中国科学院大学

University of Chinese Academy of Sciences

计算智能：神经网络大作业

BP神经网络对鸢尾花数据集分类

日期：2020.10.03　学生：陈帅华　任课老师：程龙

目 录

CONTENTS

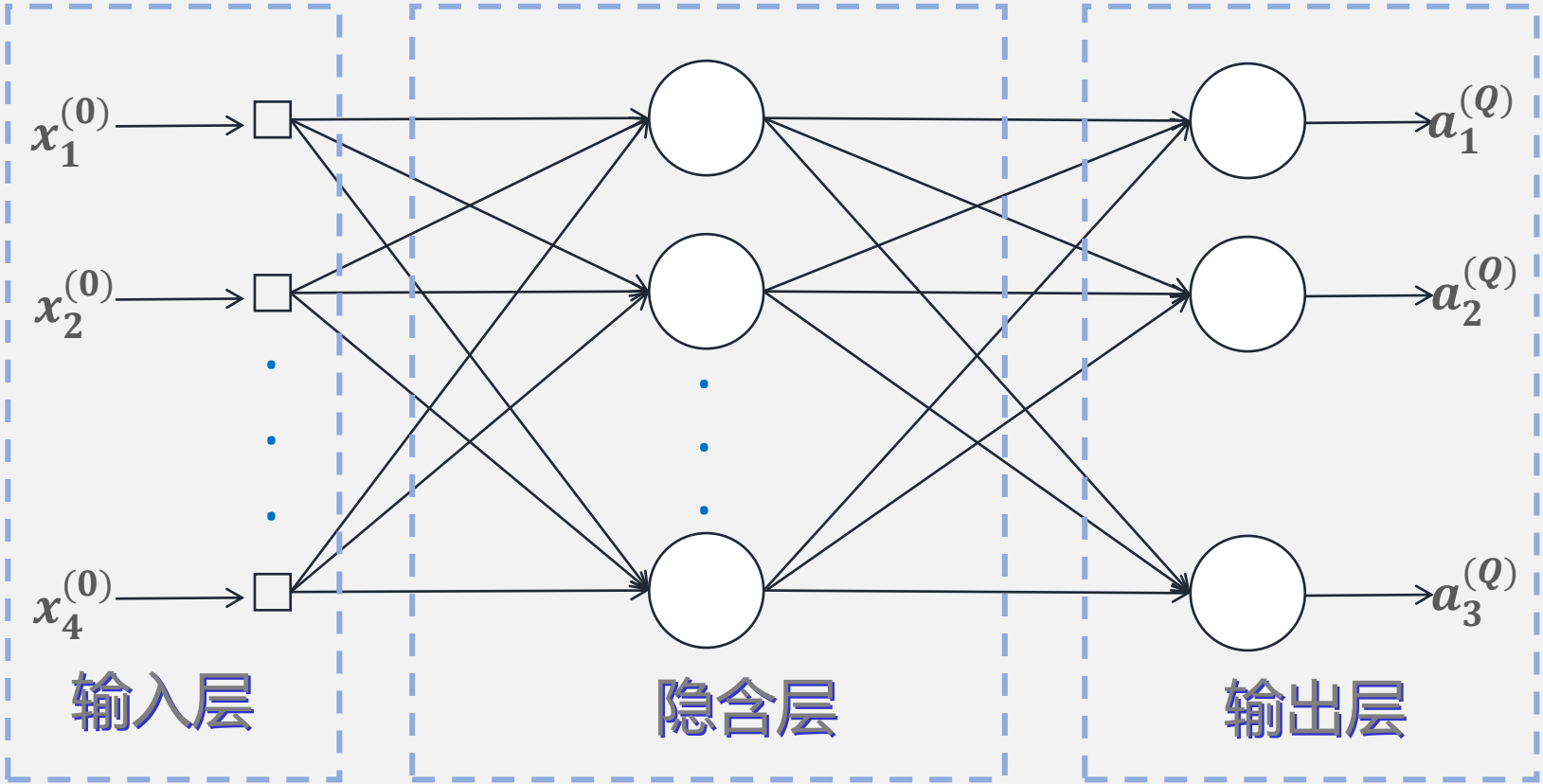
权重更新过程

Weights Update Process

中国科学院大学

University of Chinese Academy of Sciences

BP神经网络介绍



BP神经网络介绍

N个样本总的损失函数：

$$L = -\sum_{i=1}^{N} y_i \ln a_i$$

其中的一项表示一个节点的

N 样本个数数目

y_i 第i个样本的标签(0/1)

a_i 输出层第i个样本的激活值

BP神经网络介绍

- 本作业中使用3层的BP神经网络（第0层为输入层，第2层为输出层）
- 输入层有4个神经元，输出层有3个神经元，隐含层有n_h个神经元
- 考虑隐含层到输出层，隐含层第k个神经元到输出层的第i个神经元的连接权重为w_{ik}，

$$\frac{\partial L}{\partial z_i} = \sum_j \left(\frac{\partial L}{\partial a_j}\frac{\partial a_j}{\partial z_i}\right)$$

BP神经网络介绍

其中：

$$\frac{\partial L}{\partial a_j}=\frac{\partial(-y_j \ln a_j)}{\partial a_j}=-y_j\frac{1}{a_j}$$

$$\frac{\partial a_j}{\partial z_i}=\frac{\partial(\frac{e^{zi}}{\sum_k e^{zk}})}{\partial z_i}=\frac{\sum_k e^{zk}e^{zi}-(e^{zi})^2}{(\sum_k e^{zk})^2}=(\frac{e^{zi}}{\sum_k e^{zk}})(1-\frac{e^{zi}}{\sum_k e^{zk}})=a_i(1-a_i)$$ i=j时

$$\frac{\partial a_j}{\partial z_i}=\frac{\partial(\frac{e^{zj}}{\sum_k e^{zk}})}{\partial z_i}=\frac{-e^{zj}e^{zi}}{(\sum_k e^{zk})^2}=-a_i a_j$$ i≠j时

BP神经网络介绍

$$\frac{\partial L}{\partial z_i}=\sum_j\left(\frac{\partial L_j}{\partial a_j}\frac{\partial a_j}{\partial z_i}\right)=\sum_{i\neq j}\frac{\partial L_j}{\partial a_j}\frac{\partial a_j}{\partial z_i}+\sum_{i=j}\frac{\partial L_j}{\partial a_j}\frac{\partial a_j}{\partial z_i}$$

