[0,1] 之间但无法训练

$$P(y=1 \mid \boldsymbol{x}) = \max\left\{0, \min\{1, \boldsymbol{w}^\top \boldsymbol{h} + b\}\right\}. \tag{6.18}$$

这的确定义了一个有效的条件概率分布，但我们并不能使用梯度下降来高效地训练它。任何时候当 $\boldsymbol{w}^\top \boldsymbol{h} + b$ 处于单位区间外时，模型的输出对它的参数的梯度都将为 $\mathbf{0}$。梯度为 $\mathbf{0}$ 通常是有问题的，因为学习算法对于如何提高相应的参数没有了指导。

与之相对的，最好是使用一种不同的方法来保证无论何时模型给出了错误的答案时总能有一个很强的梯度。这种方法是基于使用sigmoid输出单元结合最大似然来实现的。

sigmoid输出单元定义为

$$\hat{y} = \sigma\left(\boldsymbol{w}^\top \boldsymbol{h} + b\right), = \frac{1}{1+e^{-z}}. \tag{6.19}$$

输出为 (0,1) 间来

sigmoid $\longrightarrow$ Bernoulli 分布