

模式识别导论第三次作业——编程题

杨登天-202028015926089-微电子研究所

第一类 10 个样本 (三维空间):

(1.58, 2.32, -5.80) (0.67, 1.58, -4.78) (1.04, 1.01, -3.63) (-1.49, 2.18, -3.39)
(-0.41, 1.21, -4.73) (1.39, 3.16, 2.87) (1.20, 1.40, -1.89) (-0.92, 1.44, -3.22)
(0.45, 1.33, -4.38) (-0.76, 0.84, -1.96)

第二类 10 个样本 (三维空间):

(0.21, 0.03, -2.21) (0.37, 0.28, -1.8) (0.18, 1.22, 0.16) (-0.24, 0.93, -1.01)
(-1.18, 0.39, -0.39) (0.74, 0.96, -1.16) (-0.38, 1.94, -0.48) (0.02, 0.72, -0.17)
(0.44, 1.31, -0.14) (0.46, 1.49, 0.68)

第三类 10 个样本 (三维空间):

(-1.54, 1.17, 0.64) (5.41, 3.45, -1.33) (1.55, 0.99, 2.69) (1.86, 3.19, 1.51)
(1.68, 1.79, -0.87) (3.51, -0.22, -1.39) (1.40, -0.44, -0.92) (0.44, 0.83, 1.97)
(0.25, 0.68, -0.99) (0.66, -0.45, 0.08)

1. 请编写两个通用的三层前向神经网络反向传播算法程序, 一个采用批量方式更新权重, 另一个采用单样本方式更新权重。其中, 隐含层结点的激励函数采用双曲正切函数, 输出层的激励函数采用 sigmoid 函数。目标函数采用平方误差准则函数。

解答:

任务一: 推导两个通用的三层前向神经网络反向传播算法;

步骤一、预定义变量: 训练数据输入输出对 $\{x_i^k, t_j^k\}$, x_i^k 表示输入信号, 在本题中 $i \in (1, 2, 3)$, 也就是样本的参数维度 $d = 3$, 偏置为 x_0 , 为了方便, 将偏置设置为 x_0^k ; 在本题中 $j \in (1, 2, 3)$, 也就是样本的分类类别数目为 3, 并计划输出信号的监督 t_j^k 采用独热码 (考虑到 Sigmoid 函数的值域 $[0, 1]$), 如下分别表示

$$\begin{aligned}(t_1^k, t_2^k, t_3^k) &= (1, 0, 0) \quad \text{if } k \in \text{Class } 1 \\(t_1^k, t_2^k, t_3^k) &= (0, 1, 0) \quad \text{if } k \in \text{Class } 2 \\(t_1^k, t_2^k, t_3^k) &= (0, 0, 1) \quad \text{if } k \in \text{Class } 3\end{aligned}$$

其中 k 表示第 k 个样本, 本题中每类样本 $k \in (1, 2 \dots K)$ 。

输出层的结点输出 z_j^k , 隐含层结点的输出 y_h^k , 输入端点数目为 $d + 1$, 输入层结点 i 至隐含层结点 h 的权重为 w_{ih} , 隐含层结点 h 到输出层结点 j 的权重为 w_{hj} 。其中 h 为隐含层结点个数, 并且 $h \in (1, 2 \dots H)$ 。

根据题目要求隐含层结点的激励函数采用双曲正切函数, 因而

$$y_h^k = \tanh(\text{net}_h^k) = \tanh\left(\sum_{i=0}^3 x_i^k w_{ih}\right)$$

根据题目要求输出层的激励函数采用 sigmoid 函数, 因而

$$z_j^k = \text{sigmoid}(\text{net}_j^k) = \text{sigmoid}\left(\sum_{h=1}^H y_h^k w_{hj}\right)$$

根据题目要求目标函数采用平方误差准则函数, 因而

$$J(w) = \frac{1}{2} \sum_{j=1}^3 (t_j^k - z_j^k)^2$$

对于所有的样本而言，均希望目标函数趋近于 0，即

$$J(w) \rightarrow 0$$

步骤二、更新隐含层到输出层的权重：总体目标为希望通过监督信号和输出信号的误差关系反馈到隐含层到输出层的权重上。

$$\Delta w_{hj} = -\eta \frac{\partial J(w)}{\partial w_{hj}} = -\eta \sum_k \frac{\partial J(w)}{\partial net_j^k} \frac{\partial net_j^k}{\partial w_{hj}} = \eta \sum_k [(t_j^k - z_j^k) \times sigmoid'(\sum_{h=1}^H y_h^k w_{hj}) \times y_h^k]$$

由于

$$sigmoid'(net_j^k) = sigmoid(net_j^k) \times [1 - sigmoid(net_j^k)] = z_j^k \times (1 - z_j^k)$$

所以

$$\Delta w_{hj} = \eta \sum_k [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k]$$

其中 η 为更新步长。

那么对于 K 个样本而言，

$$w_{hj}(t+1) = w_{hj}(t) + \Delta w_{hj} = w_{hj}(t) + \eta \sum_k [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k]$$

然而对于单个样本而言，此处以第 k 个样本为例

$$w_{hj}(t+1) = w_{hj}(t) + \eta(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k$$

根据题目要求需要分为单样本权重更新和批量样本权重更新。其中单样本权重又可以分为随机单样本权重更新和遍历单样本权重更新。

1) 批量样本权重更新

这里假定所有样本，即 30 个样本作为一个批次对权重进行更新。

$$w_{hj}(t+1) = w_{hj}(t) + \eta \sum_k [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k]$$

也就是 $K = 30$

2) 随机单样本权重更新

这里假定所有样本混合，即 30 个样本作为抽样集合，每经过一个样本更新一次权重

$$w_{hj}(t+1) = w_{hj}(t) + \eta(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k$$

此处以抽样到第 k 个样本为例。

3) 遍历单样本权重更新

这里假定所有样本混合，即 30 个样本作为遍历集合，每经过一个样本更新一次权重

$$w_{hj}(t+1) = w_{hj}(t) + \eta(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k$$

