

基于 PLS 模型的葡萄酒质量评价

席 倩, 万冰越

(南京师范大学泰州学院, 江苏 泰州 225300)

摘要:利用 2012 年全国大学生数学建模竞赛 A 题所给数据,建立了偏最小二乘回归(PLS)分析模型,选取包括芳香物质在内的所有的酿酒葡萄和葡萄酒理化指标作为自变量,以葡萄酒的质量作为因变量建立模型。通过自变量和因变量提取成分的贡献率的计算,选取了可以解释自变量比率为 86.28% 的 13 个成分对,得到葡萄酒的质量与所有理化指标之间的 PLS 回归方程。利用标准化指标变量的回归方程,得到对葡萄酒的质量影响较大的理化指标有 23 个,其中 13 个正向指标,10 个负向指标,正向指标中影响系数最大是葡萄的干物质含量(影响系数为 0.361 4),负向指标中影响系数最大的是葡萄的褐变度(影响系数为-0.378 8)。用 Matlab 编程对模型的精度进行了检验。检验结果表明如果不考虑酿酒工艺等技术、环境因素的影响,利用理化指标根据建立的偏最小二乘回归模型预测葡萄酒的质量具有很高的可信度。

关键词:PLS 模型;葡萄酒质量;理化指标;精度检验

中图分类号:X53 **文献标志码:**A **文章编号:**1671-1807(2014)02-0088-06

随着人们生活品质的提高,葡萄酒已经成为我国消费者喜爱的生活饮品。葡萄酒质量是其外观、香气、口味、整体评价的综合表现。在此形势下,葡萄酒的质量评价得到关注。由于葡萄酒的化学成分复杂,其质量需考虑酿酒葡萄和葡萄酒的理化指标。

国外对葡萄酒质量评价的研究成果比较丰富,其中 Cortez^[1]用支持向量机建立了葡萄质量分类模型,分集介绍效果较好。Moreno^[2]基于红酒的矿物质元素含量,用概率神经网络成功将 54 个样本分为两类。国内对此也有较多的研究葡萄酒质量的课题。天津大学的刘延玲^[3]构建了新的 Hopfield 神经网络分类器模型,利用葡萄酒的理化性质指标和感官评价结果将其质量进行分类,基于 Hopfield 神经网络的分类器成功实现了葡萄酒质量的评价分类,进一步挖掘指标变量之间的关系。吴昊潼,薛超,刘海明^[4]建立了 BP 神经网络模型分析葡萄和葡萄酒理化指标对酒质量的影响,李运,李记明,姜忠军^[5]用分析软件 SPSS 对 40 个样本进行了聚类分析,认为统计学方法应用于葡萄酒质量分析与评价中,可以更加清楚地了解葡萄酒成分与感官质量之间的相互关系。王金甲等^[6]

提出了一种可视化的葡萄酒质量评价方法。王文静^[7]论述了感官评价在葡萄酒加工工艺、成分分析和评价结果统计分析等方面的研究进展。张贝贝等^[8]运用多元总体的均值向量和方差的检验,得到感官评价结果的显著差异性应用主成分分析,逐步回归等方法进行分析。文献[9]综述了同位素比质谱仪法(IRMS)和点特异性天然同位素分馏核磁共振技术(SNIF-NMR)2 种化学方法的原理及其在葡萄酒质量鉴别中的应用。

本文以 2012 年高教社杯全国大学生数学建模竞赛 A 题数据作为建模^[10]的历史数据,运用偏最小二乘回归(PLS)对葡萄酒质量进行评价,选取酿酒葡萄和葡萄酒理化指标作为自变量,以葡萄酒的质量作为因变量建立模型。计算自变量和因变量提取成分的贡献率,得到相应的回归方程,并检验了模型的精度,分析了葡萄酒理化指标对葡萄酒质量的影响,探讨了利用葡萄酒的理化指标对葡萄酒质量进行评价的可行性。

1 模型建立

偏最小二乘回归分析^[11]是一种多对多线性回归

收稿日期:2013-12-10

基金项目:江苏省高校大学生创新训练计划项目(201313843012Y);江苏省软科学项目(BR2013039)

