

用于地方高校学费分析 BP 神经网络数学模型方法研究

董亚楠

(兰州交通大学电子与信息工程学院 兰州 730070)

摘要 主要研究分析影响地方高校学费的因素。利用主成分分析法和 BP 神经网络建立数学模型,对 31 个省、直辖市的 7 项数据进行了分析。分析结果表明,前 3 个主成分的贡献率累计达到 87.39%,这表明,前 3 个主成分已基本能表述所有信息。为进一步提高精确度,使用前 5 项数据作为 BP 网络的输入数据,将各省的学费作为 BP 网络的输出。经过 6 次训练,输出误差可降至 10^{-15} ,基本达到了完全拟合的效果,为进一步预测学费提供了一种数学途径。

关键词 BP 神经网络,主成分分析,数学模型,高校学费
中图分类号 O212.4 文献标识码 A

Research on BP Neural Network Mathematical Model Used for Analyzing Local University Tuition

DONG Ya-nan

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract One mathematical way was introduced in this article to study and analyze the factors that affect the local university tuition. It was brought forth that principal component analysis (PCA) and BP neural network were confirmed to create a mathematical model to analysis the 7 data of 31 provinces or province-level municipality. The results of PCA indicated that the first 3 principal components accumulated contribution rates to 87.39% which showed that the first 3 principal can express all the information. It used the first 5 data as the input data of BP neural network in order to improve accuracy further, and the provincial tuition as BP network output. After six training sessions, the output error can be reduced to 10^{-15} which basically achieved the full effect of fitting, and also provided a mathematical way to predict further university tuition.

Keywords BP neural network, PCA, Mathematical model, University tuition

我国高等学校学费合理性已经成为社会关注的热点问题,已有不少文献利用数学建模的方法对这一问题进行了详细的研究^[8],不过进行定量分析的比较少。本文收集一些可能影响学费的相关数据,建立一个数学模型,对学费标准进行定量分析,并给出相对合理的学费标准。

高等院校的学费与很多因素有关,我们可以把这些因素归结为 3 个方面:国家对教育的支持情况、人民对学费的承受能力及高校培养学生的情况。这 3 方面的因素可细化为以下 7 个因素:国家生均拨款、人均 GDP、家庭收入、生均培养费用、高校非学费收入占事业性收入的比重、事业性经费支出占教育经费收入比重、预算内事业性教育经费拨款占教育经费收入比重。

使用 MATLAB7.11.0 软件工具,首先通过主成分分析法,对这 7 个因素进行数据处理、分析、判定,得出对学费影响较大的几个因素。然后建立 BP 神经网络模型,利用这种模型强有力的学习能力,实现输入与输出之间的高度非线性映射,找出地方高校学费与这些因素之间的关系,为相关部门的决策提供理论依据。

1 分析方法

1.1 主成分分析法

主成分分析是将多个相关变量简化为少数几个不相关变

量的一种多元统计方法。其目的在于简化统计数据,减少次要因素,并揭示变量之间的相互关系。每个主成分是初始变量的线性组合,所有的主成分之间相互正交,所以没有冗余信息,它们构成数据空间的正交基。从数学的角度看,其根本思想在于降维,而降维是从简化方差和协方差的结构来考虑的。

主成分分析法的主要实现步骤如下:

(1) 求出协方差矩阵 A (由 a_{ij} 构成), $a_{ij} = cov(x_i, x_j)$, 其中 $i, j = 1, 2, \dots, p$ 。

(2) 求矩阵的特征值,并依次排序得 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ 。

(3) λ_i 对应的单位正交化的特征向量为 $e_1, e_2, e_3, \dots, e_p$, $Y_i = e_i^T X, L_i = e_i$, 且满足:

a) $C = \lambda_k = e_k^T \sum e_k$;

b) $cov(Var(Y_i), Var(Y_j)) = 0, i \neq j$;

c) $|L_k| = 1$, 并称 $Y_k = e_k^T X$ 为 X 的第 k 个变量的主成分;

d) $\sum_{k=1}^p Var(Y_k) = \sum_{k=1}^p \lambda_k$, 称 $\lambda_k / \sum_{i=1}^p \lambda_i$ 为第 k 个主成分的贡献率。