此处以进行到第 k 个样本为例。

步骤三、更新输入层到隐含层的权重：总体目标为希望通过监督信号和输出信号的误差关系反馈到输入层到隐含层的权重上。

$$\begin{aligned} \Delta w_{ih} &= -\eta \frac{\partial J(w)}{\partial w_{ih}} = -\eta \sum_{k,j} \frac{\partial J(w)}{\partial z_j^k} \frac{\partial z_j^k}{\partial w_{ih}} = -\eta \sum_{k,j} \frac{\partial J(w)}{\partial z_j^k} \frac{\partial z_j^k}{\partial net_j^k} \frac{\partial net_j^k}{\partial w_{ih}} \\ &= -\eta \sum_{k,j} \frac{\partial J(w)}{\partial z_j^k} \frac{\partial z_j^k}{\partial net_j^k} \frac{\partial net_j^k}{\partial y_h^k} \frac{\partial y_h^k}{\partial w_{ih}} = -\eta \sum_{k,j} \frac{\partial J(w)}{\partial z_j^k} \frac{\partial z_j^k}{\partial net_j^k} \frac{\partial net_j^k}{\partial y_h^k} \frac{\partial y_h^k}{\partial net_h^k} \frac{\partial net_h^k}{\partial w_{ih}} \\ &= \eta \sum_{k,j} [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times tanh'(net_h^k) \times x_i^k] \end{aligned}$$

由于

$$y_h^k = \tanh(\text{net}_h^k) = \frac{e^{2 \times \text{net}_h^k} - 1}{e^{2 \times \text{net}_h^k} + 1}$$

$$\tanh'(\text{net}_h^k) = \frac{4}{e^{2 \times \text{net}_h^k} + e^{-2 \times \text{net}_h^k} + 2} = \frac{4}{\frac{1 + y_h^k}{1 - y_h^k} + \frac{1 - y_h^k}{1 + y_h^k} + 2}$$

$$= 4 \times \frac{1 - (y_h^k)^2}{2 + 2(y_h^k)^2 + 2 - 2(y_h^k)^2} = 1 - (y_h^k)^2$$

所以

$$\Delta w_{ih} = \eta \sum_{k,j} [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k]$$

其中 η 为更新步长。

那么对于 K 个样本而言，

$$w_{ih}(t+1) = w_{ih}(t) + \Delta w_{ih}$$

$$= w_{ih}(t) + \eta \sum_{k,j} [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k]$$

然而对于单个样本而言，此处以第 k 个样本为例

$$w_{ih}(t+1) = w_{ih}(t) + \eta \sum_j [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k]$$

根据题目要求需要分为单样本权重更新和批量样本权重更新。其中单样本权重又可以分为随机单样本权重更新和遍历单样本权重更新。

1) 批量样本权重更新

这里假定所有样本，即 30 个样本作为一个批次对权重进行更新。

$$w_{ih}(t+1) = w_{ih}(t) + \eta \sum_{k,j} [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k]$$

也就是 $K = 30$

2) 随机单样本权重更新

这里假定所有样本混合，即 30 个样本作为抽样集合，每经过一个样本更新一次权重

$$w_{ih}(t+1) = w_{ih}(t) + \eta \sum_j [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k]$$

此处以抽样到第 k 个样本为例。

3) 遍历单样本权重更新

这里假定所有样本混合，即 30 个样本作为遍历集合，每经过一个样本更新一次权重

$$w_{ih}(t+1) = w_{ih}(t) + \eta \sum_j [(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k]$$

此处以进行到第 k 个样本为例。

步骤四、不断以抽样或遍历的形式训练，重复步骤二、步骤三，直至满足如下条件

$$||\nabla J(w)|| < \theta$$

其中 θ 为给定的正实数，为允许的上限。如此得到训练成功的权重网络。

$$||\nabla J(w)|| = \text{sqr}t\left(\sum_h \sum_i \left(\frac{\partial J}{\partial w_{ih}}\right)^2 + \sum_j \sum_h \left(\frac{\partial J}{\partial w_{hj}}\right)^2\right)$$

其含义就是所有权重的增量在无 η 放缩下的矢量模。

步骤五、用该网络进行预测或者分类。

任务二：编辑两个通用训练网络算法的伪代码；根据前述的步骤依次编辑两个算法的伪代码

1) 批量样本权重更新

```

begin initialize:  $\theta$ ,  $w_{ih}$ ,  $w_{hj}$ ,  $H$ ,  $K$ ,  $iteration = 0$ 
do  $iteration \leftarrow iteration + 1$ 
  initialize:  $k = 0$ ,  $\Delta w_{ih} = 0$ ,  $\Delta w_{hj} = 0$ 
  do  $k \leftarrow k + 1$ 
     $\Delta w_{hj} \leftarrow \Delta w_{hj} + \eta(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k$ 
     $\Delta w_{ih} \leftarrow \Delta w_{ih} + \eta \sum_j (t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k$ 
  until  $k == K$ 
     $w_{hj} \leftarrow w_{hj} + \Delta w_{hj}$ 
     $w_{ih} \leftarrow w_{ih} + \Delta w_{ih}$ 
  until  $||\nabla J(w)|| < \theta$ 
return  $w$ 

```

2) 随机单样本权重更新

```

begin initialize:  $\theta$ ,  $w_{ih}$ ,  $w_{hj}$ ,  $H$ ,  $K$ ,  $iteration = 0$ 
do  $iteration \leftarrow iteration + 1$ 
   $x^k$  is selected randomly,  $k \in \{1, 2, 3 \dots K\}$ 
   $\Delta w_{hj} \leftarrow \eta(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k$ 
   $\Delta w_{ih} \leftarrow \eta \sum_j (t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k$ 
   $w_{hj} \leftarrow w_{hj} + \Delta w_{hj}$ 
   $w_{ih} \leftarrow w_{ih} + \Delta w_{ih}$ 
  until  $||\nabla J(w)|| < \theta$ 
return  $w$ 

