****

课程设计报告书

**大数据应用课程设计**

**学院 数学学院**

**专业 数学与应用数学（统计学）**

**学生姓名 邓睿**

**学生学号 202130320342**

**指导教师 钟友良 张威**

**课程编号 040102331-1**

**课程学分 2**

**起始日期 2023.12.24**

|  |  |
| --- | --- |
| 教  师  评  语 | 教师签名：  日期： |
| 成  绩  评  定 |  |
| 备  注 |  |

**目录**

[大数据课程设计 1](#_Toc158040182)

[一、选题背景 1](#_Toc158040183)

[二、复现证明过程 1](#_Toc158040184)

[2.1最佳线性无偏估计量的性质 2](#_Toc158040185)

[2.2岭回归 2](#_Toc158040186)

[2.3岭迹 3](#_Toc158040187)

[2.4岭回归的均方误差 4](#_Toc158040188)

[2.5岭回归的通用形式 6](#_Toc158040189)

[2.6选择一个更好的估计值 6](#_Toc158040190)

[三、应用实例 7](#_Toc158040191)

[四、课程设计总结 13](#_Toc158040192)

[附录 14](#_Toc158040193)

# 大数据课程设计

## 一、选题背景

岭回归是一种专用于共线性数据分析的有偏估计回归方法，实质上是一种改良的最小二乘估计法，通过放弃最小二乘法的无偏性，以损失部分信息、降低精度为代价获得回归系数更为符合实际、更可靠的回归方法，对病态数据的拟合要强于最小二乘法。岭回归的优点在于其能够提供更加稳定和准确的估计结果，特别是在处理共线性数据时。它通过正则化项对系数的大小施加惩罚，从而避免了过拟合和模型复杂度过高的问题。这使得岭回归在许多实际应用中成为一种受欢迎的方法，例如在生物信息学、金融分析和自然语言处理等领域。岭回归是一种有偏估计方法，即它会导致估计的系数比最小二乘法估计的系数更小。这是由于正则化项的存在，它对系数的大小施加了惩罚。因此，在应用岭回归时需要权衡偏差和方差之间的平衡，选择合适的正则化参数以获得最佳的估计结果。

Hoerl和Kennard在1970年的论文《Reige Regression: Biased Estimation for Nonorthogonal Problems》中提出了岭回归的理论和方法，并证明了其在非正交问题中的优势。

Hoerl和Kennard的论文中提出了岭回归的数学推导和理论证明。他们证明了岭回归的估计结果具有偏差，但相对于普通最小二乘估计，岭回归能够显著降低参数估计的方差。这使得岭回归在高维数据和共线性问题中具有很好的性能。在数值实验方面，Hoerl和Kennard通过模拟和实际数据集进行了实验验证。他们比较了岭回归和普通最小二乘回归在不同条件下的性能，并展示了岭回归在减小参数估计方差和提高预测性能方面的优势。

本文将复现论文中的数学推导以及数值实验。

## 二、复现证明过程

考虑多元回归的标准型，其中X为秩为n×p的矩阵，为未知的p×1的向量，满足高斯-马尔可夫假设。

### 2.1最佳线性无偏估计量的性质

当随机向量服从正态分布时，使用最小方差或最大似然估计的无偏线性估计可以得到的估计量

残差平方和的最小值为

此处仅关注不为单位阵的情况。为说明该条件对结果的影响，考虑方差-协方差矩阵和它与其期望值的距离。

这些相关属性显示了当从单位矩阵变为病态矩阵时的不确定性，其特征值用，则的期望和协方差可表示成

平均值和方差的下界分别为和。如果有一个或者多个小的特征值，则值会很大。最小二乘法估计（1）存在数学优化技术所给的点估计的不足之处；估计过程没有内置一种方法来描述解（1）对优化准则（2）的敏感性。

