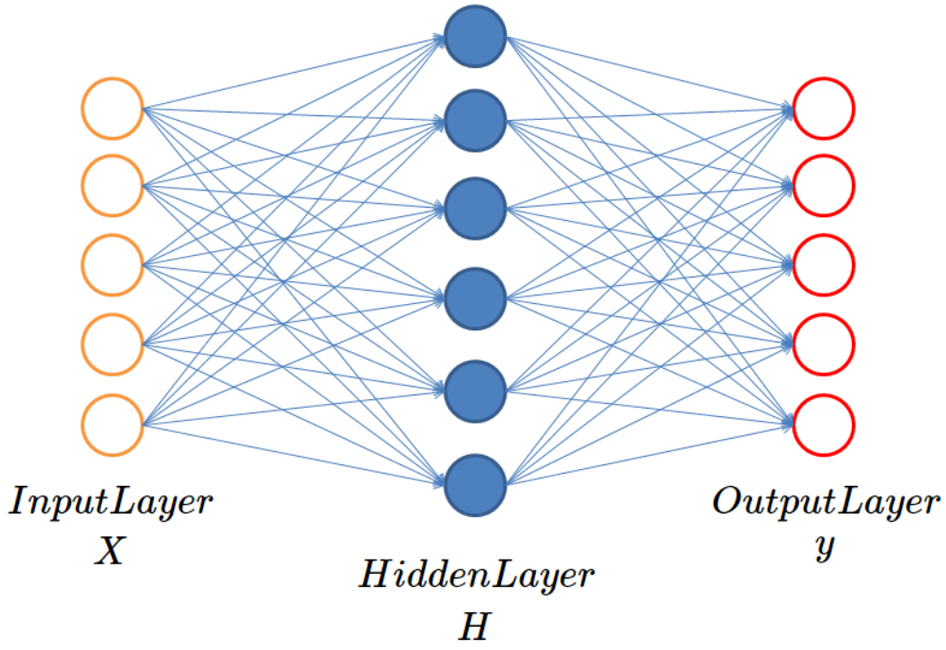


P( ) = 0.07

P( ) = 0.05

LIME

Local **I**nterpretable **M**odel-agnostic **E**xplanations



$$\sim \quad \hat{y} = w_1 x_1 + w_2 x_2$$