```

3) 遍历单样本权重更新

```

begin initialize:  $\theta$ ,  $w_{ih}$ ,  $w_{hj}$ ,  $H$ ,  $K$ ,  $iteration = 0$ 
do  $iteration \leftarrow iteration + 1$ 
  initialize:  $k = 0$ 
  do  $k \leftarrow k + 1$ 
     $\Delta w_{hj} \leftarrow \eta(t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times y_h^k$ 
     $\Delta w_{ih} \leftarrow \eta \sum_j (t_j^k - z_j^k) \times z_j^k \times (1 - z_j^k) \times w_{hj} \times (1 - (y_h^k)^2) \times x_i^k$ 
     $w_{hj} \leftarrow w_{hj} + \Delta w_{hj}$ 
     $w_{ih} \leftarrow w_{ih} + \Delta w_{ih}$ 
  until  $k = K$ 
  until  $||\nabla J(w)|| < \theta$ 
return  $w$ 

```

解答完毕

2. 请利用上面的数据验证你写的程序，分析如下几点：

- (a) 隐含层不同结点数目对训练精度的影响；
- (b) 观察不同的梯度更新步长对训练的影响，并给出一些描述或解释；
- (c) 在网络结构固定的情况下，绘制出目标函数值随着迭代步数增加的变化曲线。

解答：

训练精度在本题中被如下定义：训练样本在网络中被分类的正确率。但需要考虑的是，训练精度在本题中，对于遍历式批量更新权重网络而言无问题；对于遍历式单样本更新权重网络和随机式单样本更新权重网络而言，则需要后续所有样本测试该网络才可以作为训练精度的依据。此外，训练精度和测试精度有区别！

一些声明：

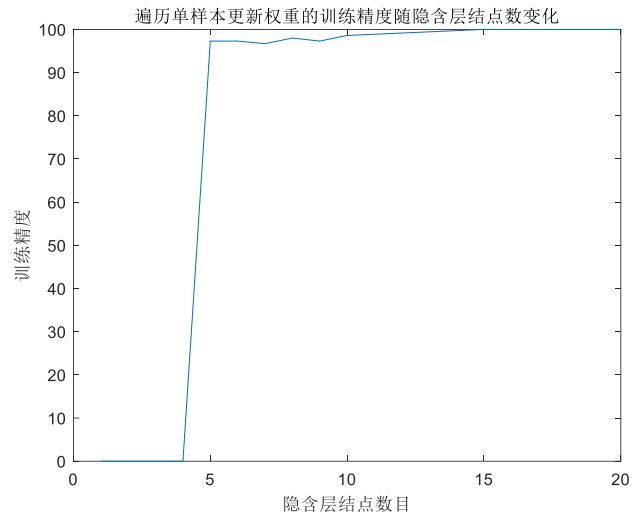
- 1、本题计划参照控制变量法回答所有问题，但权重网络的初始化无法控制每次相同。
- 2、本题所有样本都用于训练网络，除非在单样本更新权重网络模式下求解训练精度时，所有样本测试作为训练精度。
- 3、所有程序的权重网络采用(-1,1)的均匀随机生成，此外规定偏置权重为 1。
- 4、本题中样本按照独热码分类，比如(100)表示 1 类，(010)表示 2 类，(001)表示 3 类。
- 5、遍历式批量更新权重网络中以每次 30 个样本作为一次循环，并以 30 个样本输出信号对权重梯度的模作为控制 while 循环的判断语句主体。
- 6、遍历式单样本更新权重网络中采取顺序样本集。单样本选取方法则通过循环次数对 30 求余数(余数为 0 则样本编号为 30)。权重网络训练完成后对所有样本加以测试，得到的测试分类结果作为训练精度。
- 7、随机式单样本更新权重网络中采取属顺序样本集。单样本选取方法为均匀随机生成 (0,30)的随机数后由 ceil 函数向上取整。权重网络训练完成后对所有样本加以测试，得到的测试分类结果作为训练精度。
- 8、所有程序未调用任何 MATLAB 内置神经网络函数和 nntool 工具箱。

任务一：隐含层不同结点数目对训练精度的影响；

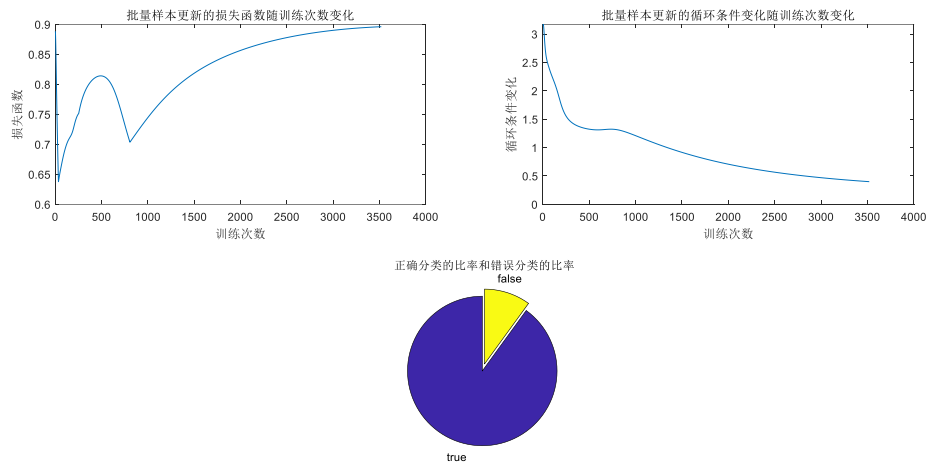
对于遍历式批量更新权重网络而言，所有误差上限设置为 0.4，学习步长为 0.01。考虑到权重产生的随机性，对于所有隐含层结点数目进行多次测试，以排除偶然性。其训练精度的结果如下(以下运行时间超过 5min 的认为无法训练，且认为其训练精度为 0)：

