Lecture3习题作业

1，假设训练样本集为，使用线性回归算法（Linear Regression Algorithm），通过广义逆来求解，并设计这两类的分类函数，讨论结果。（可通过编程计算得到广义逆的结果）

2，根据向量或矩阵的计算性质，证明：

3，总结梯度下降法、随机梯度下降法、Adagrad、RMSProp、动量法（Momentum）和Adam等方法权系数更新表达式。

Lecture3编程作业

1，分别编写一个用广义逆和梯度下降法来求最小误差平方和最佳解的算法

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）在上述数据集上分别第1题的两个算法，利用产生的训练样本集得到分类面，算法中用到的各类超参数自定。

（c）分别在训练集和测试集上统计分类正确率。

（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。

3，重复第2题的内容，但数据集和数据集的均值向量分别改为和，其他不变。

4，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。

5，单变量函数为，分别用梯度下降法、随机梯度下降法、Adagrad、RMSProp、动量法（Momentum）和Adam共6种方法，编写程序画图呈现从初始值为-4、迭代10次时及的每次变化情况，这里对所有算法学习率（或初始学习率）均为0.4，为防止分母为0时给的最小量为1e-6，RMSProp算法的=0.9，动量法的=0.9，Adam的beta1=0.9，beta2=0.999，观察不同算法的变化情况体会各自的差异。如果迭代50次，并将Adam的beta1改成0.99，其他参数不变，观察不同算法的变化结果。尝试调整上述算法的各种参数，体会上述不同方法的特点。