6.2 Multi-layer Networks and Activation Functions

In this section, we focus on two simple extensions

1. Nonlinear transformation instead of linear

   linear: $Ax = y$, nonlinear: $y = f(A, x)$

   $f(\cdot)$: specified activation function / transfer function

   • Some standard activation functions:

   $f(x) = x$ linear

   $f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 & \text{for } x > 0 \end{cases}$ binary step

   $f(x) = \dfrac{1}{1 + \exp(-x)}$ logistic (soft step)

   $f(x) = \tanh(x)$ tanh

   $f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x & \text{for } x > 0 \end{cases}$ rectified linear unit (ReLU)

2. Multiple layers.