隐含层结点数目	1	2	3	4	5	6
Test1	0	0	0	0	96.7%	100%
Test2	0	0	0	0	100%	100%
Test3	0	0	0	0	100%	96.7%
Test4	0	0	0	0	96.7%	93.3%
Test5	0	0	0	0	93.3%	96.7%
Average	0	0	0	0	97.3%	97.3%
隐含层结点数目	7	8	9	10	15	20
Test1	100%	100%	96.7%	100%	100%	100%
Test2	93.3%	100%	100%	100%	100%	100%
Test3	96.7%	96.7%	93.3%	100%	100%	100%
Test4	93.3%	93.3%	100%	93.3%	100%	100%
Test5	100%	100%	96.7%	100%	100%	100%
Average	96.7%	98%	97.3%	98.6%	100%	100%

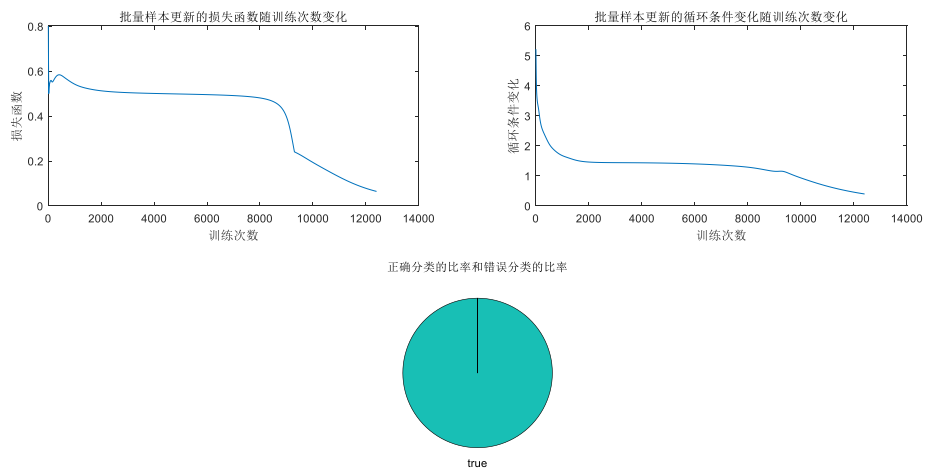
根据上述表格的结果，采用平均值(0 项不计入)，得到如下图



并给出隐含层结点数目为 5 的一个训练结果



给出隐含层结点数目为 8 的一个训练结果

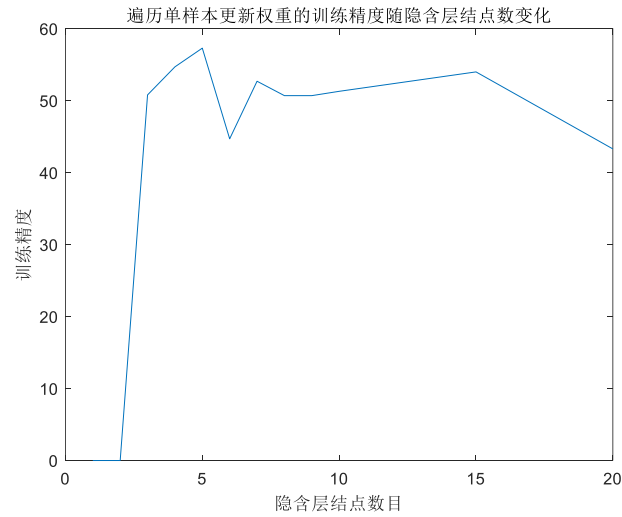


对于遍历式单样本更新权重网络而言，所有误差上限设置为 0.005，学习步长为 0.1，其结果如下：

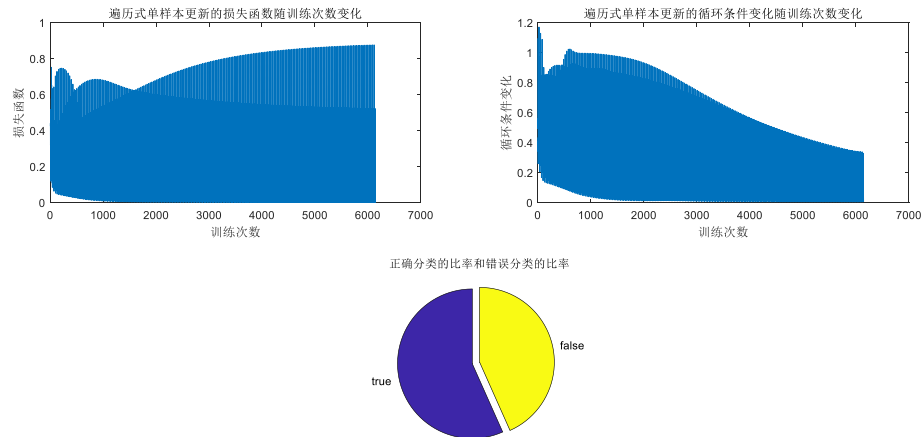
隐含层结点数	1	2	3	4	5	6
--------	---	---	---	---	---	---

Test1	0	0	33.3%	60%	60%	56.7%
Test2	0	0	60%	46.7%	60%	56.7%
Test3	0	0	63.3%	63.3%	60%	36.7%
Test4	0	0	0	60%	63.3%	33.3%
Test5	0	0	46.7%	43.3%	43.3%	40%
Average	0	0	50.8%	54.7%	57.3%	44.7%
隐含层结点数目	7	8	9	10	15	20
Test1	40%	56.7%	56.7%	46.7%	33.3%	60%
Test2	50%	40%	43.3%	63.3%	53.3%	36.7%
Test3	60%	40%	33.3%	63.3%	60%	40%
Test4	60%	60%	60%	40%	63.3%	40%
Test5	53.3%	56.7%	60%	43.3%	60%	40%
Average	52.7%	50.7%	50.7%	51.3%	54.0%	43.3%