(4) 比较 $\lambda_k / \sum_{i=1}^p \lambda_i (k=1, 2, \dots, p)$ 的大小,假设第 k 个主成分的贡献率最大,则依据 $Y_k = e_k^T X$ 判断 e_k 向量中相对较小的几个值,确定并去掉与它们对应的 x 值,从而筛选出主要因素,达到减少构成因素的个数的目的。

董亚楠(1988-),男,硕士,主要研究方向为数据分析、图像融合。

1.2 BP 神经网络

BP 网络是一种单向传播的多层前向网络,网络除输入和输出节点外,还有一层或者多层隐含节点,同层节点之间没有任何耦合,如图 1 所示。输入信号从输入层节点经隐含层传输到输出层,由于在神经网络的学习过程中,各层节点之间联接权值的调整采用的是反向传播算法,因此,这种网络又称为 BP 网络。利用 BP 网络对现有的数据进行拟合和预测,这种方法的好处在于可以将学费、人均 GDP、生均教育经费支出、国家生均拨款等因素之间的关系看作一个黑盒子模型,不必考虑这些因素之间的具体关系。经过已有数据对 BP 网络的训练,表明该 BP 网络即可具备拟合以及预测数据的能力。

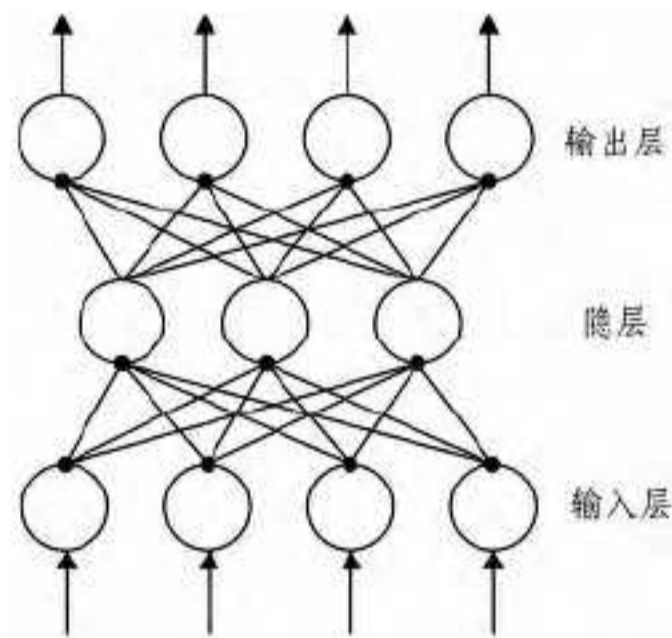


图 1 BP 网络模型结构

BP 网络的学习过程分为两个阶段:

第一个阶段是输入通过主成分分析法筛选出的影响学费的因素样本,通过设置的网络结构和前一次迭代的权值和阈值,从网络的第一层向后计算各神经元的输出。

第二个阶段是对权值和阈值进行修改,从最后一层向前计算各权值和阈值对总误差的影响,据此对各权值和阈值进行调整。

以上两个过程反复交替,直至达到收敛为止。本模型中采用 Levenberg-Marquardt 算法对网络进行训练,在训练之前对样本数据进行规范化处理,使输入和输出数据规范到 $[0, 1]$ 区间。传递函数采用 Sigmoid 型函数:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

理论上已经证明,在不限制隐层节点数的情况下,只有一个隐层的 BP 网络可以实现任意非线性映射^[5]。在模式样本相对较少的情况下,较少的隐层节点,可以实现模式样本空间的超平面划分,此时,选择两层 BP 网络就可以了。但如果模式样本数目很多时,减少网络规模,增加一个隐层是必要的,但 BP 网络隐层层数一般不超过两层。本模型选用单隐层 BP 网络。