### 2.2岭回归

为了控制最小二乘法估计所伴随的膨胀和整体不稳定，可以使用一种被称为岭回归的技术。

岭估计与普通估计之间的关系如下：

其中W、Z的特征值分别为：,

Z与W之间的关系如下式

此关系式可由Z的表达式证出。

由定义，Z为对称正定矩阵，因此有

因为为最大的特征值，由式子（11）和Z的特征值可以得当k趋于∞时Z趋于0。

的残差平方和为

这个表达式表明，是总平方和减去“回归”平方和，其中“回归”平方和是依赖于的平方长度的修改后的平方和。

### 2.3岭迹

在进行统计分析和回归建模时，特别是当数据存在多重共线性或其他复杂关系时，如何理解和处理因子空间（即自变量空间）的结构和敏感性。传统的回归分析方法可能不足以揭示这些因素之间的复杂关系，因此需要其他技术，如岭回归和主成分分析，来提供更深入的理解和更准确的预测。

令B为的估计值，残差平方和可以写成以下形式

是中心在的超椭球体的表面，其中是的普通最小二乘估计值。的值是最小值加上B-中的二次形式的值。存在一个连续的B0值，它们将满足关系 =+，其中>0是一个固定的增量。由前面论证，如果有一个小的特征值，那么从到的距离平均上会很大。特别是，的条件越差，的长度越有可能过长。另一方面，条件越差，离越远而不显著增加残差平方和。如果远离最小平方和点，移动的方向应该是缩短回归向量的长度。岭迹可以表示为穿过平方和表面的路径，对于固定的，选择一个B的值，该值具有最短的长度，即求的最小值，运用拉格朗日乘子法求解。

对B求偏导为0，求解可得岭估计值。此时残差平方和为。回归向量B的平方长度固定在时，B去时平方和值达到最小。也就是说，B去值为时函数F取最小值。

，其似然函数为

通过(14)在中，表明残差平方和的增加等价于似然函数值的减小。因此，等概率的等高线也位于以为中心的超椭球体的表面上。因此，岭迹可以解释为通过似然空间的一条路径，并且问题转变成为什么这条特定的路径可能特别重要。理由与平方和相同。虽然长的向量与较短的向量给出相同的似然值，但它们并不总是具有相同的物理意义。暗示了对的可能值的约束，这在引言中给出的通用线性模型的公式中没有明确说明。这一点将在后续章节中进一步讨论。

### 2.4岭回归的均方误差

首先求的期望

其中表示的平方距离，当k=0时取0，即Z=I为单位阵。表示参数估计的总方差和。的协方差阵为

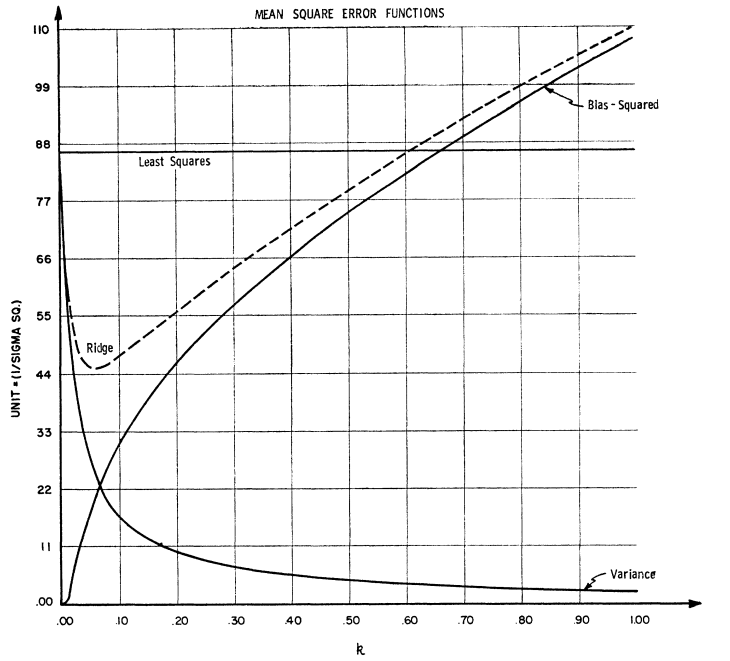
图1以定性的形式展示了方差、平方偏差和参数k之间的关系。随着k的增加，总方差会减小，而平方偏差会随着k的增加而增加。由虚线表示，它是和的和，即是，存在一种可能性，即存在某些k值（可容许值），使得的均方误差小于的均方误差。函数是关于k的单调递减函数，而是单调递增的。两个函数在原点附近的导数为