根据以上表格的结果，采用平均值(0 项不计入)，得到如下图



给出隐含层结点数为 8 的一种情况

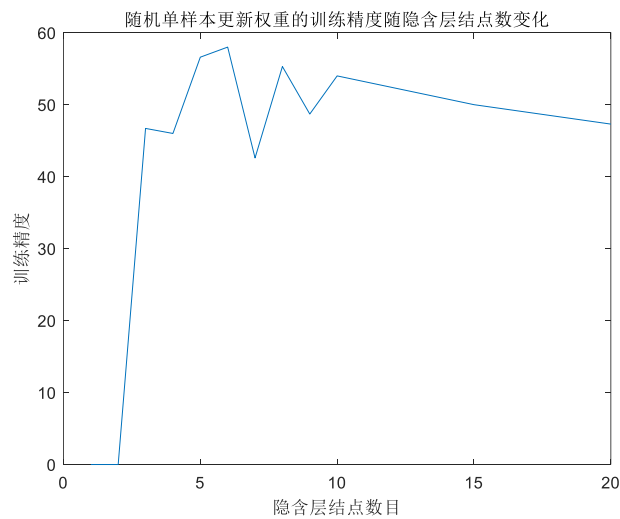


对于随机式单样本更新权重网络而言，所有误差上限设置为 0.005，学习步长为 0.1，其结果如下：

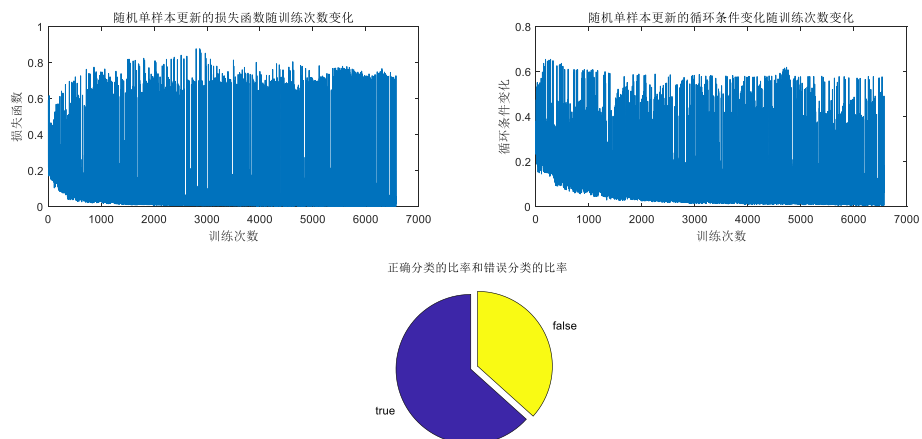
隐含层结点数目	1	2	3	4	5	6
---------	---	---	---	---	---	---

Test1	0	0	40%	63.3%	63.3%	56.7%
Test2	0	0	40%	60%	33.3%	53.3%
Test3	0	0	60%	33.3%	60%	63.3%
Test4	0	0	60%	40%	63.3%	56.7%
Test5	0	0	33.3%	33.3%	63.3%	60%
Average	0	0	46.7%	46.0%	56.6%	58%
隐含层结点数目	7	8	9	10	15	20
Test1	33.3%	60%	56.7%	53.3%	43.3%	33.3%
Test2	40%	36.7%	63.3%	56.7%	60%	53.3%
Test3	43.3%	60%	63.3%	56.7%	50%	50%
Test4	33.3%	63.3%	26.7%	60%	40%	46.7%
Test5	63.3%	56.7%	33.3%	43.3%	56.7%	53.3%
Average	42.6%	55.3%	48.7%	54%	50%	47.3%

根据以上表格的结果，采用平均值(0项不计入)，得到如下图



给出隐含层结点为 4 的一种情况



根据上图和上表得到的结论如下

结论 1: 就训练精度而言，批量样本更新权重网络更好，随机单样本权重更新和顺序单样本权重更新训练效果并不好。

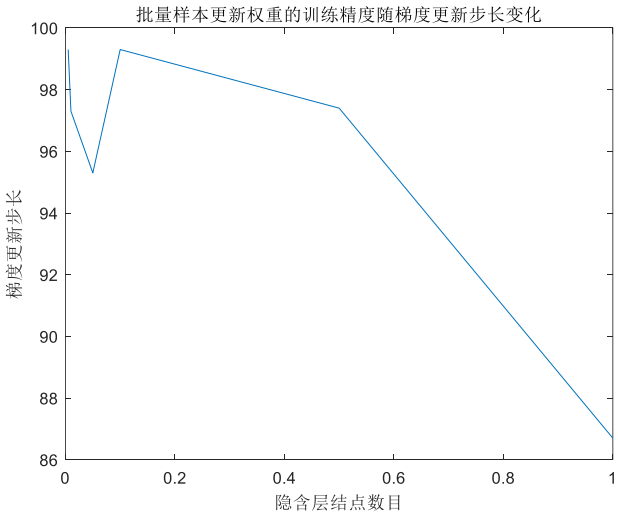
结论 2: 就训练过程看, 批量样本的损失函数变动平缓, 随机单样本和顺序单样本震荡明显。
结论 3: 就隐含层结点数目的增加来看, 批量样本的性能随着结点数目增加而更能准确分类, 但过多的隐含层结点有可能陷入过度训练; 两种单样本更新方式的训练精度则随着结点数目增加起伏较大。

