

用最小二乘法求解线性回归方程的算法研究

□ 戚龙

摘要: 如果给出的样本数据基本分布在一条直线附近, 我们可以根据样本数据求出这条直线, 求出的这条直线就叫做线性回归方程, 而线性回归方程通常用于预测数据使用。但是很多人在使用线性回归方程进行数据预测时, 并不知道线性回归方程的参数是如何得出的以及并不理解它的原理, 从而经常计算出错误的方程参数导致失败的预测结果。本文在此基础上对如何使用最小二乘法得出线性回归方程的原理进行了分析并应用线性回归方程进行数据预测算法研究。

关键词: 线性回归方程; 最小二乘法

一、最小二乘法

最小二乘法又称最小误差平方和法, 即选择一个最佳函数, 经过这个所选择函数的计算可以使实际数据与计算得出的数据之间误差平方和为最小。最小二乘法通常用于解决一些优化问题和线性回归问题。

二、线性回归方程

线性回归方程是指一条近似直线, 这条近似直线能最好地描述变量 x 与 y 之间的线性关系。换句话说线性回归方程就是试图得到一个线性模型, 用这个线性模型尽可能准确地拟合变量 x 与 y 之间的线性相关关系。观察样本点 (x, y) 的平面坐标图 (如图 1 所示) 我们可以看出所有的数据点都分布在一条直线附近, 而这样的直线可以有很多条, 但是我们要找到最能反映变量 x 与 y 之间关系的那条直线, 我们可以有以下几种方法:

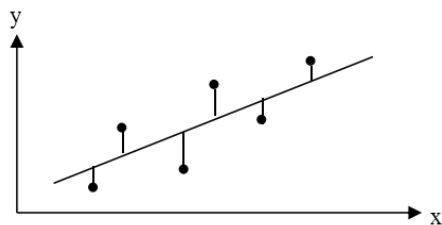


图1 (x, y) 散点图

(1) 可以使总离差等于所有离差之和, 但是这种方法因为正负离差可以相互抵消, 导致结果不准确, 一般不用这种方法求线性回归方程。(2) 可以使总离差等于所有离差的绝对

值之和, 但是求离差的绝对值不容易计算, 所以通常也不使用这种方法求线性回归方程。(3) 使用各个离差的平方和作为总离差, 当总离差得到最小值时求得的线性回归方程就是我们最终要求的线性回归方程。

三、线性回归方程的求解方法

对于一组具有线性相关关系的数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 我们设所求解的线性回归方程为 $\hat{y} = \hat{a} + \hat{b}x$, 其中 $\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$, $\hat{a} = \bar{y} - \hat{b}\bar{x}$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ 我们把 $\hat{a}$ 称为截距。

把 $\hat{b}$ 称为回归系数。

四、线性回归方程的预测算法研究

某工厂 2012 年至 2018 年的污水净化量分别为 1.020 万吨, 1.160 万吨, 1.240 万吨, 1.300 万吨, 1.450 万吨, 1.550 万吨, 1.620 万吨, 预测 2019 年该工厂污水净化量是多少。

在平面坐标图中我们可以直观的看出这些点都分布在一条直线附近, 在理论上我们就可以使用回归直线描述 y (污水净化量) 与 x (时间) 的关系。设 2012 年为第一年, 此时 $x=1$, 由基础数据可以知道: $\bar{y} = \frac{1}{7} \sum_{i=1}^7 y_i = 1.334\bar{x}$, $\bar{x} = \frac{1}{7} \sum_{i=1}^7 x_i = 4$, 设 y 对

x 线性回归方程为: $\hat{y} = \hat{a} + \hat{b}x$, 根据线性回归方程参数公式得: $\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0.0975$, $\hat{a} = \bar{y} - \hat{b}\bar{x} = 0.944$, 即 y 对 x 的线性回归方程

为: $\hat{y} = 0.944 + 0.0975x$, 2019 年对应的 $x=8$, 把 $x=8$ 代入线性回归方程计算得 $\hat{y} = 0.944 + 0.0975 \times 8 = 1.724$, 根据计算结果我们可以预测 2019 年该工厂污水净化量应该为 1.724 万吨。

参考文献

- [1] 陆健. 最小二乘法及其应用 [J]. 中国西部科技. 2007 (19).
- [2] 梁博尧. 浅析最小二乘法及其在数学建模中的应用 [J]. 中国新通信. 2018 (03).
- [3] 任建英. 一元线性回归分析及其应用 [J]. 才智. 2012 (22).
- [4] 龚循强, 李通, 陈西江. 总体最小二乘法在曲线拟合中的应用 [J]. 地矿测绘. 2012 (03).

(作者单位: 吉林化工学院信息与控制工程学院)

作者简介: 戚龙 (1979 ~), 男, 满族, 讲师, 研究方向为模式识别。

(上接第 194 页)

带学生组装一台四旋翼 450 无人机, 在组装的过程中学生就能了解一台多旋翼无人机组成的基本结构。同时我要他们说出每个主要部件的作用 (例如: 飞控、电机、电调、GPS 等), 只有了解每个部件的作用才能在出现故障时分析出是哪个部件出问题。在学习充电时, 我就把电池各参数的作用, 还有一些电学理论知识渗透进去, 学生掌握好才能在充电时设好电流电压, 保证充电安全。在装桨叶时, 我会解释桨叶为什么要正反转开装, 电机转向也要顺逆错开, 学生从中渗透学习空气动力学的基础知识。如此一来学生在动手的过程中不知不觉就学习到了一些基础理论知识。

四、实践成效与反思

经过一年多实践探索, 我的竞赛团队, 按上述训练模式进行训练, 学生水平不断提高, 先后参加 2019 年广东省职业院校技能大赛低空无人机技能与应用项目中获得一等奖第一名以及在 2019 年 Infomatrix 国际中学生资讯科技大赛中获得无人

机项目金奖的第一名的好成绩。从效果来看, 这种模式能比较有效的提高学生的无人机操控和组装技术水平。

但是这种模式也有一定的局限性, 首先对设备的要求比较高, 一般学校很难把不同类型的无人机都能买回来给学生训练, 其次这种方法只能适合学生人数比较少的情况下进行, 像无人机组装一边讲理论一边装, 这种方法在大班教学中很难开展; 最后, 学生的方向感、反应速度都是操控无人机的先天条件, 这些条件因人而异, 因此在具体训练中我们还需要针对学生特点因材施教。

参考文献

- [1] 翁诚浩. 中等职业学校无人机课程教学策略与实践 [J]. 现代职业教育. 中职中专. 2017 (04).
- [2] 柴源. 中职学校开设无人机应用专业探讨 [J]. 机电技术. 2018 (04).

(作者单位: 佛山市顺德区梁銶琚职业学校)