Lecture5习题作业

1，有人说当批量大小为1时基于随机梯度下降法（Stochastic Gradient Descent，SGD）的逻辑斯蒂回归（Logistic Regression）算法可以被看作为“软性”的感知器算法（PLA），你认同这个说法吗？请给出你的理由。

2，在Logistic regression中当标签y={+1,-1}时常用交叉熵作为损失函数： ，请推导出该函数的梯度表达式。

3， 为什么在Logistic Regression中不用作为损失函数，这里假设是*Sigmoid*函数，标签y={+1,-1}。

Lecture5编程作业

1，编程实现Logistic regression算法。

2，（a）产生两个都具有200个二维向量的数据集和。数据集的样本来自均值向量协方差矩阵的正态分布，属于“+1”类，数据集的样本来自均值向量、协方差矩阵的正态分布，属于“-1”类，其中****是一个2\*2的单位矩阵。产生的数据中80%用于训练，20%用于测试。

（b）在训练集上利用Logistic regression算法得到分类面。

（c）利用得到的分类面对测试集样本进行分类，并给出每个样本属于该类别的概率值。

（d）画出数据集和分类面。

（e）画出损失函数随epoch增加的变化曲线。

（f）改变算法中的各类超参数、样本数量、样本分布等，对于梯度下降法还要改变不同的学习率以及不同的batch size和不同epoch次数，讨论实验结果。