图 1 方差、平方偏差和参数k之间的关系

可见，的导数是负数，当正交时,k趋向于无穷大，它接近;当X'X的病态条件恶化且时，它接近-∞。另一方面，随着k从大于0的方向趋向于0，(23)表明在原点处是平坦的且为零。这些性质导致了一个结论：有可能移动到k > 0，稍微增加偏差，并大大减少方差，从而改善估计和预测的均方误差。

的性质表明它会经历一个最小值。由于在k趋向于无穷大时趋近于，这个最小值将随着的增加而向k = 0移动。由于是未知回归向量的平方长度，似乎不可能选择一个k值小于0，从而无法为确定一个上界，而实现更小的均方误差。另一方面，显然，在实践中，不会变得无穷大，应该能够找到一个或多个k值，使比更接近。

### 2.5岭回归的通用形式

可以将引言中定义的线性回归问题简化为规范形式，其中矩阵为对角矩阵。特别地，存在一个正交变换P，使得，其中是的特征值矩阵。这个结论基于线性代数中的一些基本性质，特别是矩阵的特征值和特征向量。令,则,其中,，。岭估计为，其中。通过可以得到的最优值为。

### 2.6选择一个更好的估计值

在前面已经证明了当不具有均匀的谱特征值时，回归向量的普通最小二乘估计值存在一些不足之处。通过增加的对角线上的小正数，引入了一类有偏估计器，可以描绘解对的敏感性，并形成获得具有较小均方误差的的估计值的基础。在检查的属性时，可以证明其使用相当于对的各个坐标或其平方长度做出某些有界性假设。最小二乘估计中无偏性的逻辑的替代方案是先确保的均方误差有界，而不做出任何有界性假设。如果可以对做出具体的数学假设，那么就有可能将这些假设约束在估计过程中。

使用时所做的固有有界性假设表明，无法构建一个明确的、自动的估计程序来生成一个点估计值（k的一个单一值或每个的特定值），就像可以构建来生成一样。然而，这并不妨碍其使用，因为在任何给定的数据集上，选择一个比更好的并不困难。事实上，任何一组适合使用线性回归分析的数据都隐含着对估计值的可能值的限制，这些限制与数据生成器的已知属性一致。然而，很难明确说明这些限制；尤其难以进行数学上的明确说明。对于估计分布的均值μ的问题，一组数据隐含着对值的限制，这些值可以作为逻辑上的候选生成器。当然，在线性回归中，问题要复杂得多；可能性的数量如此之大。首先，涉及的参数数量。有十个到二十个回归系数并不罕见。而且还要考虑它们的符号。然后是和不同的因子相关性以及它们之间的关联方式。然而，在最后的分析中，这些许多不同的影响可以综合起来，以评估估计值是否与数据和数据生成器的属性一致。

## 三、应用实例

Biopsy数据集是一个在医学领域中广泛使用的数据集，专门用于乳腺癌的诊断和研究。这个数据集来源于威斯康星州麦迪逊大学医院（University of Wisconsin Hospital, Madison），并且经常被用作机器学习和数据分析的教学或研究案例。这个数据集包含了699个乳腺癌患者的活检样本信息。每个样本都包含了一系列与患者乳腺肿瘤相关的特征信息，这些信息是通过细针穿刺（FNA）技术收集并进行分析的。细针穿刺是一种常用的活检方法，医生通过这种方法可以获取少量的组织样本，然后对这些样本进行病理学检查，以确定肿瘤是良性还是恶性。在这699个样本中，每个样本都有11个变量，其中一个是病人的ID（不用于分析），一个是肿瘤的诊断结果（良性或恶性），剩下的9个变量是与肿瘤特征相关的检验指标。这些检验指标包括肿块的厚度、细胞的大小均匀性、细胞形状的均匀性、边缘的粘附性、单个上皮细胞的大小、裸核的数量、核仁的分裂程度等。每个病人在这些特征上都有一个得分，这个得分是一个介于1到10之间的数值，其中1表示该特征更接近良性，10表示该特征更接近恶性。这个数据集的主要用途是建立一个能够预测乳腺肿瘤是良性还是恶性的模型。通过分析和学习这些样本的特征和诊断结果之间的关系，我们可以训练出一个机器学习模型，这个模型可以根据新患者的检验指标得分来预测其肿瘤是良性还是恶性。这对于乳腺癌的早期诊断和治疗具有重要意义，可以帮助医生更准确地判断患者的病情，并制定相应的治疗方案。