隐层节点数目的确定一直是一个复杂的问题,迄今为止,尚未有一个很好的解析式,隐层节点数目往往根据前任设计所得到的经验和自己进行试验来确定。一般认为,隐层节点数与求解问题的要求、输入输出单元数多少都有直接的关系。另外,隐层节点数太多会导致学习时间过长,而隐层节点数太少,容错性差,识别未经学习的样本能力低,所以必须综合多方面的因素进行设计。本文根据经验公式:

$$n = \sqrt{n_i + n_o} + a \quad (2)$$

确定隐层节点数。式中, n_i 为输入节点数, n_o 为输出节点数, a 为 $1 \sim 10$ 之间的常数。

为防止出现“过适配”问题,进一步提高 BP 网络的泛化能力,本文采用归一化法提高 BP 网络的泛化能力,以保证训练后的 BP 网络能够对训练集之外的数据具有良好的预测作用。

归一化法的基本思想是修正网络误差性能函数。典型的前向反馈网络误差性能函数为均方误差函数:

$$F = mse = \frac{1}{N} \sum_{i=1}^N e_i^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (3)$$

如果增加一项,该项包含网络权值和阈值的均方值,利用该项修正误差性能函数,就可以提高网络的泛化能力,即

$$msereg = \gamma \cdot mse + (1 - \gamma)msw \quad (4)$$

式中, γ 为误差性能调整率,而

$$msw = \frac{1}{n} \sum_{j=1}^n x_j^2 \quad (5)$$

用式(4)修正的误差性能函数,可以使网络获得较小的权值和阈值,从而迫使网络的响应变得平滑,从而减小过适配现象,为训练好的 BP 网络进行学费的准确预测打下基础。

2 实验结果及分析

2.1 主成分分析结果

通过对地方高校学费统计表中的数据进行主成分分析,得到 7 个不同成份的贡献率分别为: 0.4422, 0.2223, 0.2094, 0.0585, 0.0400, 0.0217, 0.0055。

所得地方高校学费分析的主成分载荷如表 1 所列。

表 1 地方高校学费分析的主成分载荷

	主成分 1	主成分 2	主成分 3
载荷 1	0.862	-0.262	0.334
载荷 2	0.413	0.610	-0.515
载荷 3	0.875	0.076	-0.315
载荷 4	0.967	-0.000	0.054
载荷 5	-0.478	-0.665	-0.420
载荷 6	-0.052	0.152	0.901
载荷 7	0.501	-0.803	-0.011

主成分载荷表示的意义是指主成分与原因素之间的关系程度,绝对值越大,表明此主成分对原因素的代表性越大。由表 1 分析可知,前 3 项的贡献率累计之和达到 87.39%,已基本包含了全部因素的相关信息,在这里,为了能够进一步提高拟合效果,在 BP 神经网络训练中,用前 5 项数据作为其输入,用各省学费作为 BP 神经网络模型的输出。即舍弃人均 GDP、非学费收入占事业收入比重 2 个因素。

2.2 神经网络实验结果

根据主成分分析结果,地方院校学费水平主要受国家生均拨款、家庭收入、生均教育经费支出、事业性经费支出占教育经费收入比重和预算内事业性教育经费拨款占教育经费收入比重 5 个因素的影响,因此,将各省的这 5 个数据作为 BP 网络的输入数据,将各省的学费作为 BP 网络的输出,对 BP 网络进行训练,经过 6 次训练,输出误差可降至 10^{-15} 。利用训练好的 BP 网络对各省的地方院校数据进行拟合,可以得到良好的效果,如图 2 所示。