作者简介:席倩(1992—),女,江苏南京人,南京师范大学泰州学院数学科学与应用学院,理学学士,研究方向:数学与应用数学。

建模的方法,特别当两组变量的个数很多,且都存在多重相关性,而观测值的数量又较少时,用偏最小二乘回归建立的模型更具应用价值。

PLS 回归模型的原理和建模步骤在文献[11]中有较为详细的叙述,下面我们给出一种较为简洁的计算方法。

记自变量和因变量组标准化数据矩阵分别为 E_0 和 F_0

$$E_0 = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}, F_0 = \begin{bmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{np} \end{bmatrix}$$

Step1:求矩阵 $E_0^T F_0 F_0^T E_0$ 最大特征值所对应的特征向量 w_1 ,求得成分得分向量 $t_1 = E_0 w_1$ 和残差矩阵 $E_1 = E_0 - t_1 \alpha_1^T$,其中 $\alpha_1 = E_0^T t_1 / \|t_1\|^2$ 。

Step2:求矩阵 $E_1^T F_0 F_0^T E_1$ 最大特征值所对应的特征向量 w_2 ,求得成分得分向量 $t_2 = E_1 w_2$ 和残差矩阵 $E_2 = E_1 - t_2 \alpha_2^T$,其中 $\alpha_2 = E_1^T t_2 / \|t_2\|^2$ 。如此计算下去,至第 r 步,求矩阵 $E_{r-1}^T F_0 F_0^T E_{r-1}$ 最大特征值所对应的特征向量 w_r ,求得成分得分向量 $t_r = E_{r-1} w_r$ 。

Step3:进行交叉有效性检验,定义交叉有效性 $Q_h^2 = 1 - PRESS(h)/SS(h-1)$

其中 $PRESS(h) = \sum_{i=1}^p PRESS_j(h)$, $SS(h) =$

$\sum_{j=1}^p SS_j(h)$, $PRESS_j(h)$ 表示抽取 h 个成分时第 j 个因变量 y_j 的预测误差平方和; $SS_j(h)$ 表示采用所有的样本点,拟合含 h 个成分的回归方程时 y_j 的误差平方和。设定限制值为 0.05,在建模的每一步计算结束时,计算 Q_h^2 的值,如果在第 h 步有 $Q_h^2 < 1 - 0.95^2 = 0.0985$,则模型达到精度要求,对应的 h 即为所求的成分个数。

$$\begin{pmatrix} 1.0000 & 0.0235 & -0.1208 & \cdots & 0.3558 & -0.0827 & 0.1492 \\ 0.0235 & 1.0000 & -0.0143 & \cdots & 0.0466 & -0.2565 & 0.5175 \\ -0.1208 & -0.0143 & 1.0000 & \cdots & -0.3680 & -0.1932 & -0.0703 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0.3558 & 0.0466 & -0.3680 & \cdots & 1.0000 & -0.1288 & 0.2446 \\ -0.0827 & -0.2565 & -0.1932 & \cdots & -0.1288 & 1.0000 & 0.1545 \\ 0.1492 & 0.5175 & -0.0703 & \cdots & 0.2446 & 0.1545 & 1.0000 \end{pmatrix}$$

利用相关系数矩阵,可以对各个指标进行聚类分析,利用 MATLAB 编程得到各个指标间的聚类分析

Step4:根据第 3 步确定提取 r 个成分 $t_1, \dots, t_r$,

则求 F_0 在 $t_1, \dots, t_r$ 上的普通最小二乘回归方程为

$$F_0 = t_1 \beta_1^T + t_2 \beta_2^T + \cdots + t_r \beta_r^T + F_r$$

把 $t_k = \tilde{w}_{k1} x_1 + \cdots + \tilde{w}_{km} x_m (k = 1, 2, \dots, r)$,代入 $Y = t_1 \beta_1 + \cdots + t_r \beta_r$,即得 p 个因变量的偏最小二乘回归方程式

$$y_j = a_{j1} x_1 + a_{j2} x_2 + \cdots + a_{jm} x_m, (j = 1, 2, \dots, p)$$

其中 $\tilde{w}_h = (\tilde{w}_{h1}, \dots, \tilde{w}_{hm})$ 满足 $t_h = E_0 \tilde{w}_h$, $\tilde{w}_h = \prod_{j=1}^{h-1} (I - w_j \alpha_j^T) w_h$ 。

