homework#2: 理论题

By: 黄哲昊 (518021910660)

2020 年 10 月

1 理论题

1.1 现假设样本来自三个类，某次训练中的一个 batch 包含 3 个训练样本$\boldsymbol{x_1}$，$\boldsymbol{x_2}$，$\boldsymbol{x_3}$，分别来自第 1，2，3 类

1. 试推导采用单热向量编码时该 batch 交叉熵损失函数表达式。(提示：设该 batch 对应网络输出为$\boldsymbol{y_1}$,$\boldsymbol{y_2}$,$\boldsymbol{y_3}$)
2. 如果网络输出为$\boldsymbol{y_1}= (0.65, 0.43, 0.11)$，$\boldsymbol{y_2}= (0.05, 0.51, 0.18)$，$\boldsymbol{y_3}= (0.33, 0.21, 0.72)$，计算交叉熵损失函数值。

1.1.1

假设第 i 个训练样本对应网络输出为$\boldsymbol{y_i}= (y_{i1}, y_{i2}, y_{i3})$，则属于第 j 类 (j=1,2,3) 的概率为

$$P(j|\boldsymbol{y_i}) = \frac{e^{y_{ij}}}{\sum_{k=1}^{3} e^{y_{ik}}} \tag{1}$$

三个样本真实值的单热向量输出分别为$\boldsymbol{\hat{y}_1}$=$(1, 0, 0)$，$\boldsymbol{\hat{y}_2}$=$(0, 1, 0)$，$\boldsymbol{\hat{y}_3}$=$(0, 0, 1)$。则这个 batch 的交叉熵损失为：

$$\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{3} H(\hat{y}_i, P(\boldsymbol{y_i})) && (2)\\
&= -(\log \frac{\boldsymbol{\hat{y}_1}(e^{\boldsymbol{y_1}})^T}{\sum_{k=1}^{3} e^{y_{1k}}} + \log \frac{\boldsymbol{\hat{y}_2}(e^{\boldsymbol{y_2}})^T}{\sum_{k=1}^{3} e^{y_{2k}}} + \log \frac{\boldsymbol{\hat{y}_3}(e^{\boldsymbol{y_3}})^T}{\sum_{k=1}^{3} e^{y_{3k}}}) && (3)\\
&= -(\log \frac{e^{y_{11}}}{\sum_{k=1}^{3} e^{y_{1k}}} + \log \frac{e^{y_{22}}}{\sum_{k=1}^{3} e^{y_{2k}}} + \log \frac{e^{y_{33}}}{\sum_{k=1}^{3} e^{y_{3k}}}) && (4)
\end{aligned}$$

1.1.2

将网络输出$\boldsymbol{y_1}= (0.65, 0.43, 0.11)$，$\boldsymbol{y_2}= (0.05, 0.51, 0.18)$，$\boldsymbol{y_3}= (0.33, 0.21, 0.72)$，代入上式 (4)，得：

$$\begin{aligned}
\mathcal{L} &= -(\log \frac{e^{0.65}}{e^{0.65} + e^{0.43} + e^{0.11}} + \log \frac{e^{0.51}}{e^{0.05} + e^{0.51} + e^{0.18}} + \log \frac{e^{0.72}}{e^{0.33} + e^{0.21} + e^{0.72}}) && (5)\\
&= 2.54692 && (6)
\end{aligned}$$

1.2 假设输入有 2 个样本 x_1，x_2

1. 请画出计算 x_1，x_2 标准差的详细计算图；
2. 标出当 $x_1 = -1$，$x_2 = 3$ 时，输出对图中每个节点输入变量的梯度值，并求出 x_1，x_2 总的梯度值。