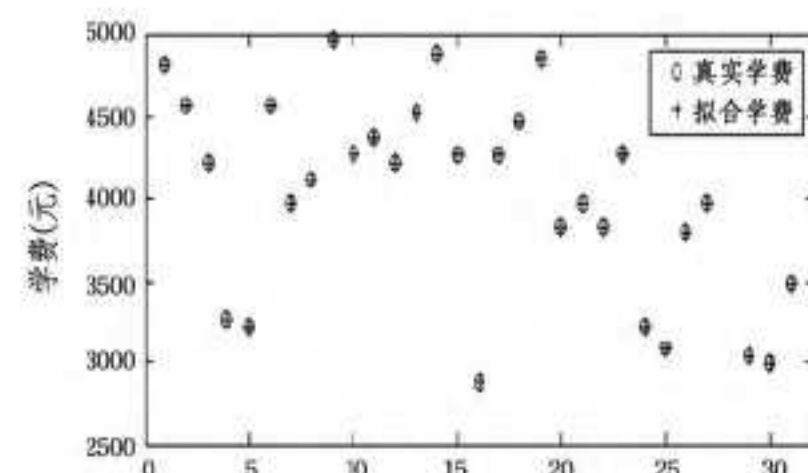


图 2 利用 BP 网络对地方院校数据进行拟合

由于采用了归一化算法提高该网络的泛化能力,因此可假定该网络对非训练集数据具有良好的预测能力,故可用该网络预测各因素单独变化时对学费水平的影响情况。

图3所示为国家生均拨款增加10%时,各省地方院校学费水平的变化情况。

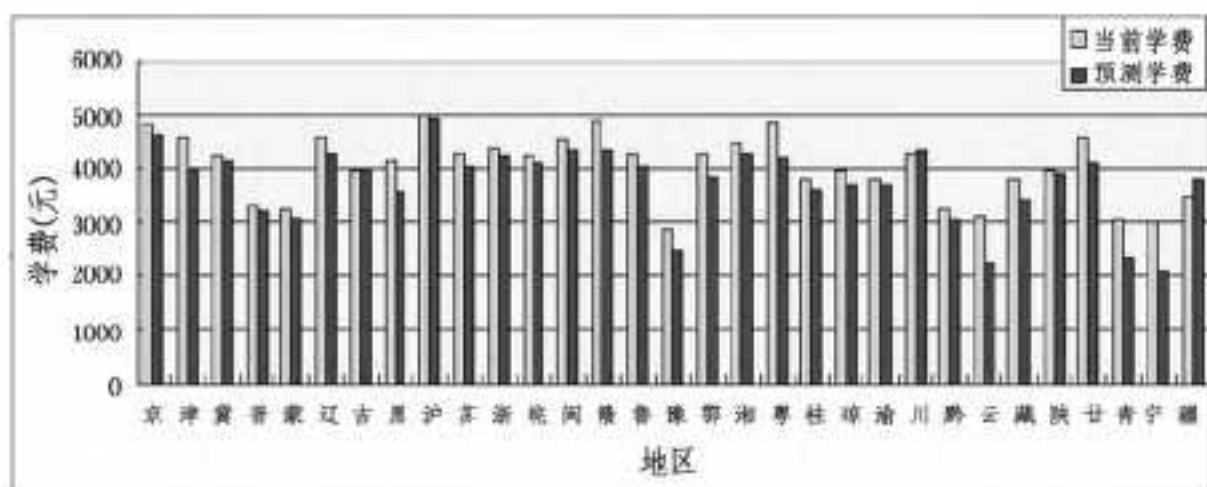


图3 利用BP网络预测国家生均拨款增加10%时学费的变化情况

从图3中可以看出,绝大部分省份的院校随着国家生均拨款的增加,学费有不同程度的下降,平均减少286元。其中,甘肃、青海、宁夏、江西等欠发达地区学费的降幅较大。