2 实证分析

利用 2012 年全国大学生数学建模竞赛 A 题^[10]数据,采用如下的步骤建立模型,利用葡萄酒的理化指标对葡萄酒的质量进行评价。

1)数据的选取。由于要分析酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响,我们以分析红葡萄为例,选取红葡萄的 31 个理化指标、红葡萄的芳香物质指标、红葡萄酒的 9 个理化指标和红葡萄酒的芳香物质指标,共 41 个指标作为红葡萄和红葡萄酒的所有的分析指标,依次分别用 $x_1, x_2, \dots, x_{41}$ 表示;选取红葡萄酒的质量指标作为因变量,用 y 表示。

2)数据的标准化处理。为提高计算精度,我们需要对数据进行标准化处理,其具体方法如下:

$$x_{ij}^* = \frac{|x_{ij} - \bar{x}_j|}{\sqrt{S_{jj}}}, i = 1, 2, \dots, n, j = 1, 2, \dots, p,$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \sqrt{S_{jj}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2},$$

$j = 1, 2, 3, \dots, p$ 。

3)求相关系数矩阵。从相关系数矩阵可以看出各个指标间是正相关还是负相关的关系,同时可以分析因变量与自变量间的正负相关性。

图像如图 1 所示。由图 1 可以看出,全部指标大致可以分成五类,其中第 1 类的有 3;第 2 类的有 14、28、

29、30、38;第 3 类的有 2、4、6、8、9、10、11、12、13、15、19、21、25、26、32、33、34、35、36、37、42;第 4 类的有 1、5、7、16、17、18、22、31、40;第 5 类的有 20、23、24、27、39、41。

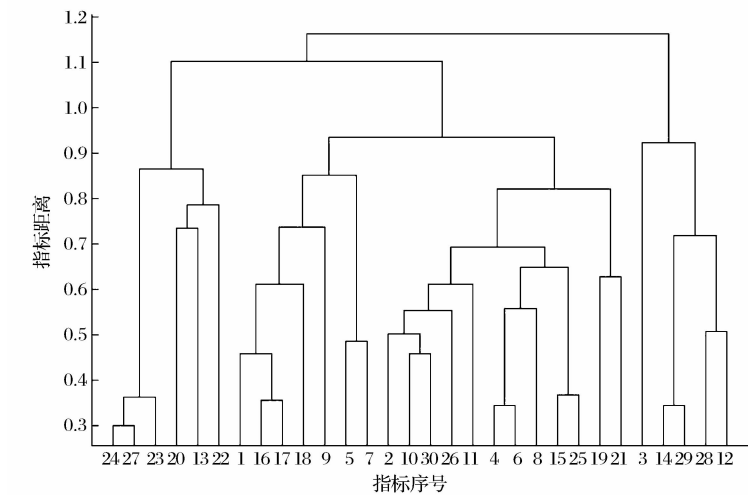


图 1 指标聚类结果

4) 分别提取自变量组和因变量的成分。利用 matlab 计算,自变量和因变量提取成分的贡献率如

表 1 自变量和因变量提取成分的贡献率

自变量	0.244 9	0.140 2	0.081 1	0.062 5	0.069 8	0.069 2	0.026 1
因变量	0.612 7	0.214 9	0.065 8	0.047 1	0.018 5	0.010 2	0.015 2
自变量	0.028 3	0.037 9	0.034 5	0.022 5	0.023 2	0.022 6	...
因变量	0.006 8	0.002 3	0.001 6	0.000 7	0.000 3	0.000 2	...