任务二：观察不同梯度更新步长对训练的影响，并给出描述或者解释；

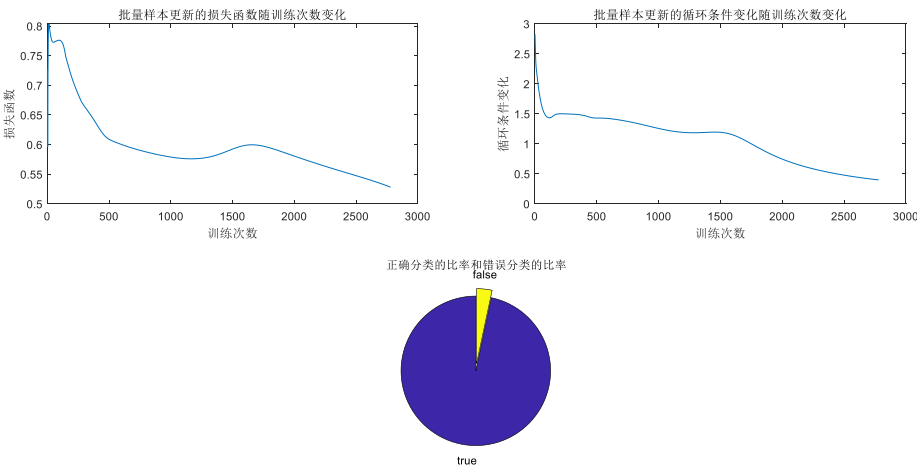
对于遍历式批量更新权重网络而言，所有误差上限设置为 0.4，隐含层结点数目为 6。其结果如下(以下运行时间超过 5min 的认为无法训练，且认为其训练精度为 0)：

梯度更新步长	0.005	0.01	0.05	0.1	0.5	1
Test1	96.7%	100%	96.7%	96.7%	100%	90%
Test2	100%	100%	93.3%	100%	96.7%	93.3%
Test3	100%	96.7%	93.3%	100%	96.7%	90%
Test4	100%	93.3%	93.3%	100%	96.7%	83.3%
Test5	100%	96.7%	100%	100%	96.7%	76.7%
Average	99.3%	97.3%	95.3%	99.3%	97.4%	86.7%

根据以上表格的结果，采用平均值(0 项不计入)，得到如下图



给出梯度更新步长为 0.05 的一种情况

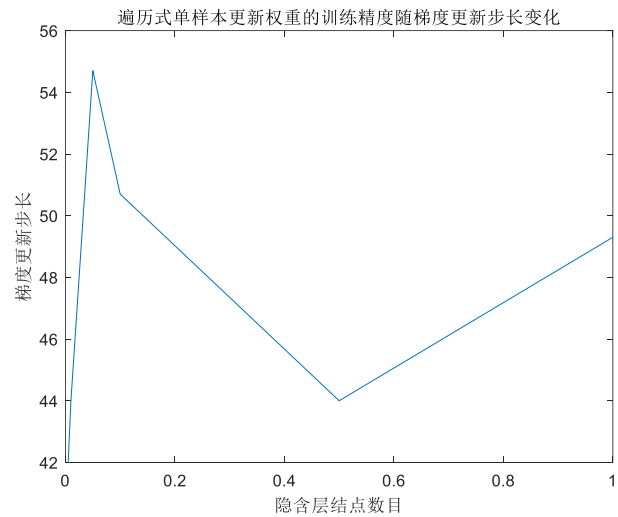


对于遍历式单样本更新权重网络而言，所有误差上限设置为 0.005，隐含层结点数目为 9。

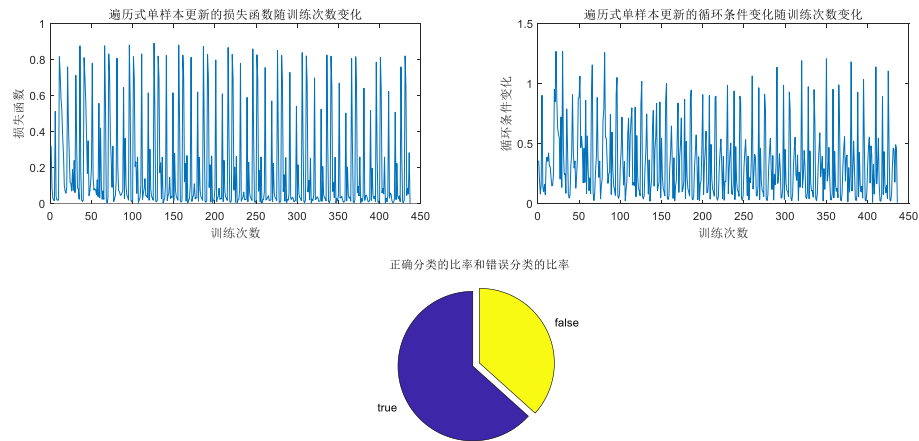
其结果如下：

梯度更新步长	0.005	0.01	0.05	0.1	0.5	1
Test1	36.7%	33.3%	63.3%	56.7%	36.7%	36.7%
Test2	40%	40%	60%	43.3%	36.7%	53.3%
Test3	36.7%	33.3%	33.3%	33.3%	56.7%	53.3%
Test4	33.3%	56.7%	56.7%	60%	43.3%	60%
Test5	63.3%	56.7%	60%	60%	46.7%	43.3%
Average	42%	44%	54.7%	50.7%	44.0%	49.3%

