**华 中 科 技 大 学**

**《智能数据处理》**

**实 验 报 告**

**实验名称：波士顿房价预测**

**院(系)名：软件学院 专业：软件工程**

**班级:软工1603 姓名：刘俊傲**

**学号：U201617047**

**提交日期：2019-10-25**

# 实验内容

## 1.1 实验内容

经典的线性回归模型主要用来预测一些存在着线性关系的数据集。回归模型可以理解为：存在一个点集，用一条曲线去拟合它分布的过程。如果拟合曲线是一条直线，则称为线性回归。如果是一条二次曲线，则被称为二次回归。线性回归是回归模型中最简单的一种。 本实验使用PaddlePaddle建立起一个房价预测模型。

## 1.2实验环境

### 1.2.1 硬件信息

### CPU： 2

运存： 8GB

### 1.2.2 环境配置

Windows10

Jupyter Notebook

Python 3.5

PaddlePaddle 1.5

## 1.3实验步骤

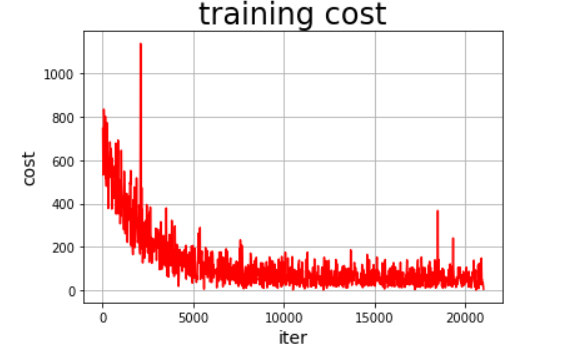
1. **准备数据**：PaddlePaddle内置了uci\_housing数据集，并且提供了读取训练集和测试集的接口，分别为paddle.dataset.uci\_housing.train()和paddle.dataset.uci\_housing.test()，直接调用即可。当使用批量梯度下降时，需要将数据组织成分批形式，使用到train\_reader()和test\_reader()，每次从缓存中随机读取批次大小的数据。
2. **模型配置**：根据数据集的形式和问题的描述，用简单的回归模型解决即可。由于数据量小，问题简单，先假定回归模型中最简单的线性模型。分别定义输入和输出张量，以及网络评估的损失函数和模型优化的优化函数。这些都相当于是模型的超参数配置，配置完毕后得到两个fluid.Program:fluid.default\_startup\_program() 与fluid.default\_main\_program()。
3. **模型训练与评估**：首先需要创建模型训练的场所，是串行还是并行，是使用CPU还是GPU运算。我们创建Exceutor定义运算场所，从而接受传入的Program，通过run()方法运行我们的模型。其次是定义数据维度，DataFeeder负责将数据提供器(之前提到的train\_reader和test\_reader)返回的数据转变成一种特殊的数据结构，从而适应Exceutor。然后定义绘制训练过程中损失值变化趋势的方法，从而可以观察模型的训练情况。最后Exceutor接受训练模型与训练数据，进行训练并将模型保存。可以通过观察损失值曲线判断模型的训练情况，并优化超参数等。
4. **模型预测**：通过比较训练结果，从所有已保存模型中取出表现结果最佳的模型，创建预测用Exceutor后对测试集进行预测，并呈现最终结果。
5. **程序代码及程序运行结果图**





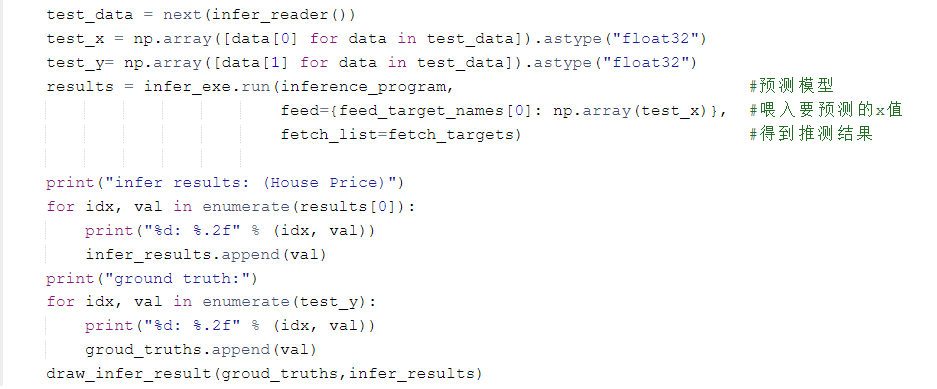


最后几轮的损失值和全程的损失值变化趋势：

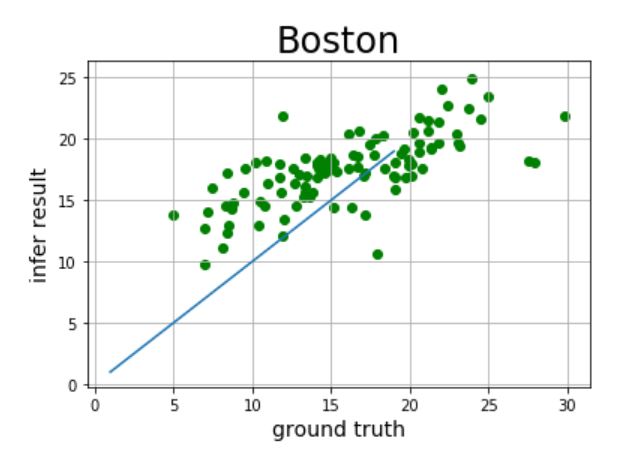








最后利用训练好的模型，对测试数据进行预测



# 3.实验遇到的主要问题及解决方案

不知如何使用提供的数据接口，对PaddlePaddle 的API不是很熟，不知如何才能查看数据详情，或送入模型训练：

通过查看教程知，可使用next()查看数据样例，并需要通过数据提供器来组织数据。