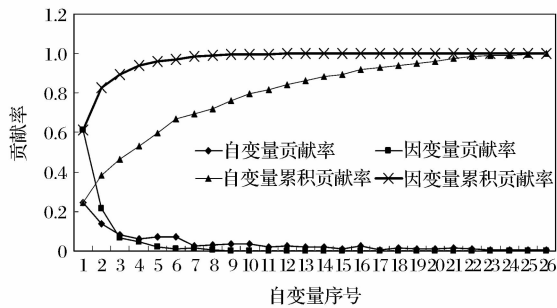


图 2 自变量和因变量提取成分的贡献率及累积贡献率

由图 2 可知,前 13 个成分解释自变量的比率为 86.28%,因此我们取 13 对成分进行分析。求得 13 个成分分别为:

$$\begin{aligned} u_1 &= -0.0298x_1 + 0.0307x_2 - 0.0018x_3 + \cdots - 0.0034x_{41} \\ u_2 &= -0.0480x_1 - 0.0194x_2 - 0.0129x_3 + \cdots - 0.0338x_{41} \\ &\vdots \\ u_{13} &= 0.0595x_1 + 0.1898x_2 - 0.1493x_3 + \cdots - 0.0195x_{41} \end{aligned}$$

13 个成分的回归系数图像如图 3 所示。

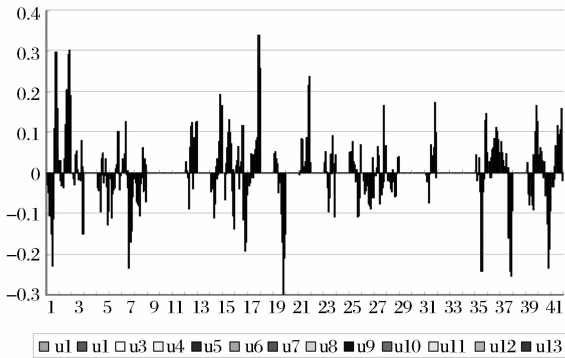


图 3 回归系数

5) 求取 13 个成分对时,标准化指标变量与成分变量之间的回归方程。求得自变量组和因变量与 13 个成分 $u_1, u_2, \cdots, u_{13}$ 之间的回归方程分别为

$$\begin{aligned} \bar{x}_1 &= 1.3641u_1 - 0.9225u_2 + 2.2687u_3 - 2.4993u_4 + \cdots - 0.5014u_{13} \\ \bar{x}_2 &= 3.3269u_1 + 0.3268u_2 + 0.0349u_3 - 0.0741u_4 + \cdots + 0.0845u_{13} \\ &\vdots \\ \bar{x}_{41} &= 0.9810u_1 + 0.0469u_2 - 0.6500u_3 + 1.8259u_4 + \cdots - 0.9684u_{13} \\ \bar{y} &= 3.9911u_1 + 2.3640u_2 + 1.3077u_3 + 1.1066u_4 + \cdots + 0.1359u_{13} \end{aligned}$$

6)计算因变量与自变量组之间的回归方程。将步骤 4)中成分 $u_i, i = 1, 2, \cdots, 13$ 代入步骤 5)中 $\bar{y}_i$ 的回归方程,得到标准化指标变量之间的回归方程

$$\begin{aligned} \bar{y} = &-0.1853\bar{x}_1 + 0.2783\bar{x}_2 + 0.0202\bar{x}_3 + \\ &0.0437\bar{x}_4 - 0.0940\bar{x}_5 - 0.1436\bar{x}_6 - 0.0698\bar{x}_7 - \\ &0.3788\bar{x}_8 - 0.0204\bar{x}_9 + 0.0985\bar{x}_{10} + 0.1352\bar{x}_{11} + \\ &0.1134\bar{x}_{12} + \cdots + 0.1721\bar{x}_{41} \end{aligned}$$

将标准化变量 $\bar{y}, \bar{x}_j (j = 1, 2, 3 \cdots 41)$ 分别还原成原始变量 y, x_j , 得到回归方程

$$\begin{aligned} y = &63.4839 - 0.0005x_1 + 0.0244x_2 + 0.0411x_3 + \\ &0.0019x_4 - 0.1162x_5 - 0.1501x_6 - 0.3760x_7 - \\ &0.1537x_8 - 0.0002x_9 + 3.4960x_{10} + 0.0811x_{11} + \\ &0.0681x_{12} + 0.0961x_{13} - 0.1256x_{14} + 0.0140x_{15} + \\ &0.0005x_{16} + 0.0338x_{17} - 0.0131x_{18} - 0.4150x_{19} - \\ &0.3794x_{20} + 0.2353x_{21} - 0.0746x_{22} + 0.0022x_{23} - \\ &0.0024x_{24} + 0.1695x_{25} - 0.1250x_{26} + 15.5231x_{27} - \\ &0.4477x_{28} - 0.1732x_{29} - 0.6472x_{30} + 0.0001x_{31} - \\ &0.0025x_{32} + 0.3334x_{33} + 0.0762x_{34} - 0.2447x_{35} + \\ &0.4598x_{36} - 1.3043x_{37} + 0.0243x_{38} - 0.0259x_{39} + \\ &0.0276x_{40} + 0.0082x_{41} \end{aligned}$$

