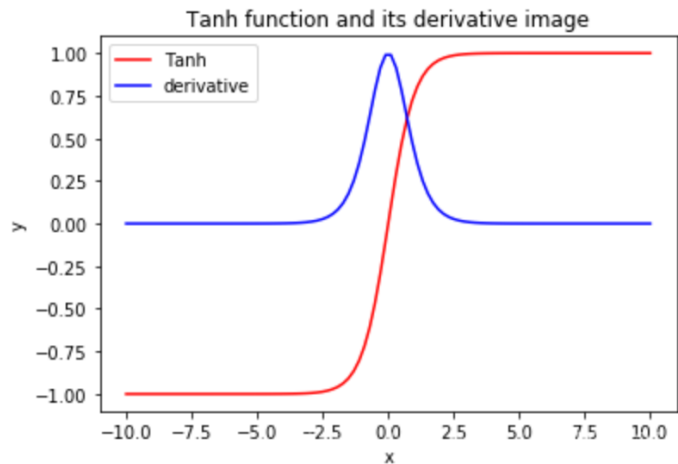


网络应用中，训练过程使用的是反向传播算法，通过链式法则回传的梯度不断相乘，因此，Sigmoid 函数作为激活函数时会导致梯度消失问题，尤其在网络比较深时会达不到训练效果。

(2) 对于双曲正切函数， $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

$$\frac{\partial \tanh(x)}{\partial x} = \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2} = 1 - \tanh^2(x)$$

双曲正切函数作为激活函数的优点是中心对称、单调、连续、可导且具有非线性特点。但其缺点是它的导数函数曲线见右图， x 变化更



小的范围内，导数才有值且不大于 1，在神经网络应用中，训练过程使用的是反向传播算法，通过链式法则回传的梯度不断相乘，因此，双曲正切函数作为激活函数时同样会导致梯度消失问题，尤其在网络比较深时会达不到训练效果。

(3) 对于 ReLU 函数， $ReLU(x) = \max(0, x)$

$$\frac{\partial ReLU(x)}{\partial x} = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

ReLU 函数作为激活函数的优点是梯度不会饱和，解决了梯度消失问题，没有指数计算，计算复杂度低。它的缺点是非中心对称，负数部分的梯度为

