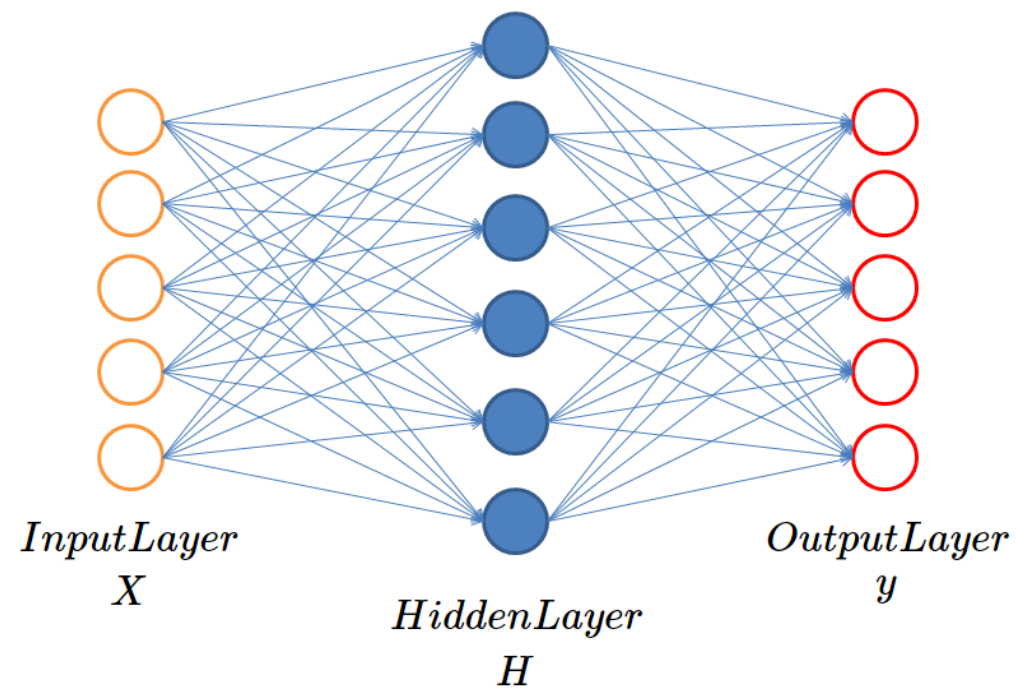


L_{ocal} **I**_{nterpretable} **M**_{odel-agnostic} **E**_{xplanations}



\sim

$$\hat{y} = w_1x_1 + w_2x_2$$

Integrated Gradients



Axiomatic Attribution for Deep Networks
(Jun 2017) by Sundararajan et al