根据以上表格的结果，采用平均值(0 项不计入)，得到如下图



给出梯度更新步长为 0.01 的一种情况

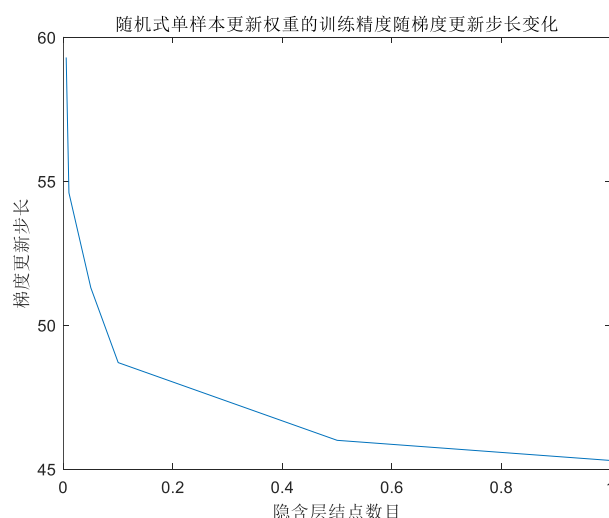


对于随机式单样本更新权重网络而言，所有误差上限设置为 0.005，隐含层结点数目为 9。其结果如下：

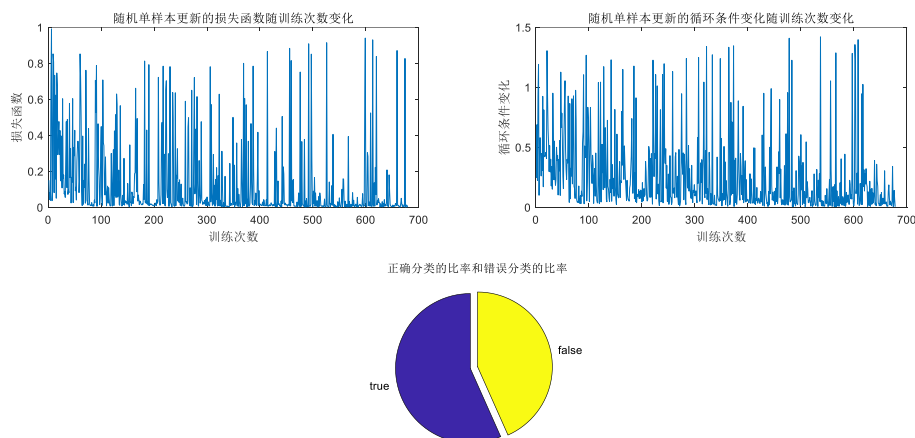
梯度更新步长	0.005	0.01	0.05	0.1	0.5	1
Test1	53.3%	63.3%	40%	56.7%	50%	33.3%
Test2	60%	50%	53.3%	63.3%	33.3%	26.7%
Test3	63.3%	63.3%	43.3%	63.3%	43.3%	56.7%
Test4	56.7%	43.3%	63.3%	26.7%	40%	53.3%
Test5	63.3%	53.3%	56.7%	33.3%	63.3%	56.7%

Average	59.3%	54.6%	51.3%	48.7%	46.0%	45.3%
---------	-------	-------	-------	-------	-------	-------

根据以上表格的结果，采用平均值(0 项不计入)，得到如下图



给出梯度更新步长为 1 的一种情况



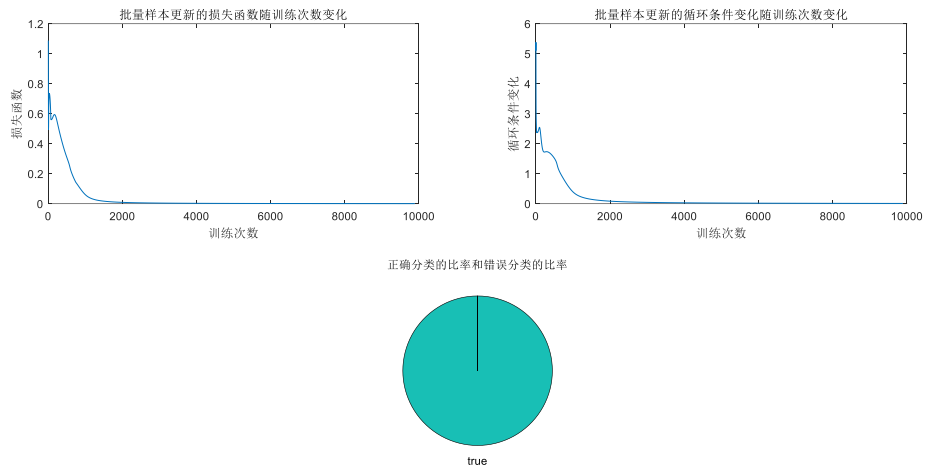
根据上图和上表得到的结论如下

结论：总体上都是随着步长的增大而训练精度由小变大再变小，其应该与步长较小时，权重网络更新速度慢，导致前后两次的误差变化并不明显，从而容易满足误差上限条件而退出循环；当步长较大时，会使得权重网络更速速度快，前后两次误差的变化有明显的起伏，在伏状态容易满足误差上限条件而退出循环。

