机器学习第四章神经网络编程报告

162050127 颜劭铭

2022 年 5 月 19 日

1 报告

1.1 sigmoid，deriv_sigmoid，mse_loss 函数编写

对于 sigmoid 函数，deriv_sigmoid 函数和 mse_loss 函数由于书上已经写好了公式：

$$sigmoid = \frac{1}{1+\exp^{-x}}$$

$$deriv\_sigmoid = sigmoid * (1 - sigmoid)$$

$$mse\_loss = \frac{1}{m}\sum_{k=1}^{m}\left(\frac{1}{2}\sum_{j=1}^{l}\left(\hat{y_j^k} - y_j^k\right)^2\right)$$

所以只需要按照公式编写即可. 其中由于训练集的输入是向量，所以在提取的时候要使用 zip 函数进行提取，这个点一开始困扰了很久，后面看了下面的代码才明白.

1.2 向前传播过程

向前传播的过程主要是基于提供的如图所示的简单神经网络进行编写（感觉图画的有问题，b_1, b_2, b_3 指的应该是阈值，所以我重新画了一个）.

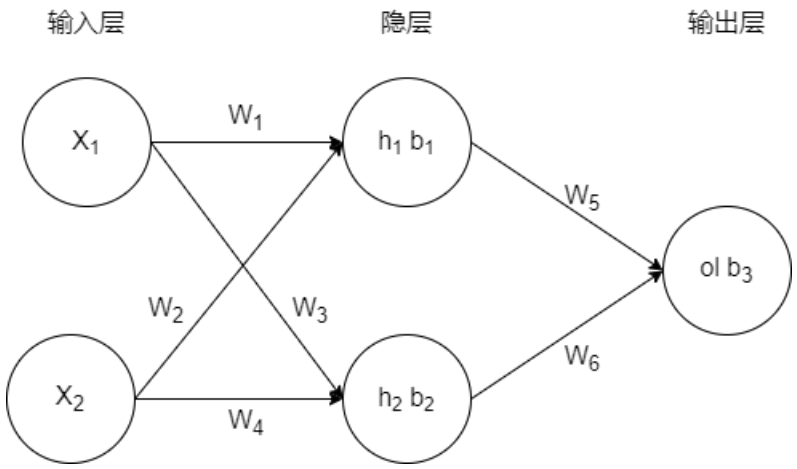

图 1: 简单神经网络示意图

由代码可知，输入的训练集二维特征向量用 X 表示，因此图 1中的 x_1 和 x_2 指的是第一维向量 X[0] 和第二维向量 X[1].

对于隐层结点和输出层结点的输入，根据箭头可以知道：

$$sum\_h_1 = w_1 \times x_1 + w_2 \times x_2$$
$$sum\_h_2 = w_3 \times x_1 + w_4 \times x_2$$
$$sum\_ol = w_5 \times h_1 + w_6 \times h_2$$

由于隐层结点和输出层结点的激活函数都为 sigmoid 函数，因此对于隐层结点和输出层结点的输出表示如下：

$$h_1 = sigmoid(sum\_h_1 + b_1)$$
$$h_2 = sigmoid(sum\_h_2 + b_2)$$
$$y\_pred = sigmoid(sum\_ol + b_3)$$

由上述公式可以表达出整个向前传播的过程得到预测值 ol，这部分没有遇到什么问题，除了一开始看到图上有 6 个阈值搞晕了，后面搞清楚了以后就编写得很快.

1.3 反向传播更新参数

BP 神经网络更新参数主要是基于链式求导进行更新，所以这边给出简要的公式推导.(等到编写完了以后才发现这个代码的变量名设计得非常方便，对于链式法则求导引用变量名非常

方便)

$$d\_L\_d\_ypred = \frac{\partial E_k}{\partial \hat{y_j^k}} = y\_pred - y\_true$$

输出层梯度

$$d\_ypred\_d\_w5 = \frac{\partial \hat{y_j^k}}{\partial w_5} = h_1 \times deriv_sigmoid(sum_ol + b_3)$$

隐层梯度

$$d\_h1\_d\_w1 = \frac{\partial h_1}{\partial w_1} = x_1 \times deriv_sigmoid(sum_h1 + b_1)$$

更新输出层梯度

$$w_5 = w_5 - learn\_rate \times d\_L\_d\_ypred \times d\_ypred\_d\_w5$$

更新隐层梯度

$$w_1 = w_1 - learn\_rate \times d\_ypred\_d\_h1 \times d\_h1\_d\_w1$$

因此只需要按照公式进行编写就可以将参数进行更新

1.4 选定阈值进行转化

由于输出对率结果已经被存入 y_pred 列表中，因此只需要选定一个合适的阈值将输出对率结果转化为预测结果，这里由于没有测试集的标签，所以也不知道选取恰当的阈值是多少，只好去常用的 0.5 进行转化，大于等于 0.5 的输出对率结果转化为预测结果 1，小于 0.5 的输出对率结果转化为预测结果 0.

1.5 改进方法

由于没有测试集标签，所以只能写了一个检验用网络预测训练集的准确度，但是这个是没啥用的，毕竟训练集精度太高就过拟合了.

①新增了一层隐层结点，输入和输出以及梯度更新方法如 1.2 和 1.3 所述，最后训练集精度实际上和单隐层时差不多，没什么大变化，增加 0.5% 左右.

②对于损失函数进行了重新编写，采用了交叉熵和正则化的方法，把预测值和真实值当做两种概率分布，而交叉熵就体现了两种概率分布的差异；正则化，我将网络中的所有参数相加求和用以描述网络复杂度.

③本来准备编写防止过拟合的方法，但是由于没有测试集的标签，而 early stop 等方法需要测试集进行交叉验证，但是由于缺少测试集的标签，所以放弃了这种方法.

④使用 Adam 进行优化，但是感觉效果不是很好，不知道是不是自己的编写方法存在问题.