1.2.1

计算 x_1，x_2 标准差的计算图如下：

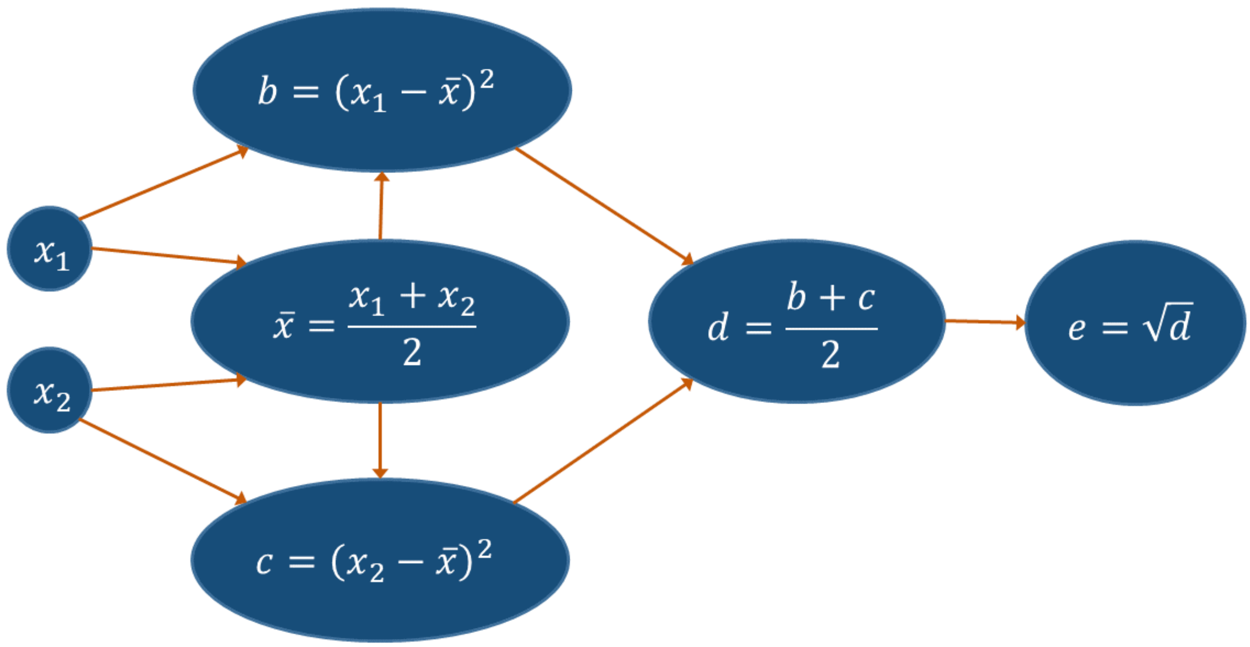

图 1: 计算图

1.2.2

当 $x_1 = -1$，$x_2 = 3$ 时，正向传播和反向传播后得到的计算图如图所示，其中红色数字表示每个节点的输出结果，蓝色数字则表示每个节点计算得到的上游和下游梯度：

对梯度进行计算得到：

$$\begin{aligned}\frac{\partial e}{\partial \bar{x}} &= \frac{\partial e}{\partial e}\frac{\partial e}{\partial d}\frac{\partial d}{\partial b}\frac{\partial b}{\partial \bar{x}} + \frac{\partial e}{\partial e}\frac{\partial e}{\partial d}\frac{\partial d}{\partial c}\frac{\partial c}{\partial \bar{x}} &(7)\\ &= 1 \times \frac{1}{4} \times \frac{1}{2} \times 4 + 1 \times \frac{1}{4} \times \frac{1}{2} \times (-4) &(8)\\ &= 0 &(9)\end{aligned}$$

$$\begin{aligned}\frac{\partial e}{\partial x_1} &= \frac{\partial e}{\partial e}\frac{\partial e}{\partial d}\frac{\partial d}{\partial b}\frac{\partial b}{\partial x_1} + \frac{\partial e}{\partial \bar{x}}\frac{\partial \bar{x}}{\partial x_1} &(10)\\ &= 1 \times \frac{1}{4} \times \frac{1}{2} \times (-4) + 0 \times \frac{1}{2} &(11)\\ &= -0.5 &(12)\end{aligned}$$

$$\begin{aligned}\frac{\partial e}{\partial x_2} &= \frac{\partial e}{\partial e}\frac{\partial e}{\partial d}\frac{\partial d}{\partial b}\frac{\partial b}{\partial x_1} + \frac{\partial e}{\partial \bar{x}}\frac{\partial \bar{x}}{\partial x_2} &(13)\\ &= 1 \times \frac{1}{4} \times \frac{1}{2} \times 4 + 0 \times \frac{1}{2} &(14)\\ &= 0.5 &(15)\end{aligned}$$