7)模型的精度检验。为了考察我们建立的回归方程的模型精度,我们以 $(\bar{y}_i, y_i)$ 为坐标值,对所有的样本点绘制预测图,其中 $\bar{y}_i$ 表示在第 i 个样本点的

预测值,在预测图上,如果所有的点都能在图的对角线附近均匀分布,则说明回归方程的拟合值与原值差异很小,说明模型的拟合效果就是令人满意的。利用 Matlab,我们得到葡萄酒质量的预测图 4,从图上可以清晰的看到,所有的预测点都在对角线附近,因此,我们建立的回归模型的精度是令人满意的。

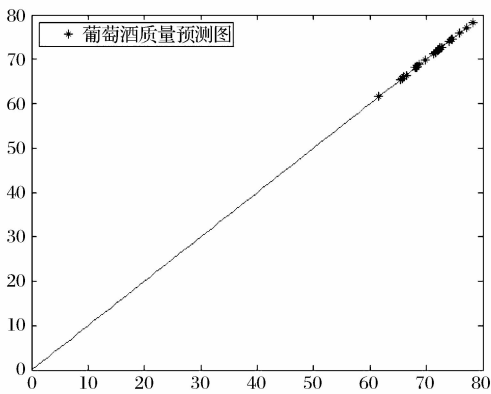


图 4 葡萄酒质量预测图

为进一步分析模型的精度,下面我们使用三种检验方式对模型的精度进行检验,分别是残差大小检验、后验差检验及关联度检验。残差大小检验是一种直观的逐点进行比较的算术检验。后验差检验属于统计概念,它按照残差的概率分布进行检验。关联度检验则属于几何检验,它检验的是模型曲线与行为曲线的集合相似程度^[10]。首先给出模型检验等级参照表(表 2)。

表 2 模型检验等级参照表

指标临界值 精度等级	相对误差 α	关联度 ξ_0	均方差比值 c_0	小概率误差 p_0
一级(好)	0.01	0.90	0.35	0.95
二级(合格)	0.05	0.80	0.50	0.80
三级(勉强)	0.10	0.70	0.65	0.70
四级(不合格)	0.20	0.60	0.80	0.60

利用模型模拟 27 组样品的质量得分,进行 Matlab 编程计算(程序见附录),可得模拟预测值与实际

值的误差分析参数 α, ξ_0, c_0, p_0 , 结果见表 3 和图 5。

表 3 模型精度检验结果

样品	1	2	3	4	5	6	7	8	9
实际值	68.1	74	74.6	71.2	72.1	66.3	65.3	66	78.2
预测值	68.054 5	73.979 7	74.583 7	71.175 9	71.960 1	66.455 6	65.426 9	66.007 9	78.306 9
样品	10	11	12	13	14	15	16	17	18
实际值	68.8	61.6	68.3	68.8	72.6	65.7	69.9	74.5	65.4
预测值	68.739	61.601 5	68.299 1	68.723 3	72.713 6	65.731 7	69.843 3	74.299 5	65.348 8
样品	19	20	21	22	23	24	25	26	27
实际值	72.6	75.8	72.2	71.6	77.1	71.5	68.2	72	71.5
预测值	72.397 3	75.809 4	72.146 3	71.6	77.234 8	71.568 1	68.221 8	72.086 9	71.614 3

$\alpha = 0.001, \xi_0 = 0.9853, c_0 = 0.0239, p_0 = 1$

对照模型检验等级参照表(表 2),可以得到模型的相对误差,关联度,均方差比值,小概率误差均达到一级(好)的标准,因此模型的模拟精度高,可以利用模型进行分析,下面我们利用模型反映的结果研究酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响。

