

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**   **张缘**

**学 号 201530613696**

**邮 箱 zyatl@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年12月8日**

## 1. 实验题目: 线性回归、线性分类与梯度下降

## 2. 实验时间：2017年 12 月 8 日

## 3. 报告人:张缘

## 4. 实验目的:

1. 进一步理解线性回归和梯度下降的原理。
2. 在小规模数据集上实践。
3. 体会优化和调参的过程

## 5. 数据集以及数据分析：

线性回归使用的是LIBSVM Data中的Housing数据，包含506个样本，每个样本有13个属性,并切分为训练集，验证集。   
线性分类使用的是LIBSVM Data中的australian数据，包含690个样本，每个样本有14 个属性，并切分为训练集，验证集。

## 6. 实验步骤:

1)线性回归和梯度下降

1. 读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。
2. 将数据集切分为训练集和验证集，本次实验不切分测试集。使用train\_test\_split函数切分数据集。
3. 线性模型参数初始化，可以考虑全零初始化，随机初始化或者正态分布初始化。
4. 选择Loss函数及对其求导，过程详见课件ppt。
5. 求得**所有样本**对Loss函数的梯度G。
6. 取梯度G的负方向，记为D。
7. 更新模型参数，W(t)=W(t-1)+nD。n为学习率，是人为调整的超参数。
8. 在训练集上测试并得到Loss函数值L(train)，在验证集上测试并得到Loss函数值L(validation)。
9. 重复步骤5-8若干次，**画出**L(train)**和**L(validation)**随迭代次数的变化图**。

2)线性分类和梯度下降

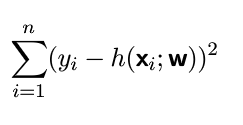
1. 读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。
2. 将数据集切分为训练集和验证集，本次实验不切分测试集。使用train\_test\_split函数切分数据集。
3. 支持向量机模型参数初始化，可以考虑全零初始化，随机初始化或者正态分布初始化。
4. 选择Loss函数及对其求导，过程详见课件ppt。
5. 求得**所有样本**对Loss函数的梯度G。
6. 取梯度G的负方向，记为D。
7. 更新模型参数，W(t)=W(t-1)+nD。n为学习率，是人为调整的超参数。
8. 选择合适的阈值，将计算结果**大于阈值的标记为正类，反之为负类**。在训练集上测试并得到Loss函数值L(train)，在验证集上测试并得到Loss函数值L(validation)。
9. 重复步骤5-8若干次，**画出L(train)和L(validation)随迭代次数的变化图**。

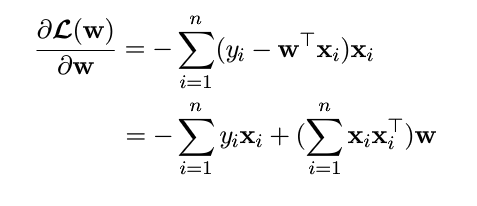
## 7. 代码内容:

## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

## 9. 模型参数的初始化方法:全零初始化

## 10.选择的loss函数及其导数:

L(w)= 



## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

## 评估结果（根据选择的评估方法）：

## 预测结果（最佳结果）：

## loss曲线图：

## 12.实验结果分析:

## 13.对比线性回归和线性分类的异同点：

## 14.实验总结：