哈爾濱工業大學

人工智能数学基础实验报告

题　　目	线性方程组求解
学　　院	计算机科学与技术
专　　业	人工智能
学　　号	2022113416
学　　生	刘子康
任 课 教 师	刘绍辉

哈尔滨工业大学计算机科学与技术学院

2024.3

实验一:数据拟合

一、实验内容或者文献情况介绍

使用 RANSAC 方法和最小二乘法拟合直线和曲线方程。

1.1 直线拟合

根据直线方程$ax+by+c=0$和噪声生成一系列随机散点，分别使用 RANSAC 方法和最小二乘法拟合直线方程中的参数，如果有一系列平行直线$ax+by+c_1=0, ax+by+c_2=0, ax+by+c_3=0$,并对直线上的点添加类似的噪声，拟合这些平行直线；

1.2 曲线拟合

设计曲线方程（例如圆方程或椭圆方程），添加适当的噪声（例如高斯噪声），分别采用 RANSAC 方法和最小二乘法进行拟合，然后添加一些外点，改进所使用的方法。

二、 算法简介及其实现细节

2.1 RANSAC 方法

2.1.1 算法简介：RANSAC 算法是从一组含有外点(outliers)的数据中正确估计数学模型参数的迭代算法。其中“外点”一般指数据中的噪声，例如匹配中的误匹配和估计曲线中的离群点。

2.1.2 实现细节（以直线方程为例）：随机选择两点（确定一条直线所需要的最小点集），由该两点确定一条线 l；根据阈值 t，确定与直线 l 的几何距离小于 t 的数据点集 S(l)，并称之为直线 l 的一致集；重复若干次随机选择，得到直线l_1, l_2, …, l_n和相应的一致集 S(l_1), S(l_2), …, S(l_n)；求最大一致集的最佳拟合直线，作为数据点的最佳匹配直线。

2.2 最小二乘法

2.2.1 算法简介：最小二乘法是一种数学优化技术，通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。

2.2.2 实现细节：采用多项式$y(x,w)=w_0+w_1x+\cdots+w_M\ x^M=\sum_{j=0}^{M}w_jx^j$进行逼近。为确定系数$w_i$，将原问题形式化为一个最优化问题$min\ E(w)$，其中误差平方和$E(w)=\frac{1}{2}\sum_{n=1}^{N}[y(x_n,w)-t_n]^2$，调整参数 M 的值以达到最佳拟合效果，并使数据大小为参数的 5-10 倍左右。为防止模型过拟合，使用正则化方法限定权系数和约束拟合模型，新误差方程为$\tilde{E}(w)=\frac{1}{2}\sum_{n=1}^{N}[y(x_n,w)-t_n]^2+$

$\frac{\lambda}{2}\|w\|_2^2$，调整 λ 的值并观察拟合结果。

三、 实验设置及结果分析（包括实验数据集）

3.1 直线拟合

以直线方程$3x-y-20=0$、$3x-y+60=0$、$3x-y+140=0$为例，利用 Numpy 库的 random 模块生成一系列随机 x 值及代入方程后求得的 y 值，对 y 值添加均值为 0、方差为 4 的高斯噪声。

3.1.1 RANSAC 方法：设置阈值 t=8，添加 40%的外点，200 次随机选点并计算一致集，输出最大一致集对应的直线方程参数；

3.1.2 最小二乘法：利用 scipy 库的 stats 模块中的 linregress 函数求解直线方程参数。

```
原始方程：y=3x+-20  y=3x+60  y=3x+140
最小二乘法结果：y=2.97x+-19.36(E=0.0450)  y=2.99x+59.58(E=0.0698)  y=3.10x+140.55(E=0.0583)
RANSAC方法结果：y=2.48x+-11.22  y=3.42x+50.95  y=2.84x+141.65
```