首先，为了探究多个自变量与因变量之间的关系，构建多元线性回归模型。模型构建的过程中，我们使用了lm()函数。这个函数在R语言中是进行线性回归分析的标准函数，它允许我们根据给定的自变量和因变量来拟合一个线性模型。该函数基于最小二乘法原理，通过最小化预测值与实际值之间的平方误差来估计模型的参数。模型中因变量是class，而数据集中的其他所有变量被选为自变量。

在多元线性回归分析中，自变量之间的多重共线性是一个重要的问题。当自变量之间存在高度相关性时，它们可能会相互影响，导致模型的不稳定和预测能力的下降。为了检查模型的多重共线性，我们使用了car包中的vif()函数。vif()函数计算了每个自变量的方差膨胀因子（Variance Inflation Factor，简称VIF）。VIF是一种评估多重共线性的度量，其值越大，表明自变量之间的相关性越高，多重共线性问题越严重。通常情况下，如果VIF值大于5或10，则认为存在多重共线性问题。对数据集构造的多元线性模型的多重共线性的检验如下表1。

表 1 多元线性模型的多重共线性的检验

|  |  |  |  |
| --- | --- | --- | --- |
|  | GVIF | Df | GVIF^(1/(2\*Df)) |
| thick | 1.02486E+15 | 629 | 1.02785574 |
| u.size | 44.54261654 | 1 | 6.674025512 |
| u.shape | 264.4384901 | 1 | 16.26156481 |
| adhsn | 260.3471206 | 1 | 16.13527566 |
| s.size | 257.1090922 | 1 | 16.03462167 |
| nucl | 169.7704954 | 1 | 13.02960074 |
| chrom | 95.37658566 | 1 | 9.766093675 |
| n.nuc | 164.3110064 | 1 | 12.81838548 |
| mit | 393.5257097 | 1 | 19.83748244 |
| `NA` | 271.4915493 | 1 | 16.47700062 |

GVIF (Generalized Variance Inflation Factor)：方差膨胀因子，用于衡量多变量共线性（Multicollinearity）的程度。如果GVIF值很大，则说明存在严重的多变量共线性问题。Df (Degrees of Freedom)：自由度，在统计模型中表示可用于估计的参数数量。GVIF^(1/(2\*Df))：这是一个与GVIF相关的分数位数，它可以帮助我们更好地理解GVIF值的大小和其对模型的影响。在给定的代码中，我们通过vif()函数得到了每个自变量的方差膨胀因子（GVIF）值。这些值提供了关于自变量之间多重共线性的重要信息。

每一列代表一个统计量，如"thick"、"u.size"等，它们是模型中的自变量。每个自变量都有对应的GVIF值、自由度和GVIF的分数位数。以"thick"为例，其GVIF值为1.024860e+15，这是一个非常高的值，表明该变量与其他变量之间可能存在严重的多变量共线性问题。这意味着"thick"变量与其他变量之间的相关性非常高，导致模型的不稳定性增加。同样地，其他的统计量如"u.size"、"u.shape"等也显示出了相对较高的GVIF值。这意味着这些变量之间也存在多变量共线性问题。GVIF的分数位数提供了额外的信息，用于评估每个自变量的预测能力受到多变量共线性的影响程度。例如，"thick"的分数位数为1.027856，这表明该变量的预测能力可能会受到多变量共线性的较大影响。由于多个统计量显示出了高的GVIF值，这表明存在多变量共线性问题。在这种情况下，我们不能简单地使用线性模型进行预测，因为模型的不稳定性可能导致预测结果不准确。为了处理多重共线性问题，我们可以采用其他回归方法，如岭回归（Ridge Regression），即前面进行理论证明的方法。

在处理具有多重共线性的数据时，我们首先决定采用岭回归方法来建立模型。将模型的输出结果保存在名为“ridge”的对象中。这里需要注意的是，glmnet包在计算λ值之前，已经对输入数据进行了标准化处理，即所有特征值都被缩放到了同一尺度上，以确保模型的一致性和准确性。此外，为了适应我们的二元结果，我们需要将响应变量的分布设定为“二项式”。同时，为了进行岭回归，我们需要将alpha参数设定为0。在glmnet包中，默认的迭代次数是100次，如果连续的两个λ值的百分比偏差改善不明显，那么算法会在达到100次迭代之前停止。这表明算法已经收敛到了最优解，不需要再进行更多的计算。利用print()函数输出非零回归系数的值、解释的百分比偏差及相应的λ值，部分结果见表2，完整数据见附录4。

