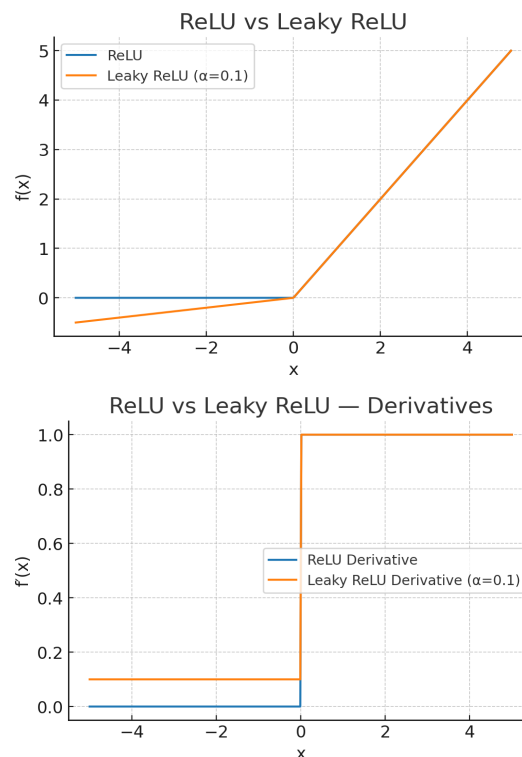


# Activation Functions & Derivatives — Real World Analogies

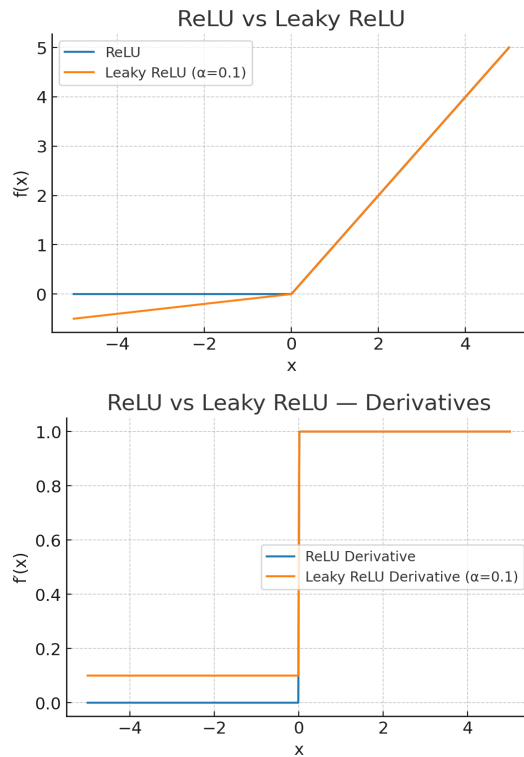
## ReLU (Rectified Linear Unit)

Activation: Like an automatic door that only opens if you push forward (positive). If you push backwards (negative), it stays shut. Derivative: Once the door is closed ( $x < 0$ ), no matter how hard you push, it doesn't move (slope = 0). When it's open ( $x > 0$ ), it moves freely (slope = 1).



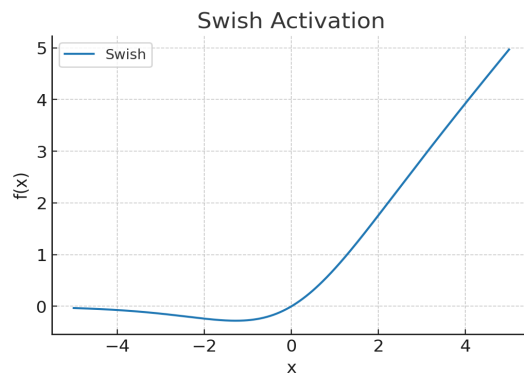
## Leaky ReLU

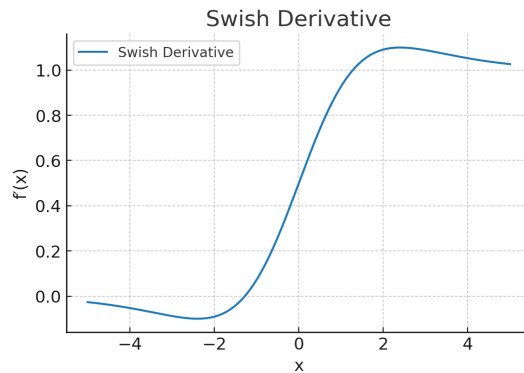
Activation: Same door, but with a small side vent — even if you push backwards, a little airflow comes through. Derivative: That tiny airflow means the gradient never completely dies — there's always some signal to adjust weights.



## Swish

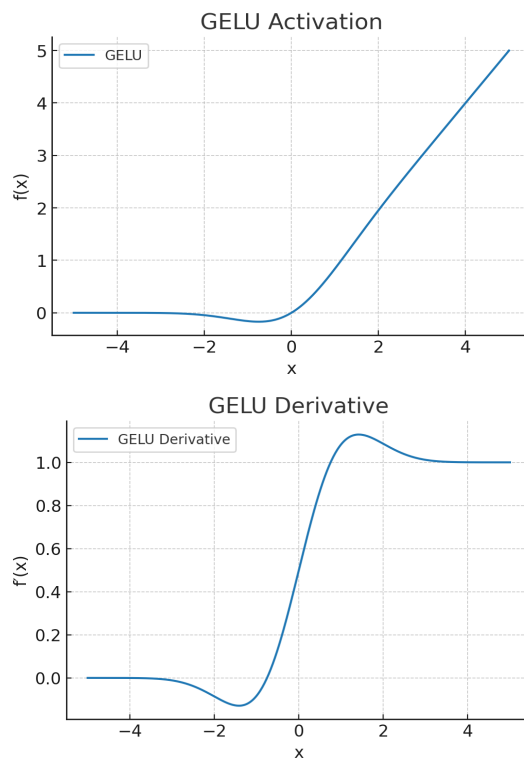
Activation: Think of a water tap with a smooth knob. Small pushes give a trickle, bigger pushes give a steady stream — no sudden ON/OFF. Derivative: Since flow changes smoothly, the adjustment (gradient) is never flat zero. Training feels smoother — less jerky than ReLU.





## GELU (Gaussian Error Linear Unit)

Activation: Imagine a smart filter in a call center. Calls (signals) get through based on how confident the filter is that they're important. Derivative: The filter doesn't fully block weak signals, it just reduces them softly. That keeps some learning signal alive, but prioritizes stronger inputs.



## Sigmoid (for contrast)

Activation: Like a saturation dial on a photo editor. Small values adjust brightness, but after a point the image looks unchanged. Derivative: Once saturated (very dark or very bright), no matter how much you

turn, nothing changes (gradient  $\approx 0$ ).

