作业一：MLQP

超并行机器学习与海量数据挖掘 EI328
2022 年春季工程实践与科技创新课程 IV-J

姓名：李子龙 学号：518070910095 日期：2022 年 3 月 2 日

1 推导

问题 1. Suppose the output of each neuron in a multi- layer perceptron is:

$$x_{kj} = f\left(\sum_{i=1}^{N_{k-1}} (u_{kji}x_{k-1,i}^2 + v_{kji}x_{k-1,i}) + b_{kj}\right) \tag{1}$$

where both u_{kji} and v_{kji} are the weights connecting the i^{th} unit in the layer $k-1$ to the j^{th} unit in the layer k, b_{kj} is the bias of the j^{th} unit in the layer k, N_k is the number of units if $1 \leq k \leq M$ and $f(\cdot)$ is the sigmoidal activation function.

Please derive a back-propagation algorithm for multilayer quadratic perceptron (MLQP) in on-line or sequential mode.

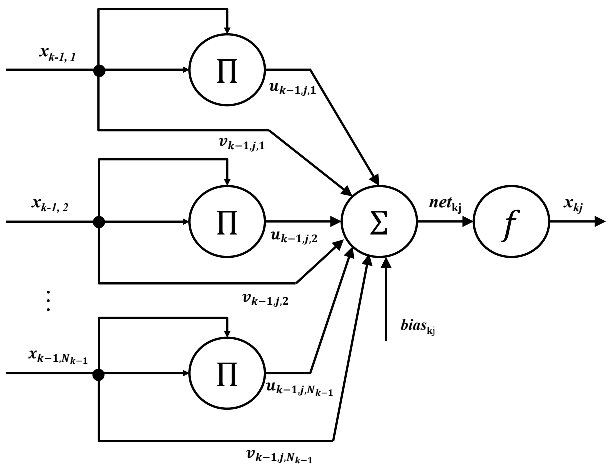

图 1: MLQP

1.1 输出神经元

对于输出神经元信号 x_{Mj} 来说，令 d_{Mj} 为目标值，则对应的误差为

$$e_{Mj} = d_{Mj} - x_{Mj}$$

使用均方误差计算损失函数

$$\mathcal{E} = \sum_{j=1}^{N_M} \frac{1}{2} e_{Mj}^2 \tag{2}$$

则对于 net_{Mj} 的梯度为

$$\begin{aligned}\frac{\partial \mathcal{E}}{\partial net_{Mj}} &= \frac{\partial \mathcal{E}}{\partial e_{Mj}}\frac{\partial e_{Mj}}{\partial x_{Mj}}\frac{\partial x_{Mj}}{\partial net_{Mj}} \\ &= -e_{Mj}f'(net_{Mj})\end{aligned}$$

定义对应的局部梯度（local gradient）为

$$\delta_{Mj} = -\frac{\partial \mathcal{E}}{\partial net_{Mj}} = e_{Mj}f'(net_{Mj}) \tag{3}$$

1.2 隐藏层

对于隐藏神经元上的 $net_{k-1,j}$ 而言，假设后一层反向传播来的局部梯度已知：

$$\delta_{kj} = -\frac{\partial \mathcal{E}}{\partial net_{kj}}$$

则该层的局部梯度为

$$\begin{aligned}\delta_{k-1,i} = -\frac{\partial \mathcal{E}}{\partial net_{k-1,i}} &= -\frac{\partial \mathcal{E}}{\partial x_{k-1,i}}\frac{\partial x_{k-1,i}}{\partial net_{k-1,i}} \\ &= -f'(net_{k-1,i})\sum_{j=1}^{N_j}\frac{\partial \mathcal{E}}{\partial net_{kj}}\frac{\partial net_{kj}}{\partial x_{k-1,i}} \\ &= -f'(net_{k-1,i})\sum_{j=1}^{N_j}\delta_{kj}(2u_{kji}x_{k-1,i} + v_{kji})\end{aligned} \tag{4}$$

1.3 更新权值

而根据公式 (1)，

$$\frac{\partial net_{kj}}{\partial u_{kji}} = x_{k-1,i}^2$$
$$\frac{\partial net_{kj}}{\partial v_{kji}} = x_{k-1,i}$$

设定学习率分别为 η_1, η_2，则对应权重的修正值

$$\Delta u_{kji} = -\eta_1\frac{\partial \mathcal{E}}{\partial u_{kji}} = -\eta_1\frac{\partial \mathcal{E}}{\partial net_{kj}}\frac{\partial net_{kj}}{\partial u_{kji}} = \eta_1\delta_{kj}x_{k-1,i}^2$$
$$\Delta v_{kji} = -\eta_2\frac{\partial \mathcal{E}}{\partial v_{kji}} = -\eta_1\frac{\partial \mathcal{E}}{\partial net_{kj}}\frac{\partial net_{kj}}{\partial v_{kji}} = \eta_2\delta_{kj}x_{k-1,i}$$

为了提高学习速率，引入动量

$$\Delta u_{kji}(t+1) = \alpha_1\Delta u_{kji}(t) + \eta_1\delta_{kj}x_{k-1,i}^2 \tag{5}$$
$$\Delta v_{kji}(t+1) = \alpha_2\Delta v_{kji}(t) + \eta_2\delta_{kj}x_{k-1,i} \tag{6}$$

这里 α_1 和 α_2 分别为其动量常数。这个结果与 [1] 一致。

2 实现

问题 2. Please implement an on-line BP algorithm for MLQP (you can use any programming language), train an MLQP with one hidden layer to classify two spirals problem, and compare the training time and decision boundaries at three different learning rates.

参考文献

[1] LU B L, BAI Y, KITA H, et al. An efficient multilayer quadratic perceptron for pattern classification and function approximation[C/OL]//Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan): volume 2. IEEE, 1993: 1385-1388. DOI: 10.1109/ijcnn.1993.716802.