图4所示为家庭收入增加10%时,各省地方院校学费水平的变化情况。

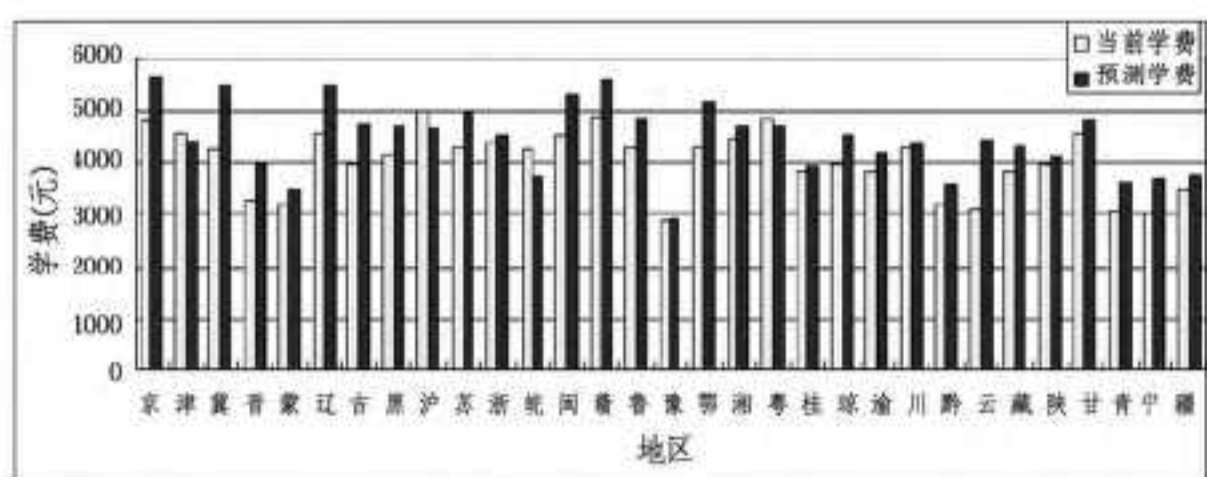


图4 利用BP网络预测家庭收入增加10%时学费的变化情况

从图4中可以看出,绝大部分省份的院校随着家庭收入的增加,学费有不同程度的增加,平均增加430元。

图5所示为生均教育经费支出增加10%时,各省地方院校学费水平的变化情况。

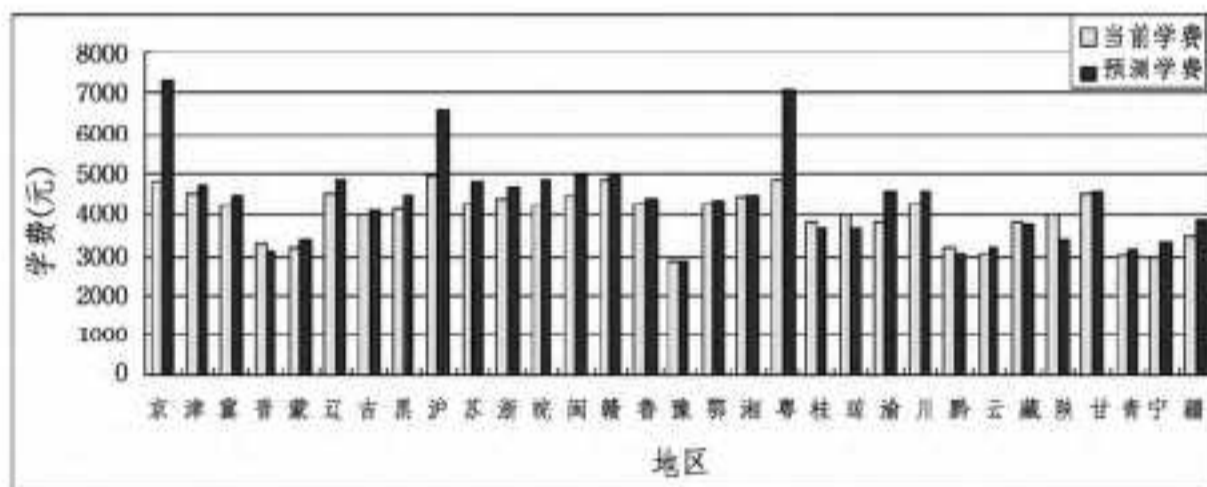


图5 利用BP网络预测生均教育经费支出增加10%时学费的变化情况

从图5中可以看出,绝大部分省份的院校随着生均教育经费的增加,学费有不同程度的增加,平均增加345元。其中,北京、上海、广东较发达地区增幅较大。

