

assignment3

Xiaoma

2022 年 10 月 23 日

题目 1.

解: 理想中的激活函数是阶跃函数, 但阶跃函数非连续, 在 0 处不可导, 线性激活函数没法完全拟合阶跃函数, 在 0 的周围 sigmoid 函数的变化急剧, 满足要求。

题目 2.

解:

- 对于

$$\text{softmax} : \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)}$$

假设 $x = c$, softmax 的理论值为 $\frac{1}{n}$, 若 c 趋近负无穷, 此时 softmax 分母为 0, 下溢, 若 c 趋近于正无穷, 此时 softmax 上溢。

解决方法: 令 $z = x - \max x_i$, 则

$$\text{softmax} : \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)}$$

当 c 较小时, 分母至少有一项为 1, 解决了下溢。当 c 较大时, 相当于分子分母同时除一个非常大的数, 解决了上溢。

- 对于

$$\log \sum_{j=1}^n \exp(x_j)$$

假设 $x = c$, $\log \sum_{j=1}^n \exp(x_j)$ 的理论值为 $c + \log n$, 若 c 趋近负无穷, 此时 $\log 0$ 未定义, 则发生下溢, 若 c 趋近于正无穷, 则发生上溢。

解决方法: 令 $z = x - \max x_i$, 当 c 较小时, 则底数至少为 1, 解决了下溢, 当 c 较大时, 相当于指数都减去了一个大的数, 解决了上溢。

题目 3.

解: 令 $f(x) = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)}$, 则

$$\frac{\partial f(x)}{\partial x_k} = \begin{cases} -\frac{\exp(x_i+x_k)}{\sum_{j=1}^n \exp(x_j)^2} & k \neq i \\ -\frac{\exp(x_k) \sum_{j=1, j \neq k}^n \exp(x_j)}{\sum_{j=1}^n \exp(x_j)^2} & k = i \end{cases}$$

令 $g(x) = \log \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)}$, 则

$$\frac{\partial g(x)}{\partial x_k} = \begin{cases} -\frac{\exp(x_k)}{\sum_{j=1}^n \exp(x_j)} & k \neq i \\ 1 - \frac{\exp(x_k)}{\sum_{j=1}^n \exp(x_j)} & k = i \end{cases}$$

题目 4.

解: 设初始权重 $w1 = 0.6, w2 = 0.1, w3 = 0.2, w4 = 0.7, w5 = 0.5, w6 = 0.8$ 。

$$\begin{aligned} error_3 &= \frac{\partial L}{\partial w} = y - \hat{y} \\ error_1 &= error_3 \times w_5 \times \frac{\partial Relu}{\partial x} = error_3 \times w_5 \\ error_2 &= error_3 \times w_6 \times \frac{\partial Relu}{\partial x} = error_3 \times w_6 \\ w'_1 &= error_1 \times A + w_1 \end{aligned}$$

$$w'_2 = error_2 \times A + w_2$$

$$w'_3 = error_1 \times B + w_3$$

$$w'_4 = error_2 \times B + w_4$$

$$w'_5 = error_3 \times output_1 + w_5$$

$$w'_6 = error_3 \times output_2 + w_6$$

$$1. \text{ input : } 0.2 \times 0.6 + 0.3 \times 0.2 = 0.18$$

$$\text{output : } 0.18$$

$$\text{error : } 0.226 \times 0.5 = 0.113$$

$$w1 : 0.113 \times 0.2 + 0.6 = 0.6226$$

$$w2 : 0.1808 \times 0.2 + 0.1 = 0.13616$$

$$2. \text{ input : } 0.2 \times 0.1 + 0.3 \times 0.7 = 0.23$$

$$\text{output : } 0.23$$

$$\text{error : } 0.226 \times 0.8 = 0.1808$$

$$w3 : 0.113 \times 0.3 + 0.2 = 0.2399$$

$$w4 : 0.1808 \times 0.3 + 0.7 = 0.75424$$

$$3. \text{ input : } 0.18 \times 0.5 + 0.23 \times 0.8 = 0.274$$

$$\text{output : } 0.274$$

$$\text{error : } 0.5 - 0.274 = 0.226 \quad w5 : 0.226 \times 0.18 + 0.5 = 0.54068 \quad w6 :$$

$$0.226 \times 0.23 + 0.8 = 0.85198$$

$$1. \text{ input : } 0.2 \times 0.6226 + 0.3 \times 0.2399 = 0.19649$$

$$\text{output : } 0.19649$$

$$2. \text{ input : } 0.2 \times 0.13616 + 0.3 \times 0.75424 = 0.253504$$

$$\text{output : } 0.253504$$

$$3. \text{ input : } 0.19649 \times 0.54068 + 0.253504 \times 0.85198 = 0.32222$$

loss 由 0.226 降为 0.17778