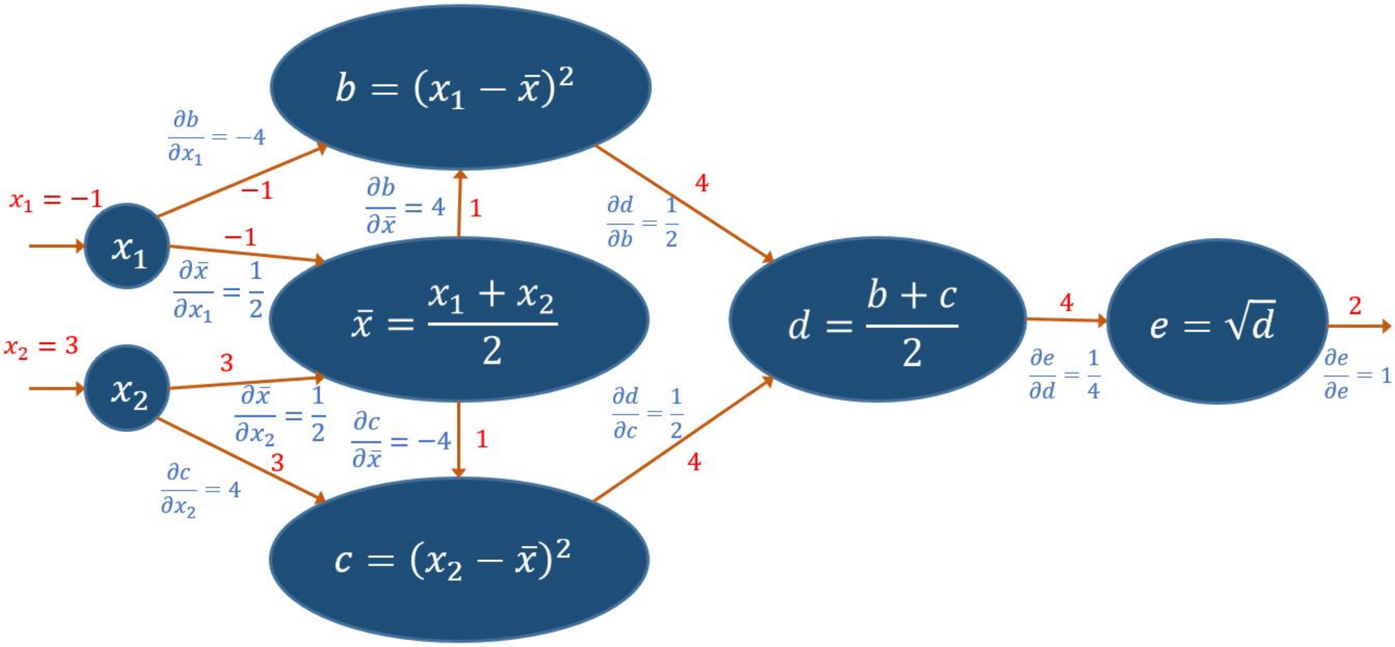
$b = (x_1 - \bar{x})^2$
$\frac{\partial b}{\partial x_1} = -4$
$x_1 = -1$
-1
$\frac{\partial b}{\partial \bar{x}} = 4$
1
$\frac{\partial d}{\partial b} = \frac{1}{2}$
4
x_1
-1
$\frac{\partial \bar{x}}{\partial x_1} = \frac{1}{2}$
$\bar{x} = \frac{x_1 + x_2}{2}$
$d = \frac{b + c}{2}$
4
$e = \sqrt{d}$
2
$x_2 = 3$
3
x_2
$\frac{\partial e}{\partial d} = \frac{1}{4}$
$\frac{\partial e}{\partial e} = 1$
$\frac{\partial \bar{x}}{\partial x_2} = \frac{1}{2}$
$\frac{\partial c}{\partial \bar{x}} = -4$
1
$\frac{\partial d}{\partial c} = \frac{1}{2}$
3
$\frac{\partial c}{\partial x_2} = 4$
4
$c = (x_2 - \bar{x})^2$

图 2: 梯度图