表 2 部分迭代次数下得到的结果

|  |  |  |  |
| --- | --- | --- | --- |
| 迭代次数 | Df | %Dev | Lambda |
| 90 | 9 | 79.73 | 0.1002 |
| 91 | 9 | 80.27 | 0.09127 |
| 92 | 9 | 80.79 | 0.08316 |
| 93 | 9 | 81.27 | 0.07577 |
| 94 | 9 | 81.72 | 0.06904 |
| 95 | 9 | 82.14 | 0.06291 |
| 96 | 9 | 82.54 | 0.05732 |
| 97 | 9 | 82.91 | 0.05223 |
| 98 | 9 | 83.26 | 0.04759 |
| 99 | 9 | 83.59 | 0.04336 |
| 100 | 9 | 83.89 | 0.03951 |

以第100行结果为例解释输出结果，第100行结果为9，8.389e-01，0.03951。其中第一个数字9模型中包含的特征数为9，这是一个在岭回归中保持不变的数字。解释偏差的百分比为0.8389，意味着模型能够解释约83.89%的数据变异，显示出较好的拟合效果。调协系数（或称正则化参数）在岭回归中也扮演着重要角色。它控制着回归系数的收缩程度，有助于防止过拟合。这个模型调协系数为0.03951，为了简化讨论和计算过程，我们通常会将测试集的λ值设置为一个固定的数值。我们选择了0.05作为测试集的λ值。

除了通过数值分析岭回归模型的参数外，我们还可以通过图形来直观地观察回归系数是如何随λ值的变化而变化的（图2）。绘制回归系数与λ值的散点图或曲线图是一种常用的可视化方法。通过图形展示，我们可以清晰地观察到回归系数随着λ值的增加而逐渐收缩的趋势。随着λ值的减小，回归系数的绝对值会增加。这是因为当正则化强度较弱时，模型更倾向于选择那些对预测有显著影响的特征，从而赋予它们更大的权重。这可能导致某些特征的系数变得更大，以弥补其他特征被压缩的情况；而随着λ值的增大，回归系数逐渐减小，模型的复杂度降低，有助于防止过拟合并提高模型的泛化能力。此外，通过图形展示回归系数与λ值的变化关系，还可以帮助我们更好地理解模型的收敛情况。如果回归系数随λ值的增加而迅速收敛到一个稳定值，说明模型在较小的λ值下已经收敛到了一个较好的解。此图显示当λ下降时，压缩参数减小，但绝对系数增加。