图6所示为事业性经费支出占教育经费收入比重增加一个百分点时,各省地方院校学费水平的变化情况。从图中可以看出,绝大部分省份的院校随着事业性经费支出占教育经费收入比重增加,学费有不同程度的增加,平均增加277元。

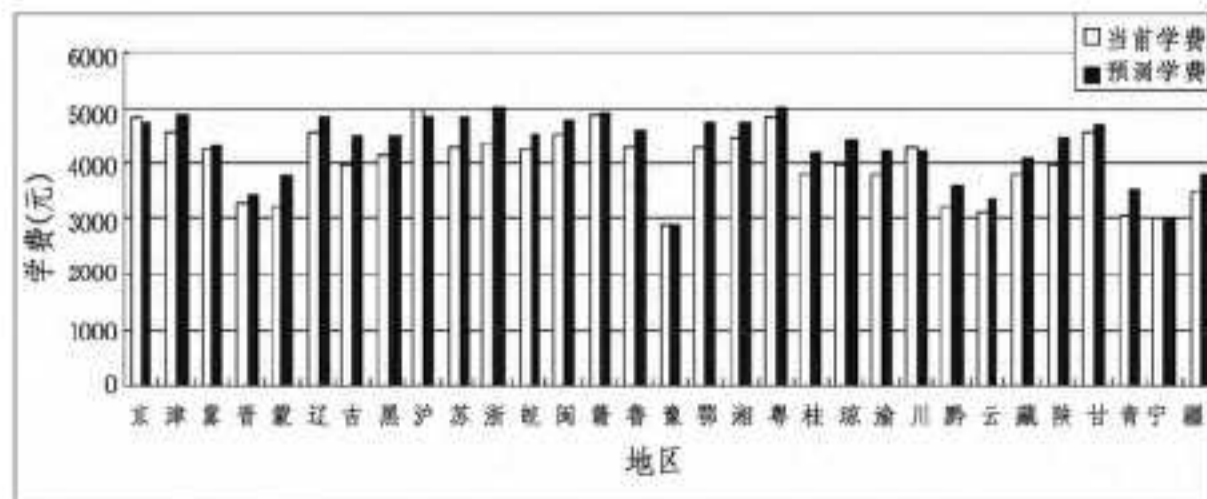


图6 利用BP网络预测事业性经费支出占教育经费收入比重增加一个百分点时学费的变化情况

图7所示为预算内事业性教育经费拨款占教育经费收入

比重增加一个百分点时,各省地方院校学费水平的变化情况。从图中可以看出,绝大部分省份的院校随着预算内事业性教育经费拨款占教育经费收入比重增加,学费有不同程度的降低,平均减少200元。其中,北京、上海、广东、福建等较发达地区降幅较大。