$$=\sum_{i\neq j}-y_j\frac{1}{a_j}(-a_i a_j)-y_j\frac{1}{a_j}(a_i(1-a_i))$$

$$=\sum_{i\neq j}a_i y_j+(a_i y_i-y_i)$$

$$=a_i\sum_j y_j-y_i$$

对于分类问题，给定的结果y_j 只有一个类别1，其余类别都是0,故：

$$\frac{\partial L}{\partial z_i}=a_i-y_i$$

$$\frac{\partial L}{\partial w_{ik}}=\sum_k x_k(a_i-y_i)$$

SGD中隐含层节点数对分类精度影响

Number of hidden layer nodes' impacts on classification

中国科学院大学

University of Chinese Academy of Sciences

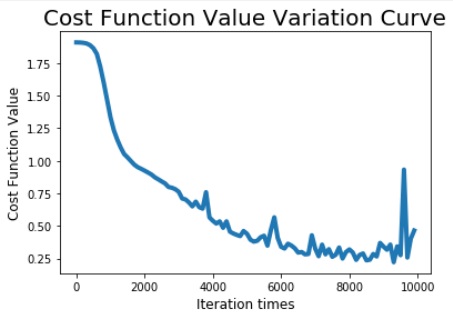
隐含层节点数目的影响

隐含层的选取的经验公式：

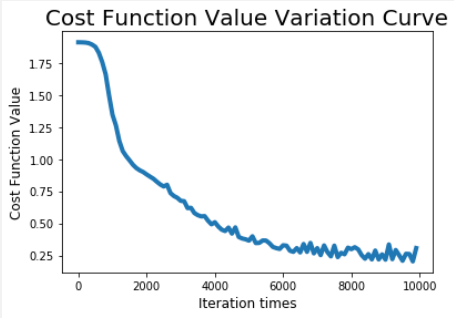
$$h = \sqrt{m+n} + a$$

- h为隐含层节点数目，m为输入层节点数目，n为输出层节点数目
- a为1~10之间的调节常数，故h可以取4~14的正整数。

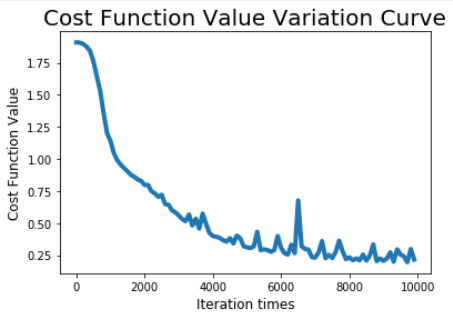
隐含层节点数目的影响



n_h=4



n_h=8



n_h=14

图2·1 损失函数随隐含层节点数目变化曲线

隐含层节点数目的影响

总结2·1

- 隐含层节点个数取不同值时，基本上都是在第8000次迭代时，损失函数收敛到一个数值(0.25左右)
- 当隐含层节点个数由4增加到14的过程中，在迭代次数靠后时，损失函数的变化曲线由粗糙变得光滑继而再变得粗糙。

隐含层节点数目的影响

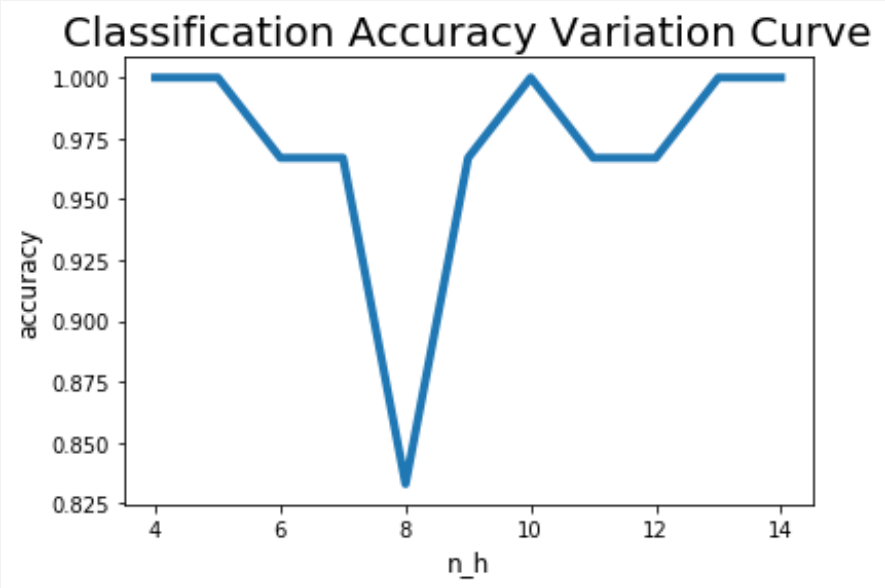


图2·2 隐含层节点数目对分类精度的影响

隐含层节点数目的影响

总结2·2

- 随着隐含层节点个数的增加(4→14)，分类精度会先减少再增加，并且在节点个数等于8时，分类精度取得最小值。
- 当隐含层节点数目过少时，神经网络将不容易建立合适的判断界，而当节点数目过大时会导致训练时间长，模型出现过拟合。
- 结合图2·2，本作业中选择节点数为10，然后进行下面的研究分析。

