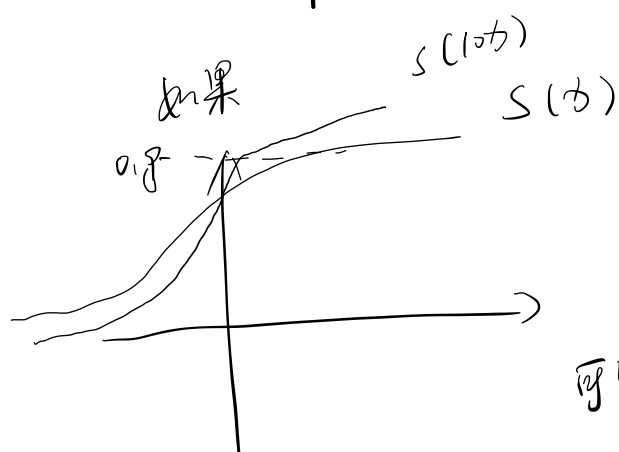


1. sigmoid 和 softmax 都对二分类有效.

$$\text{sigmoid} = \frac{1}{1 + e^{-b_1}} \rightarrow \text{学习 } -b_1$$

$$\text{softmax} = \frac{e^{b_1}}{e^{b_1} + e^{b_2}} = \frac{1}{1 + e^{b_2 - b_1}} \rightarrow \text{学习 } (b_2 - b_1)$$

2. 但是今天遇到 sigmoid 曲线的极限率在 0.8 左右,



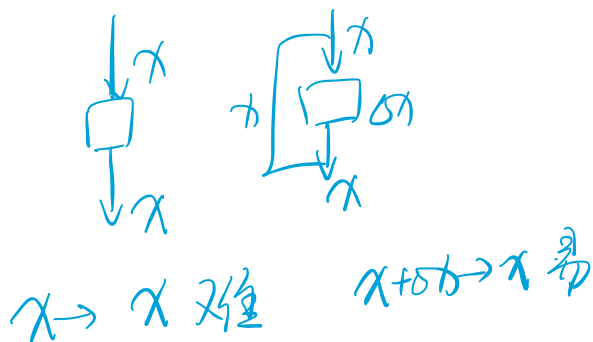
将 $\text{sigmoid}(x)$
↓

$\text{sigmoid}(100)$

强制让函数在 prob=0.8 时继续学习,
可以让 prob 上升, 但是更过拟合.

3. 如果把 sigmoid 换为 softmax, 则没有这个问题.

可解释原因:



可解释同样的思路.

学习 $-b_1$ 难,

学习 $b_2 - b_1$ 易, 因为有相互联系.