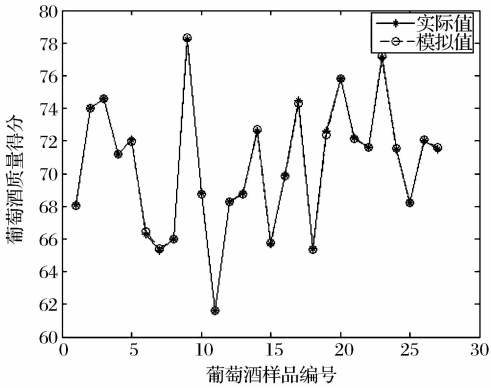


图 5 模型模拟结果对比图



图 6 标准化数据的回归系数图

为了更直观、迅速地观察各个自变量(即葡萄和葡萄酒的各项理化指标)对因变量(葡萄酒的质量)的边际作用,我们绘制回归系数图 6,图 6 给出了针对标准化数据的回归系数图。从图 6 中我们可以观察到,在影响葡萄酒的质量上有 23 个葡萄和葡萄酒的理化指标起到了非常重要的作用,具体指标见表 4。

表 4 对葡萄酒的质量影响显著的理化指标

葡萄的理化指标 (17 个指标)	蛋白质(x_2)、VC 含量(x_3)、柠檬酸(x_7)、褐变度(x_9),总酚(x_{11})、单宁(x_{12})、葡萄总黄酮(x_{13})、白藜芦醇(x_{14})、黄酮醇(x_{15})、总糖(x_{16})、可溶性固形物(x_{18})、固酸比(x_{21})、干物质含量(x_{22})、果皮质量(x_{27})、果皮颜色 L * (x_{28})、果皮颜色 a * (十红;一绿)(x_{29})、芳香性指标(x_{31})
葡萄酒的理化指标 (6 个指标)	单宁(x_{33})、总酚(x_{34})、白藜芦醇(x_{36})、DPPH 半抑制体积(x_{37})、色泽 a * (D65)(x_{39})、芳香性指标(x_{42})

其中对葡萄酒的质量影响非常显著的指标是葡萄的 VC 含量(影响系数为 0.278 3)、褐变度(影响系数为-0.378 8)、可溶性固形物(影响系数为 0.292 4)、干物质含量(影响系数为 0.361 4)、果皮质量(影响系数为-0.228 7)、果皮颜色 L * (影响系数为 0.221 1);对葡萄酒的质量影响非常显著的葡萄酒的理化指标有总酚(影响系数为 0.243 4)和 DPPH 半抑制体积(影响系数为 0.334 5)。

进一步分析,从图 4 和模型的计算结果我们可以看到,理想葡萄酒的理化指标分为正向指标和负向指

标,在影响葡萄酒的质量非常重要的 23 个指标中有 13 个是正向指标,分别是葡萄的理化指标中的:VC 含量、总酚、单宁、葡萄总黄酮、白藜芦醇、总糖、可溶性固形物、干物质含量、果皮颜色 L * ;葡萄酒的理化指标中的总酚、DPPH 半抑制体积、色泽 a * (D65)、芳香性指标。有 10 个负向指标,分别是,葡萄的理化指标中的蛋白质、柠檬酸、褐变度、黄酮醇、固酸比、果皮质量、果皮颜色 a * (十红;一绿)和芳香性指标;葡萄酒的理化指标中的单宁、白藜芦醇。

表 5 影响葡萄酒质量显著的理化指标分类表

影响质量显著 的正向指标	葡萄	VC 含量、总酚、单宁、葡萄总黄酮、白藜芦醇、总糖、可溶性固形物、干物质含量、果皮颜色 L*
	葡萄酒	总酚、DPPH 半抑制体积、色泽 a* (D65)、芳香性指标
影响质量显著 的负向指标	葡萄	蛋白质、柠檬酸、褐变度、黄酮醇、固酸比、果皮质量、果皮颜色 a* (+红;-绿)、芳香性指标
	葡萄酒	单宁、白藜芦醇