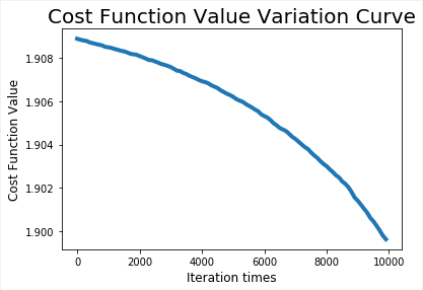
SGD中不同梯度更新步长对训练的影响

Impacts of different gradient update step size on training

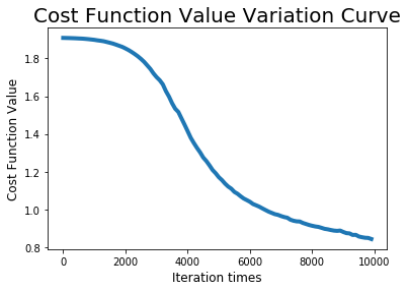
中国科学院大学

University of Chinese Academy of Sciences

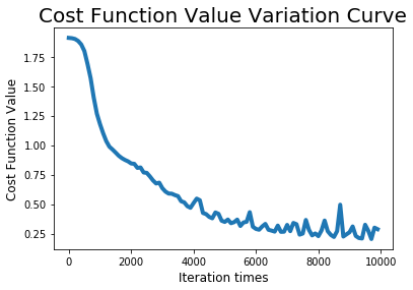
不同梯度更新步长的影响



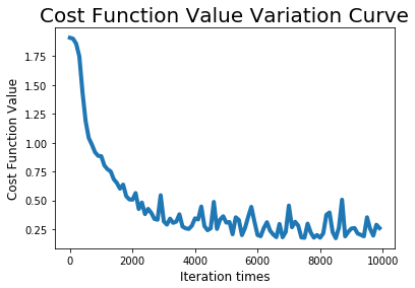
learning rate=0.0001



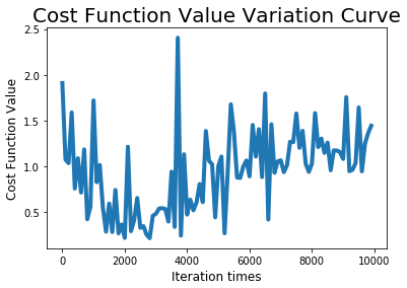
learning rate=0.001



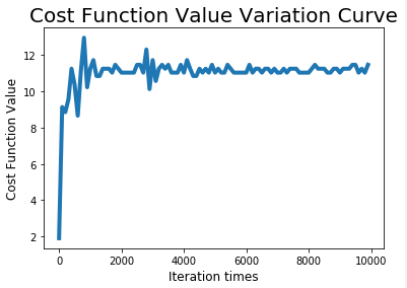
learning rate=0.005



learning rate=0.01



learning rate=0.1



learning rate=0.8

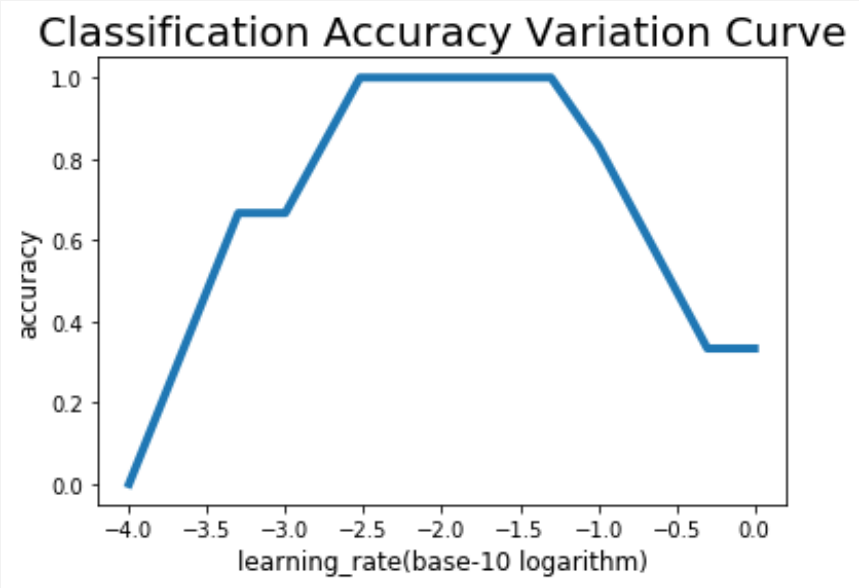
图3·1 损失函数随不同梯度更新步长变化曲线

不同梯度更新步长的影响

总结3·1

- 在梯度更新步长的增加(从0.0001增加到0.8)过程中，损失函数的下降速度加快，同时损失函数最终收敛的数值也越来越小；
- 但当梯度更新步长过大时，损失函数曲线会**不停震荡**甚至一直**保持在一个较大的数值不变**；

不同梯度更新步长的影响



注：横坐标为对梯度更新步长取了以10为底的对数

图3·2 不同梯度更新步长对分类精度的影响

不同梯度更新步长的影响

总结3·2

- 当梯度更新步长过小时，训练过程会变得非常慢，同时最终得到的分类精度也会比较低，误差比较大；
- 当梯度更新步长过大时，不但不会降低误差，反而会增加误差，使得分类精度降低；
- 只有当选择合适的梯度更新步长时，才既能有较快的训练速度，同时又有较高的分类精度。

