

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**   **杨卓东**

**学 号 20153061341**

**邮 箱**

**指导教师**

**提交日期** **2017年 月 日**

## 实验题目

## 线性回归、线性分类与梯度下降

## 实验时间

## 2017年 12 月 2 日

## 报告人

## 杨卓东

## 实验目的

进一步理解线性回归和梯度下降的原理。

在小规模数据集上实践。

体会优化和调参的过程。

## 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。

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。

## 实验步骤

编写python代码，并运行于本机中进行模型训练。

**线性回归和梯度下降**

* 读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。
* 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。
* 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。
* 选择Loss函数及对其求导，过程详见课件ppt。
* 求得所有样本对Loss函数的梯度。
* 取梯度的负方向，记为。
* 更新模型参数，。为学习率，是人为调整的超参数。
* 在训练集上测试并得到Loss函数值，在验证集上测试并得到Loss函数值。
* 重复步骤5-8若干次，画出和随迭代次数的变化图。

**线性分类和梯度下降**

* 读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。
* 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。
* 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。
* 选择Loss函数及对其求导，过程详见课件ppt。
* 求得所有样本对Loss函数的梯度。
* 取梯度的负方向，记为。
* 更新模型参数，。为学习率，是人为调整的超参数。
* 在训练集上测试并得到Loss函数值，在验证集上测试并得到Loss函数值。
* 重复步骤5-8若干次，画出和随迭代次数的变化图。

## 代码内容

见提交仓库内ClassificationExperiment.ipynb, RegressionExperiment.ipynb文件

## 实验参数

**线性回归和梯度下降**

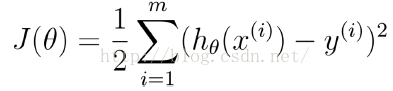
## 选择的评估方法

留出法

## 模型参数的初始化方法:

全零初始化

## 选择的loss函数及其导数:

Loss: 

导数: 

## C:\Users\Curisu\Desktop\Figure_1.png实验结果和曲线图:

## 实验结果分析:

模型在前20~30次迭代中快速收敛，呈现拐点后收敛减速

## 对比线性回归和线性分类的异同点

相同

算法的流程基本相同

使用的都为线性参数进行建模

不同

Loss函数(及导数)的选取

评估方法的不同