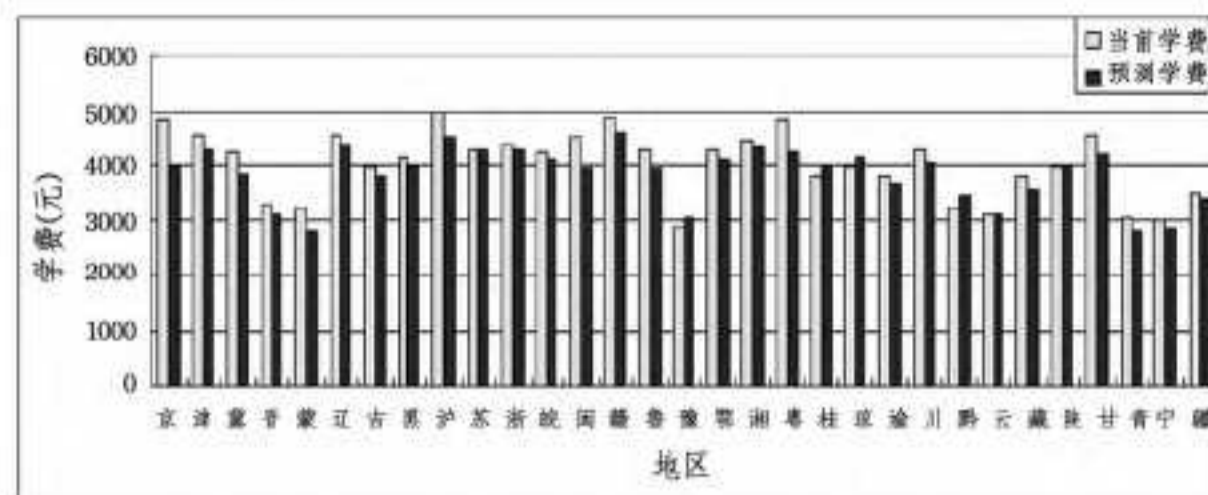


图7 利用BP网络预测预算内事业性教育经费拨款占教育经费收入比重增加一个百分点时学费的变化情况

结束语 高等学校的学费标准受多方面因素的影响,本文主要对人均GDP、国家生均拨款、家庭收入、生均教育经费支出、高校非学费收入占事业性收入的比重、事业性经费支出占教育经费收入比重和预算内事业性教育经费拨款占教育经费收入比重等7个因素对学费的影响进行了分析。

(1)从模型拟合与预测结果看,影响地方院校学费的第一因素是家庭收入,因此,在制定高校学费标准时,应该首要考虑学生的家庭收入情况,充分考虑学生的家庭负担;

(2)从分析结果看,地方院校的学费标准与人均GDP无关,而与事业性经费占教育经费支出的比例具有一定的相关度,因此,地方院校学费标准的制定应充分考虑到事业性经费占教育经费支出比例的影响,地方院校也应该充分进行科研创收等工作,以改善学费标准;

(3)生均教育经费支出对地方院校的学费标准的影响相差不多,因此在保证教育质量的前提下,适当降低生均教育经费支出,可在一定程度上改善高等院校的学费标准。

参考文献

- [1] 芮光正,许晓芊.大学生培养成本核算刍议[J].绵阳经济技术高等专科学校学报,2000,17(2):69-70
- [2] 邹婷婷,何军.高校生均培养成本模型及其系统研究[J].浙江理工大学学报,2008,25(3):305-308
- [3] 黄享苟.我国高校生均培养成本的一种计算方法[J].当代经理人,2006,(21):23
- [4] 吴群英,林亮.应用数理统计[M].天津:天津大学出版社,2004
- [5] 董长虹. Matlab神经网络与应用[M].北京:国防工业出版社,2007
- [6] 邱雅.我国普通高校学费模型的建立与分析[J].教育与经济,2006(4):32-34
- [7] 苏良军,孙便霞.高校学费影响因素及空间相关性分析[J].数理统计与管理,25(4):400-406
- [8] 艾灵芝,梅正阳,王波,等.主成分分析和回归法在高校学费标准评价中的应用[J].湖北师范学院学报,自然科学版,2009,29(3):99-103
- [9] 贾文,李金红.基于BP神经网络高等教育学费分析的模型[J].山东轻工业学院学报,2009,29(2):86-89
- [10] 王丽颖.高等教育学费标准探讨[J].白城师范学院学报,2013,27(3):13-16
- [11] 王莹.基于SPSS的南京地区高校学费水平与影响因素分析[J].2013(12):137-138
- [12] 石丹.基本BP神经网络的高校学费定价模型的建立[D].西安:西安科技大学,2013