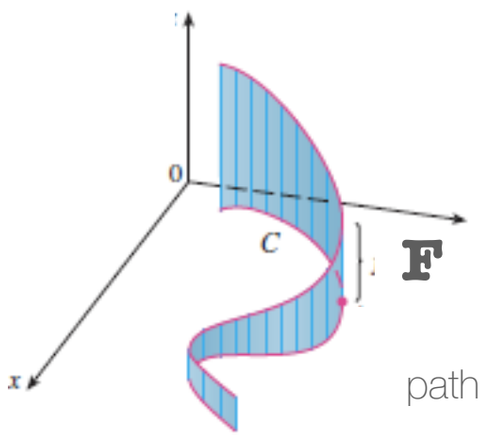
baseline input

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

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



path is from x' to x

Riemann Sums

↓

20-300 steps is enough to estimate true attribution within 5%