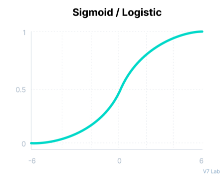
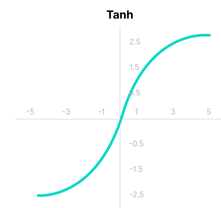
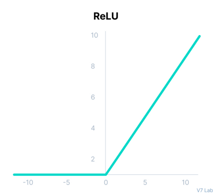
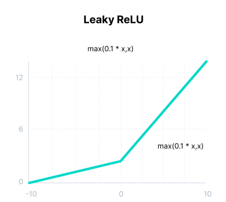
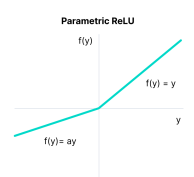
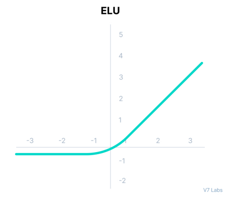
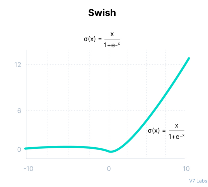
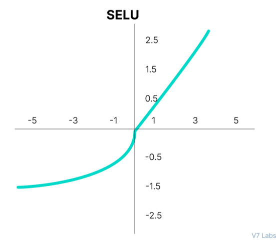
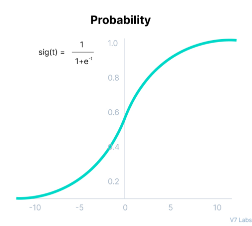
Activation functions:

1. **Sigmoid Function**: This function maps the input to a range between 0 and 1 and was commonly used in the past as an activation function for neural networks.	*Sigmoid / Logistic* $f(x) = \frac{1}{1+e^{-x}}$	
2. **Hyperbolic Tangent (tanh) Function**: Similar to the sigmoid function, but it maps the input to a range between -1 and 1, and it is still used in certain situations today.	*Tanh* $f(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})}$	
3. **Rectified Linear Unit (ReLU)**: One of the most popular activation functions, it's simple and effective, replacing negative values with zero to introduce non-linearity.	*ReLU* $f(x) = max\ (0, x)$	
4. **Leaky ReLU**: A variation of ReLU, it allows a small, non-zero gradient for negative inputs to prevent neurons from getting stuck during training.	*Leaky ReLU* $f(x) = max\ (0.1x, x)$	
5. **Parametric ReLU (PReLU)**: Similar to Leaky ReLU, but with the slope for negative inputs as a learnable parameter.	*Parametric ReLU* $f(x) = max\ (ax, x)$	

6. **Exponential Linear Unit (ELU)**: This function smooths the output for negative values and can speed up learning compared to ReLU.	ELU $\begin{cases} x & for\ x \geqslant 0 \\ \alpha\left(e^x - 1\right) & for\ x < 0 \end{cases}$	
7. **Swish**: A newer activation function that combines elements of ReLU and sigmoid to achieve better performance in some cases.	$Swish$ $f(x) \ = \ x{*}sigmoid(x)$	
8. **Scaled Exponential Linear Unit (SELU)**: A self-normalizing activation function that maintains a constant mean and variance in deep neural networks.	$SELU$ $f(\alpha, x) = \lambda \begin{cases} \alpha\left(e^x - 1\right) & for\ x < 0 \\ x & for\ x \geqslant 0 \end{cases}$	
9. **Softmax**: Commonly used in the output layer for multi-class classification, it converts the output into a probability distribution.	**Softmax** $\text{softmax}(z_i) = \dfrac{\exp(z_i)}{\Sigma_j \exp(z_j)}$	
10. **Maxout**: An activation function that generalizes ReLU and Leaky ReLU by taking the maximum value of multiple linear pieces.	$h(x) \ = \ \max\ \left(Z_1\, , Z_2\, , \ldots , Zn\right)$ $h(x) \ = \ \max\ \left(W_1 \cdot x + b_1, W_2 \cdot x + b_2, \ldots , Wn \cdot x + bn\right)$