

Default Prediction Baseline

baseline vector
 x'

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}$$

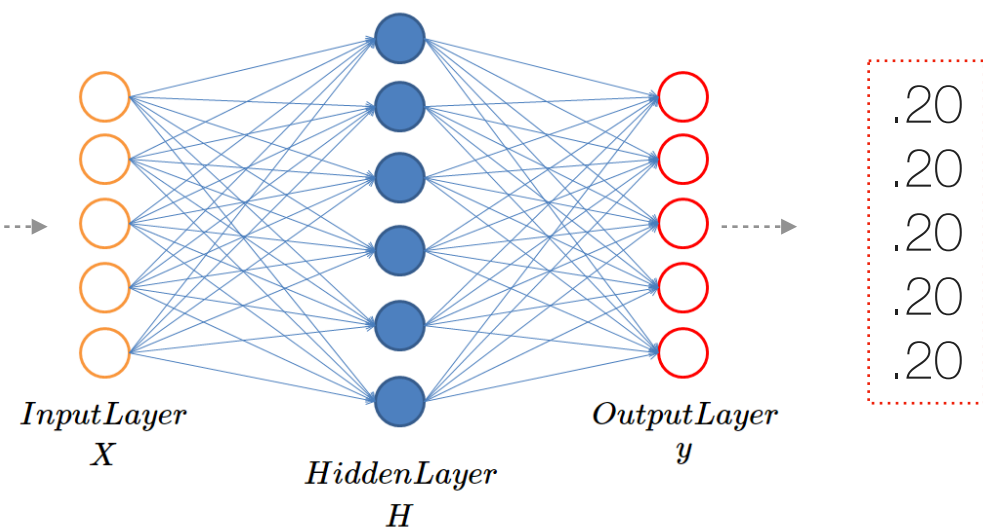
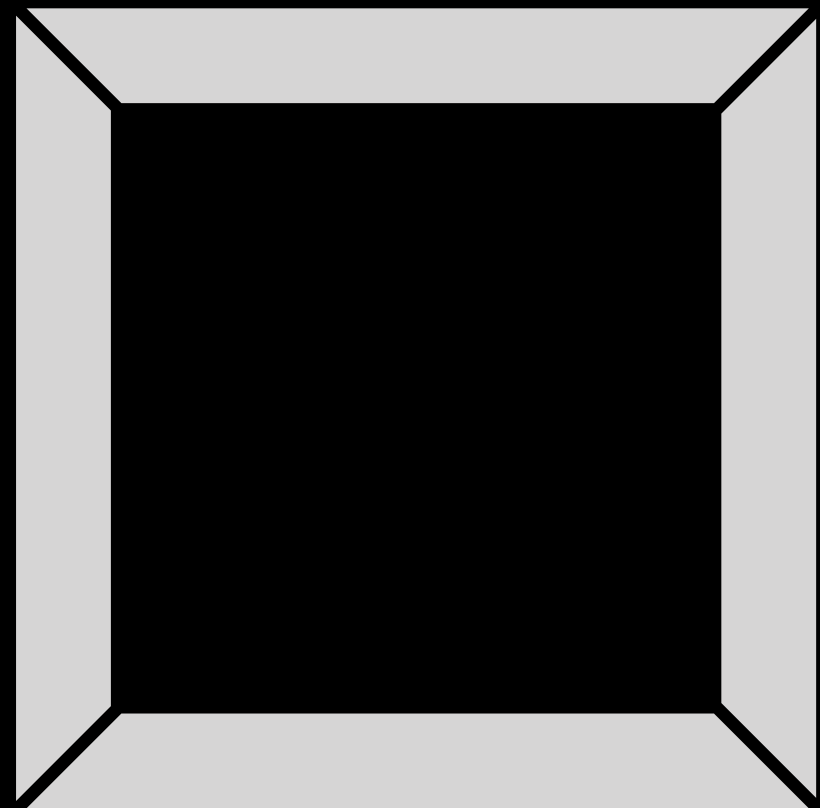


Image Classification Baseline



$$\text{IntegratedGrads}_i(x) ::= (x - x') \times \int_{\alpha=0}^1 \frac{\overset{\text{baseline}}{\partial F(x' + \alpha \times (x - x'))}}{\overset{\text{input}}{\partial x_i}} d\alpha$$

where $\frac{\partial F(x)}{\partial x_i}$ is the gradient of F along the i^{th} dimension at x .