3 结语

综合上面的分析,酿酒葡萄和葡萄酒的理化指标对葡萄酒的质量有显著的影响,从模型的模拟精度可以看出,模型的三种检验方式均能达到一级(好)的标准,因此根据葡萄和葡萄酒的理化指标利用模型对葡萄酒的质量进行预测将具有很好的可信度,也就是说如果不考虑酿酒工艺、酿酒的水平等外界因素,利用葡萄和葡萄酒的理化指标进行葡萄酒质量的评价是可行的。但事实上,影响葡萄酒的质量的因素有许多,如原产地的生态条件、葡萄品种以及人所采用的栽培、采收、酿造工艺等,都会影响葡萄酒的质量和风格,而这些因素是葡萄或葡萄酒中的理化指标所反映不到的。因此,要想利用葡萄或葡萄酒中的理化指标来准确的评价葡萄酒的质量我们还需做更为深入的工作。

参考文献

[1] PAULO CORTEZ, ANTÓNIO CERDEIRA, FERNANDO ALMEIDA, TELMO MATOS, JOSé REIS. Modeling wine preferences by data mining from physicochemical properties [J]. Decision Support Systems,2009,47(4):547—553.

[2] ISABEL M MORENO, VALERIO GUTIERREZ, ANA M CAMEÁN, ARTURO HARDISSON. Differentiation of two

Canary DO red wines according to their metal content from inductively coupled plasma optical emission spectrometry and graphite furnace atomic absorption spectrometry by using Probabilistic Neural Networks[J]. Talanta,2007,72(1):263—268.

[3] 刘延玲. 新的 Hopfield 神经网络分类器在葡萄酒质量评价中的应用[J]. 价值工程,2012,31(2):181—182.

[4] 吴昊潼,薛超,刘海明. 应用数学建模对葡萄酒评价问题的探究[J]. 数学学习与研究,2013(13):111—112.

[5] 李运,李记明,姜忠军. 统计分析在葡萄酒质量评价中的应用[J]. 酿酒科技,2009(4):79—82.

[6] 王金甲,尹涛,李静,洪文学,马崇霄. 基于物理化学性质的葡萄酒质量的可视化评价研究[J]. 燕山大学学报,2010,34(2):133—137.

[7] 王文静. 感官评价在葡萄酒研究中的应用[J]. 酿酒,2007,34(4):57—59.

[8] 张贝贝,李珊珊,杨珊. 基于多元统计的葡萄酒品质因素分析[J]. 中央民族大学学报:自然科学版,2013(22):133—139.

[9] 傅小伟. IRMS 和 SNIF-NMR 化学方法在葡萄酒质量鉴别中的应用[J]. 现代农业科技,2010(24):338—339.

[10] 2012 年高教社杯全国大学生数学建模竞赛 A 题[EB/OL]. <http://www.mcm.edu.cn/>.

[11] 罗批,郭继昌,李镔,滕建辅. 基于偏最小二乘回归建模的探讨[J]. 天津大学学报:自然科学与工程技术版,2002,35(6):783—786.

The Quality Evaluation of Wine Based on PLS Model

XI Qian, WAN Bing-yue

(College of Taizhou,Nanjing Normal University,Taizhou Jiangsu 225300,China)

Abstract: This paper established a model of partial least squares regression analysis (PLS), the use of the 2012 National Undergraduate Mathematical Contest in Modeling A title given data, including selected aromatic substances, including all of the physical and chemical indicators of wine grapes and wine as independent variables to wine quality as the dependent variable to model. Extracts calculated by the independent variables and the dependent variable contribution rate can be explained by the independent variables selected ratio was 86.28% of the 13 ingredients right, get the PLS regression equation with all the physical and chemical quality of the wine between indicators. The use of standardized indicators variable regression equation to obtain a greater impact on the quality of the wine has 23 physical and chemical indicators, the dry matter content of which 13 positive indicators, 10 negative indicators, positive indicators are the biggest factor affecting grapes (impact factor of 0.3614), negative indicator is the biggest factor affecting the degree of browning grapes (impact factor is -0.3788). Matlab programming for the accuracy of the model was tested. Test results show that if you do not consider the technical, environmental factors such as brewing process, the use of physical and chemical indicators have high credibility based on partial least squares regression model to predict the quality of the wine establishment.

Key words: PLS model; wine quality; physical and chemical indicators; accuracy test