图 1：直线拟合参数及误差评价

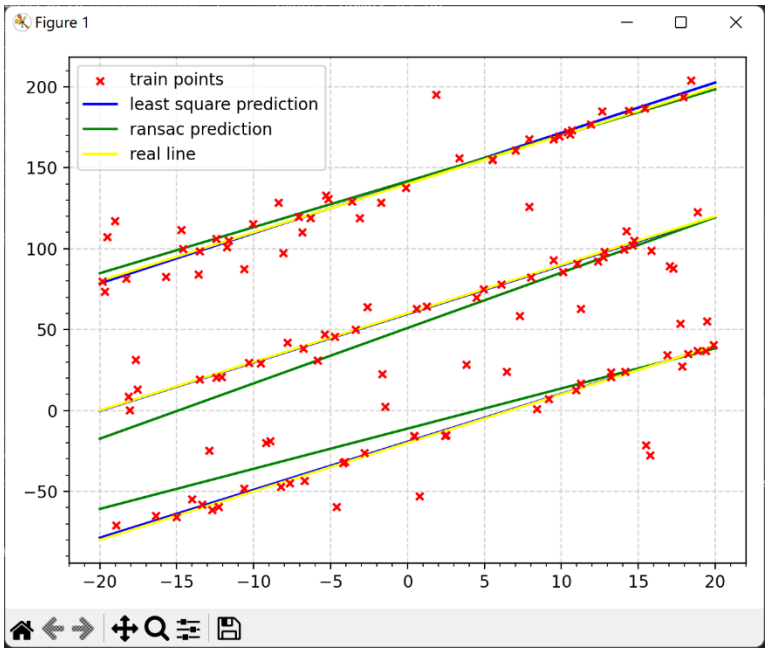


图 2：直线拟合结果

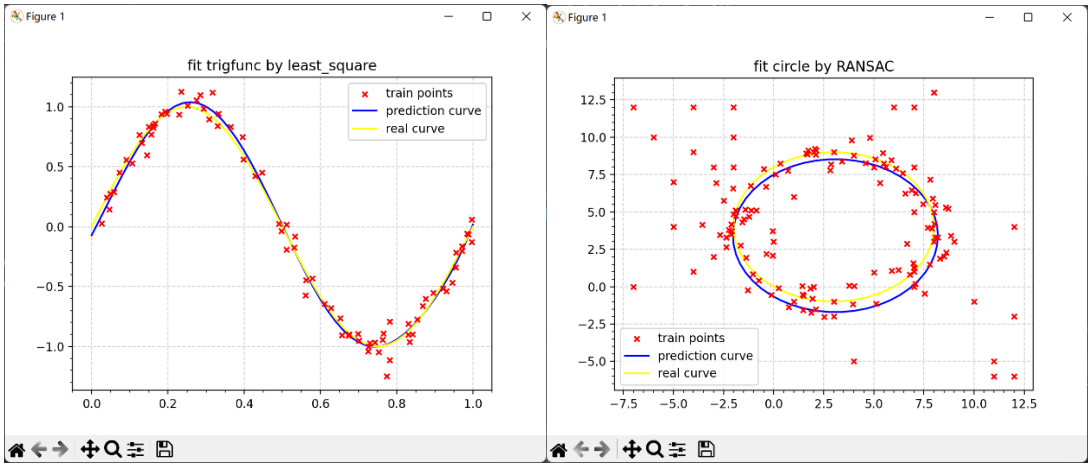
3.2 曲线拟合

3.2.1 RANSAC 方法：以圆方程$(x-3)^2+(y-4)^2=25$为例，写出圆曲线的参数方程$\begin{cases}x=5\cos\theta\\y=5\sin\theta\end{cases}$，利用 Numpy 库的 random 模块生成(0,2π)的一系列随机 θ 值并带入参数方程求得 x 值和 y 值，并添加均值为 0、方差为 0.64 的高斯噪声，设置阈值为 t=1，添加 40%的外点，随机选择三点，利用 Numpy 库的 linalg 模块的 solve 函数求解线性方程组得到圆心坐标，并计算拟合模型的半径，重复 10000 次，输出最大一致集对应的圆方程的圆心和半径；

3.2.2 最小二乘法：以三角函数曲线$y = \sin 2\pi x$为例，取其中一周期(0,2π)作为 x 的取值范围，利用 Numpy 库的 random 模块生成一系列随机 x 值及代入方程后求得的 y 值，对 y 值添加均值为 0、方差为 0.09 的高斯噪声，写出带正则化修正的误差平方和函数$\tilde{E}(w)$，利用 scipy 库的 optimize 模块中的 minmize 函数求解拟合模型的圆心和半径。

```
RANSAN方法：圆心：(3.08, 3.41) 半径：r=5.12
```

图 3：RANSAC 方法拟合圆方程参数



(a)最小二乘法　　(b)RANSAC 方法

图 4：曲线拟合结果

四、　结论

最小二乘法和 RANSAC 方法均可以较好地拟合直线或曲线方程。

最小二乘法的拟合效果与多项式阶数和数据点的分布有关，阶数过低会欠拟合，阶数过高会过拟合，并且在 y 有异常点时效果不佳。模型需要添加正则项对近似模型进行约束，通过权系数$\boldsymbol{\lambda}$控制正则项与误差项之间的均衡程度；数据量应是参数的 5-10 倍。

RANSAC 方法的拟合效果与迭代次数和阈值 t 有关，适合有外点和噪声较大的情况。可以通过调整迭代次数和自适应算法终止抽样改进模型。

五、　参考文献

[1] 泠山.RANSAC 基本原理[EB/OL].CSDN 博客,2024-03-04[2024-03-25].https://blog.csdn.net/qq_28087491/article/details/107376740

[2] Einstellung.最小二乘法和正则化[EB/OL].CSDN 博客,2019-02-03[2024-03-25].https://blog.csdn.net/Einstellung/article/details/86761710

[3] 曹连江著,电子信息测量及其误差分析校正的研究,东北师范大学出版社,2017.09,第 260 页

[4] 肖文博主编,统计信息化 Excel 与 SPSS 应用,北京理工大学出版社,2017.01,第 160 页

[5] 侯洪凤，王璨，曾维佳,管理信息系统基础,中国铁道出版社,2018.06,第 14 页