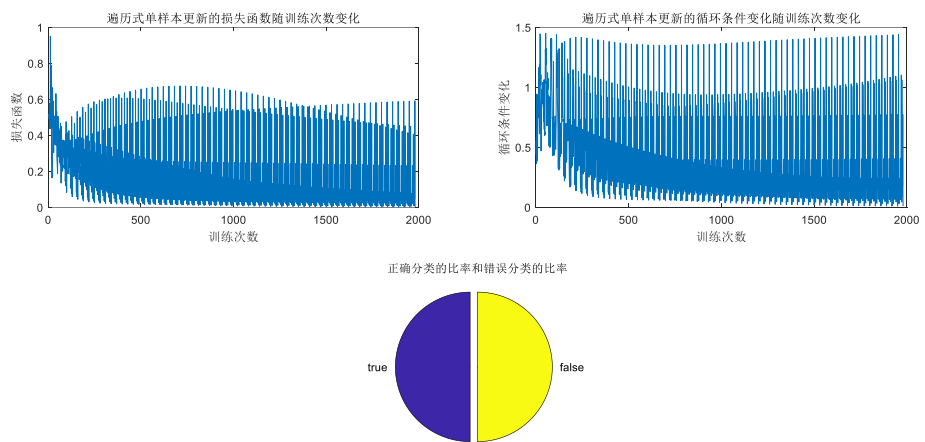
任务三：在网络结构固定的情况下，绘制出目标函数值随着迭代步数增加的变化曲线；

本题中所有误差上限设置为 0.01，隐含层结点数目为 10，学习步长为 0.1

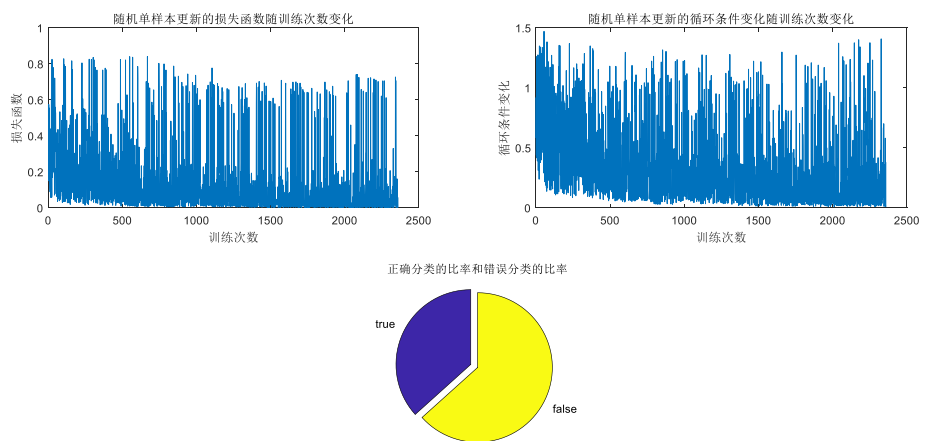
对于遍历式批量更新权重网络而言，其结果如下：



对于遍历式单样本更新权重网络而言，其结果如下：



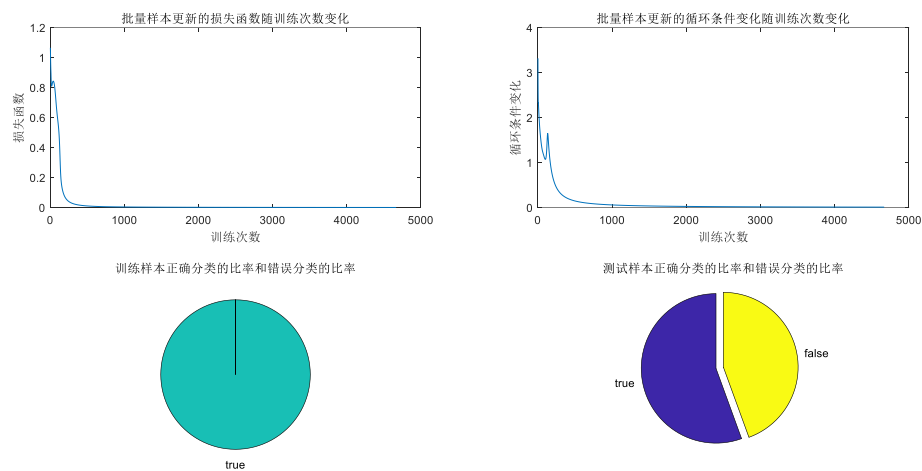
对于随机式单样本更新权重网络而言，其结果如下：



解答完毕

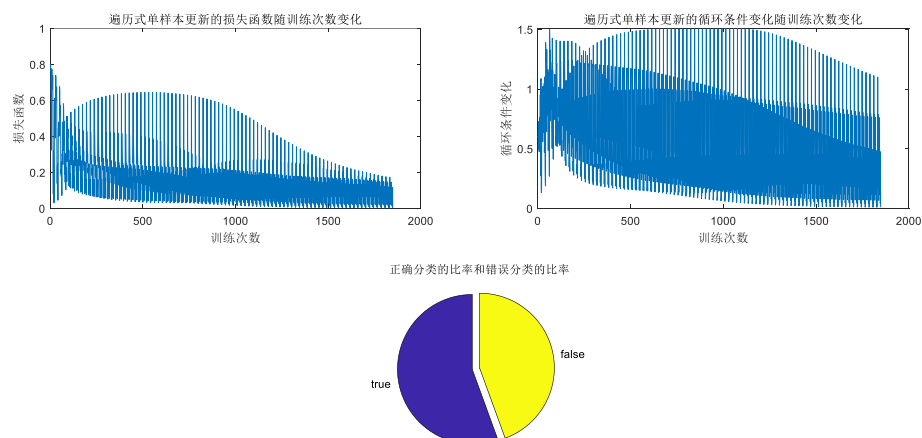
额外做点将样本分为训练集和测试集的工作

所有误差上限设置为 0.01，隐含层结点数目为 10，学习步长为 0.1。对每一类的前 7 个样本作为训练样本集，后 3 个样本作为测试样本集。展开下述三类训练。
对于遍历式批量更新权重网络而言，其结果如下：



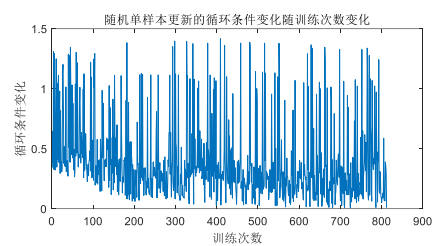
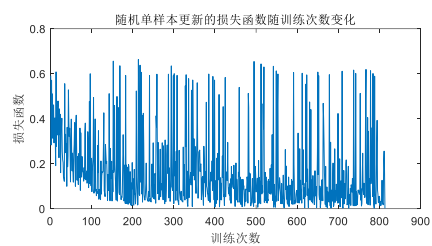
其中训练精度高达 100%，但测试精度只有 53.3%。也就是 9 个测试样本分类正确只有 5 个！

对于遍历式单样本更新权重网络而言，其结果如下：

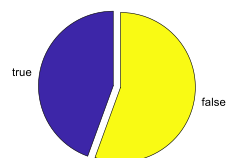


此处无训练精度，测试精度位 53.3%，即 9 个测试样本分类正确只有 5 个！（数据浮动较大，最少可以分类正确 1 个）

对于随机式单样本更新权重网络而言，其结果如下：



正确分类的比率和错误分类的比率



此处无训练精度，测试精度位 44.4%，即 9 个测试样本分类正确只有 4 个！（数据浮动较大，最少可以分类正确 1 个）