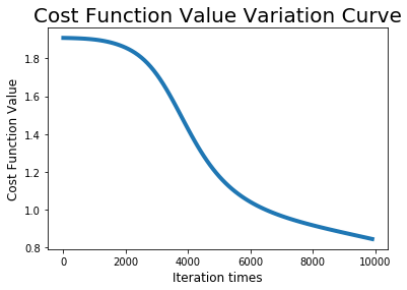
SGD与Momentum的对比

Comparison between SGD and Momentum

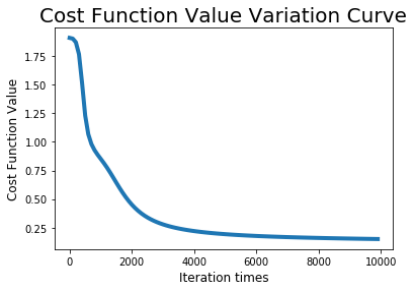
中国科学院大学

University of Chinese Academy of Sciences

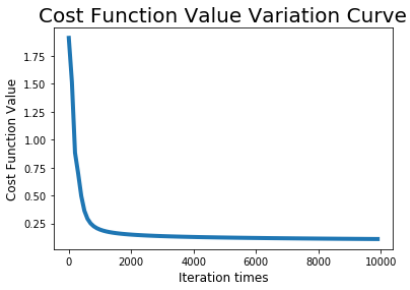
SGD与Momentum的对比



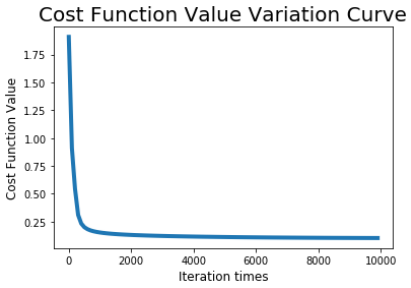
learning rate=0.0001



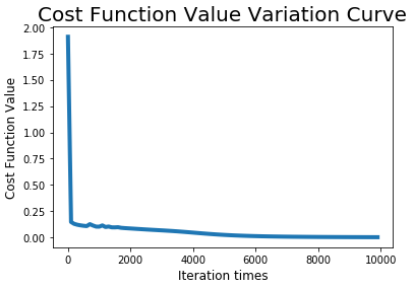
learning rate=0.001



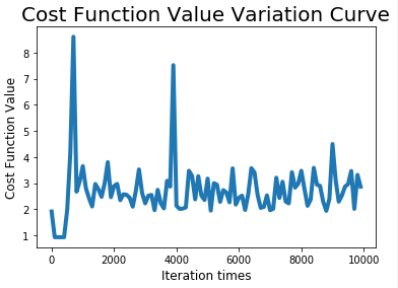
learning rate=0.005



learning rate=0.01



learning rate=0.1



learning rate=0.8

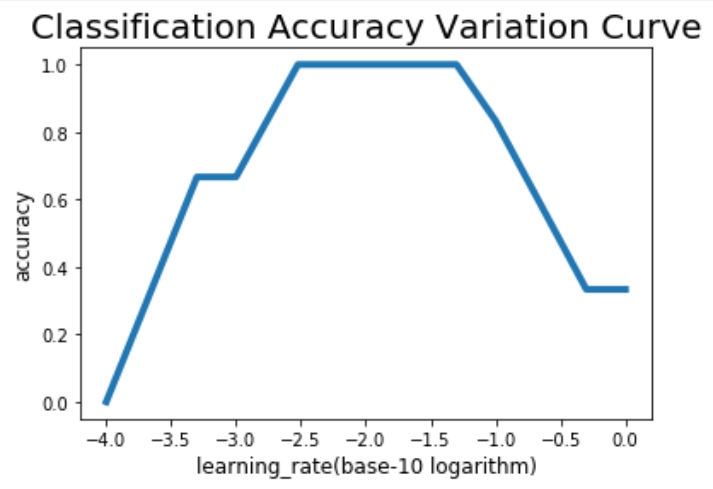
