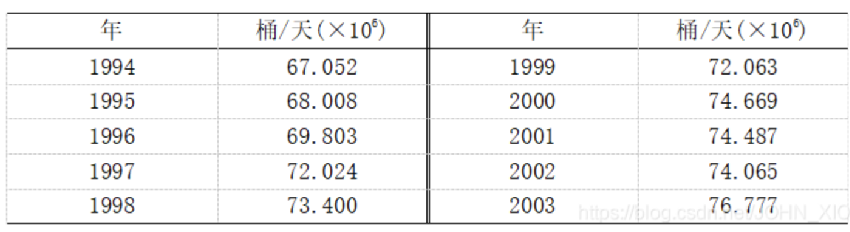
# 第四次上机实验报告

姓名：刘龙 学号：3119009062 班级：软件工程2班

## 实验 石油产量预测

世界石油产量以每天百万桶计，如下表所示，求最佳最小二乘法数值估计：



表一 历年石油产量

【实验条件】

计算机配置：64 位操作系统, 基于 x64 的处理器

CPU：Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz

内存大小：8GB

操作系统：Windows

【算法介绍】

使用多项式拟合，求出拟合曲线（采用leastsq函数）

假设“逼近”规律的近似函数为  ，即有

（1）

它与观测值 之差 称为残差。残差大大小可以作为衡量近似函数好坏的标准。按照使残差的平方和 最小的规则求得近似函数 的方

法称为最佳平方逼近，也称为曲线拟合的最小二乘法。

【代码】

import matplotlib.pyplot as plt

import numpy as np

from scipy.optimize import leastsq

plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签

plt.rcParams['axes.unicode\_minus'] = False # 用来正常显示符号

plt.figure()

# 画图,导入数据

x = np.arange(0, 10, 1, dtype='float')

y = np.array([67.052, 68.008,

69.083, 72.024, 73.400, 72.063, 74.669, 74.487, 74.065, 76.77], dtype='float')

plt.xlabel('年份')

plt.ylabel('石油产量\*10^6')

plt.plot(x, y, 'k.')

# 进行多项次拟合

# 一次多项式

param0 = [0.0, 0]

def leaner(s, x):

k1, b = s

return k1 \* x + b

# 二次多项式

param1 = [0.0, 0.0, 0.0]

def quadratic\_fun(s, x):

k1, k2, b = s

return k1 \* pow(x, 2) + k2 \* x + b

# 三次多项式

param2 = [0.0, 0.0, 0.0, 0.0]

def cubic\_fun(s, x):

k1, k2, k3, b = s

return k1 \* pow(x, 3) + k2 \* pow(x, 2) + k3 \* x + b

# 计算残差

def dist(a, fun, x, y):

return fun(a, x) - y

funs = [leaner, quadratic\_fun, cubic\_fun]

params = [param0, param1, param2]

colors = ['blue', 'red', 'yellow']

fun\_name = ['leaner', 'quadratic\_fun', 'cubic\_fun']

for i, (func, param, color, name) in enumerate(zip(funs, params, colors, fun\_name)):

var = leastsq(dist, np.array(param, dtype='float'), args=(func, x, y)) # 求出残差和最小的系数

# 检验拟合效果，画图

plt.plot(x, func(var[0], x), color)

# 预测2010年石油产量

print("{}预测的产量为{}".format(name, func(var[0], 2010 - 1994)))

print("[%s] 二范数：%.4f,abs(bias):%.4f,bias-std:%.4f" % (name,

(pow(y - func(var[0], x), 2).sum()),

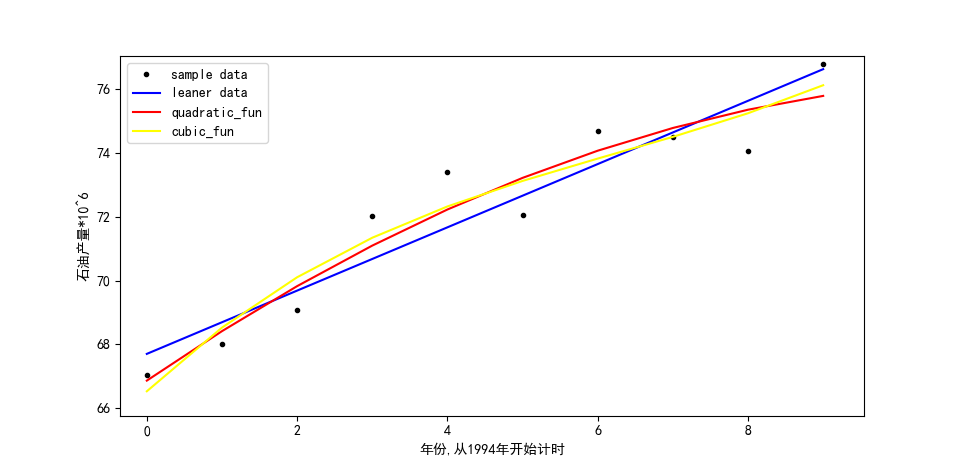
(y - func(var[0], x)).std(),

(abs(y - func(var[0], x))).mean()))

plt.legend(['sample data', 'leaner data', 'quadratic\_fun', 'cubic\_fun'], loc='upper left')

plt.show()

【实验结果分析】

1. 三种拟合函数的拟合情况

图一 拟合情况

1. 分析残差

|  |  |  |  |
| --- | --- | --- | --- |
| 多项式类型 | 二范数的平方 | 残差的标准差 | 残差的绝对值的均值 |
| 一次 | 9.96 | 0.85 | 0.99 |
| 二次 | 7.39 | 0.77 | 0.86 |
| 三次 | 6.58 | 0.76 | 0.83 |

表2 各拟合多项式残差等相关统计量分布

从表中可以看出，三者的预计差别很大，但是由于三次多项式的二范数平方，残差的标准值以及其均值比其他两者较小

1. 预计2010年每天石油产量

|  |  |
| --- | --- |
| 多项式类型 | 每天产石油（天/桶（\*10^6）） |
| 一次 | 83.55 |
| 二次 | 74.90 |
| 三次 | 92.83 |

通过查询真实2010年石油数据【1】，为每天 桶，和三次多项式的预测最接近，所以选择三次多项式

【引用】

[1] 2010年全球石油产量 https://data.gotohui.com/show-151157