



华南理工大学

South China University of Technology

机器学习实验报告

学院：软件学院

专业：软件工程

作者:

王子通

指导教师:

谭明奎

学号:

202030484151

班级:

2020 级 2 班

2022-10-12

线性回归、随机梯度下降

摘要—回归 (regression) 是能为一个或多个自变量与因变量之间关系建模的一类方法。在自然科学和社会科学领域，回归经常用来表示输入和输出之间的关系。在这里通过随机梯度下降与闭式解，得到线性回归的参数预测。

I. 介绍

输入特征为 13，输出预测值为 1。

首先线性回归基于几个简单的假设：

- ◇ 假设自变量 x 和因变量 y 之间的关系是线性的，即 y 可以表示为 x 中元素的加权和，这里通常允许包含观测值的一些噪声
 - ◇ 假设任何噪声都比较正常，如噪声遵循正态分布。
- 通过线性回归的闭式解与随机梯度下降的方式，得到了模型的参数估计值。

最终结果 $loss_train, loss_val$ 展示在 table I 中，参数估计值展示在 table II 中

II. 方法和理论

◇ 解析解：

将偏置 b 合并到参数 w 中，合并方法是在包含所有参数的矩阵中附加一列。预测问题是最小化 $\|y - Xw\|^2$ 。这在损失平面上只有一个临界点，这个临界点对应于整个区域的损失极小点。将损失关于 w 的导数设为 0，得到解析解：

$$w^* = (X^T X)^{-1} X^T y$$

◇ 随机梯度下降：

梯度下降最简单的用法是计算损失函数（数据集中所有样本的损失均值）关于模型参数的导数（在这里也可以称为梯度）。但实际中的执行可能会非常慢：因为在每一次更新参数之前，我们必须遍历整个数据集。因此，我们通常会在每次需要计算更新的时候随机抽取一小批样本，这种变体叫做小批量随机梯度下降。在每次迭代中，我们首先随机抽样一个小批量 B ，它是由固定数量的训练样本组成的。然后，我们计算小批量的平均损失关于模型参数的导数（也可以称为梯度）。最后，我们将梯度乘以一个预先确定的正数 η ，并从当前参数的值中减掉。更新过程如下：

$$(w, b) \leftarrow (w, b) - \frac{\eta}{|B|} \sum_{i \in B} \partial_{(w, b)} l^{(i)}(w, b)$$

III. 实验

A. 数据集

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。输出为单个值。并按照 2: 1 切分册成训练集与验证集。

B. 实现

实验中的所有详细实现：

- ◇ 初始化: `load_svmlight_file()` 函数读取数据，并规格化数据格式。并利用 `train_test_split()` 将数据集切分成训练集与验证集

◇ 处理过程：

- 解析解：直接带入公式计算即可得
- 随机梯度下降：在每次迭代中，读取一小批量训练样本，并通过我们的模型来获得一组预测。计算完损失后，我们开始反向传播，存储每个参数的梯度。最后，我们调用优化算法 `sgd` 来更新模型参数。即不断执行以下循环：

1. 计算梯度：

$$g \leftarrow \partial_{(w, b)} \frac{1}{|B|} \sum_{i \in B} l(x^{(i)}, y^{(i)}, w, b)$$

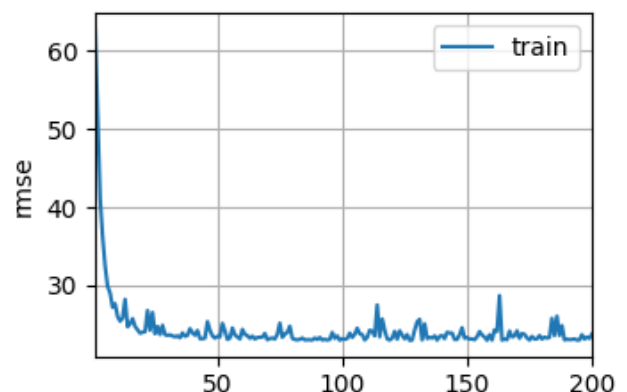
2. 更新参数：

$$(w, b) \leftarrow (w, b) - \eta g$$

◇ 结果：

训练随机梯度损失下降过程如 figure I 所示：

FIGURE I
Model Training Process



闭式解得出训练集与测试集的损失值如 table I

TABLE I
Loss Table of Different Ways

	<i>loss_train</i>	<i>loss_val</i>
analytical solution	22.947321	20.642921
stochastic gradient descent	23.785135	20.711567

参数估计如 table II:

TABLE II
SIMULATION PARAMETERS

	analytical solution	stochastic gradient descent
Bias	10.42259376	
W1	-5.72741864	20.711567
W2	1.89116185	20.711567
W3	0.79399826	20.711567
W4	1.6193348	20.711567
W5	-3.92926638	20.711567
W6	10.1824044	20.711567
W7	-0.62390586	20.711567
W8	-7.8200281	20.711567
W9	2.70081962	20.711567
W10	-2.15189053	20.711567
W11	-4.36499552	20.711567
W12	2.33379236	20.711567
W13	-9.92189944	20.711567

IV. 结论

收获与感悟:

- 使用小批量的随机梯度下降适用于数据量比较大的数据集，所以对于本次实验只有 506 个样本而言，没有必要。
- 在写代码过程中，数据格式着实令人头疼，在编写代码过程中要有良好的习惯去统一格式。