图4·1 损失函数随不同梯度更新步长变化曲线

SGD与Momentum的对比

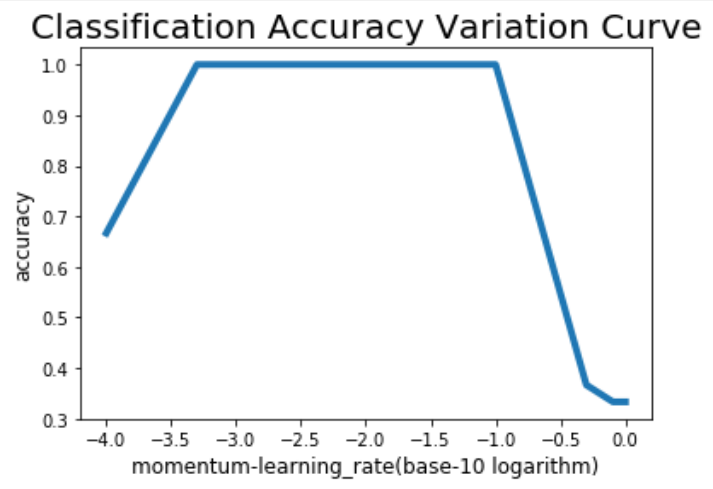
总结4·1

- 将图4·1和图3·1对比之后，可以发现在不同的梯度更新步长下，动量法的收敛速度要**远快于**随机 梯度下降法；
- 同时模型收敛后由动量法得到的损失函数值也要低于由随机梯度下降法得到的函数值；
- 当梯度更新步长较大时，动量法仍能使得模型收敛，而此时随机梯度下降法已经震荡了。

SGD与Momentum的对比



随机梯度下降法



动量法

图4·2 不同梯度更新步长对分类精度的影响

SGD与Momentum的对比

总结4·2

- 通过图4·2可以看出，在**相同**的梯度更新步长下，由动量法得到的分类精度要**高于**由随机梯度下降法得到的分类精度，因而动量法的**整体分类精度高于**随机梯度下降法。

中国科学院大学

University of Chinese Academy of Sciences

请老师批评指正

Hoping for criticism and suggestions

日期：2020.10.06　学生：陈帅华　任课老师：程龙