为检查岭回归对数据的拟合程度，自定义误分类误差函数(misClassError)。

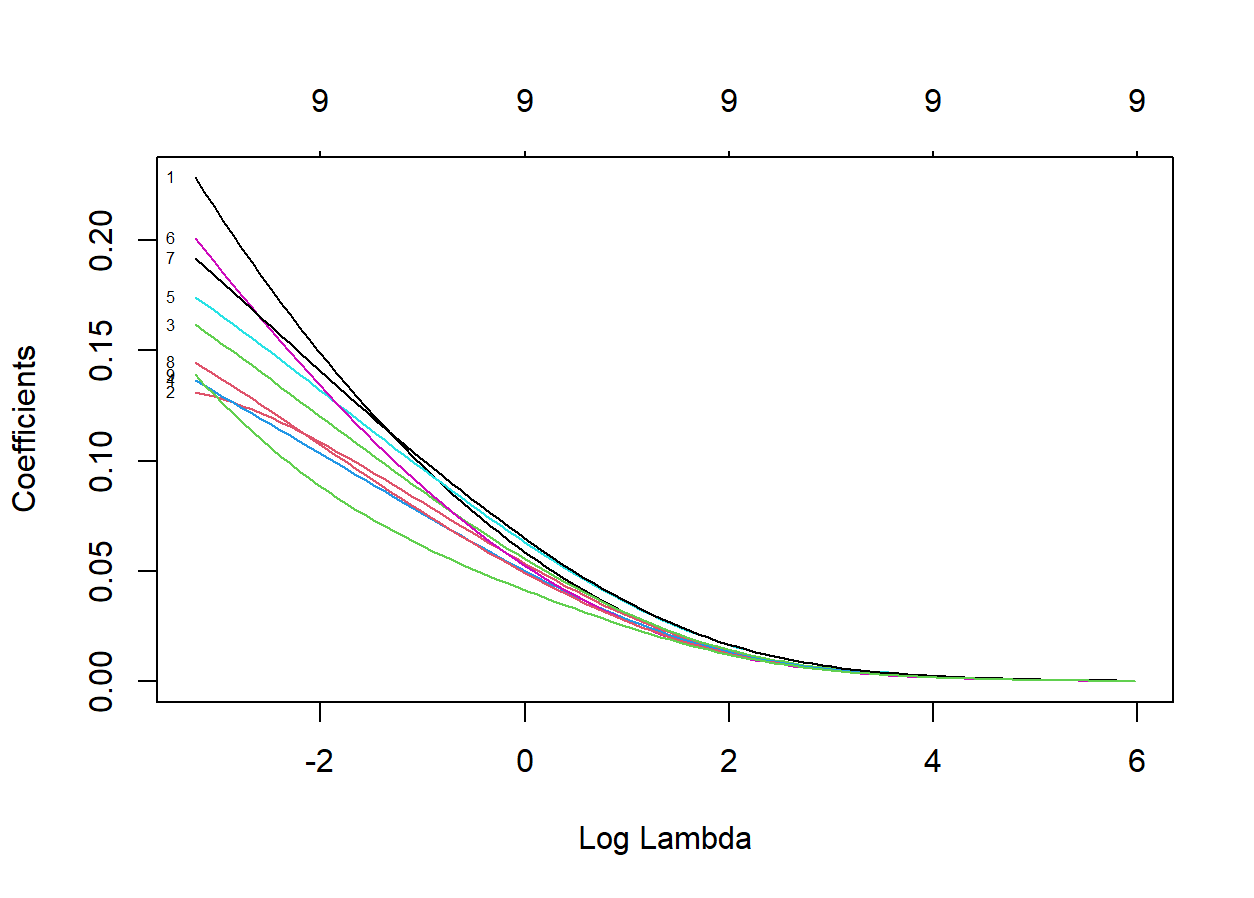
这个函数是为了计算二元分类模型的误分类误差而定义的。它基于真正例率（TPR）和假正例率（FPR）来计算误分类误差。函数内部，通过条件判断和计数，计算了真正例、假正例、真负例和假负例的数量。使用这些计数，进一步计算了TPR和FPR。最后，使用1 - TPR + FPR的公式计算了误分类误差，运行得到误分类误差值为0.044，即在测试数据中大约有4.4%的样本被错误分类。这个值越小，表示模型的分类性能越好，误分类的可能性越低。

图 2 回归系数与λ值的关系

ROC曲线是一种用于评估二元分类模型性能的重要工具。它通过绘制真正例率（TPR）和假正例率（FPR）之间的关系，展示了模型在不同阈值下的性能表现。ROC曲线下的面积（AUC）可以用来评估模型的性能，AUC越接近于1表示模型性能越好，而AUC越接近于0.5表示模型性能较差。ROC曲线的起点位于左下角（FPR=0，TPR=0），表示当模型判定所有样本为负例时，没有任何真正例被识别出来；曲线的终点位于右上角（FPR=1，TPR=1），表示当模型判定所有样本为正例时，所有真正例都被识别出来。ROC曲线随着阈值的改变而发生变化。如果曲线非常接近左上角，说明模型在较宽松的阈值下就能较好地识别真正例，但同时也会产生较多的假正例；如果曲线非常接近右下角，说明模型在较严格的阈值下才能较好地识别真正例，但同时也会漏掉较多的真正例。理想的ROC曲线应该是尽量向上翘起，尽可能接近右上角。AUC是ROC曲线下的面积，其值介于0.5和1之间。AUC越接近于1，表示模型性能越好；AUC越接近于0.5，表示模型性能越差。一般来说，AUC值大于0.75可以认为模型性能较好，AUC值在0.6到0.75之间表示模型性能一般，AUC值小于0.5则认为模型性能较差。

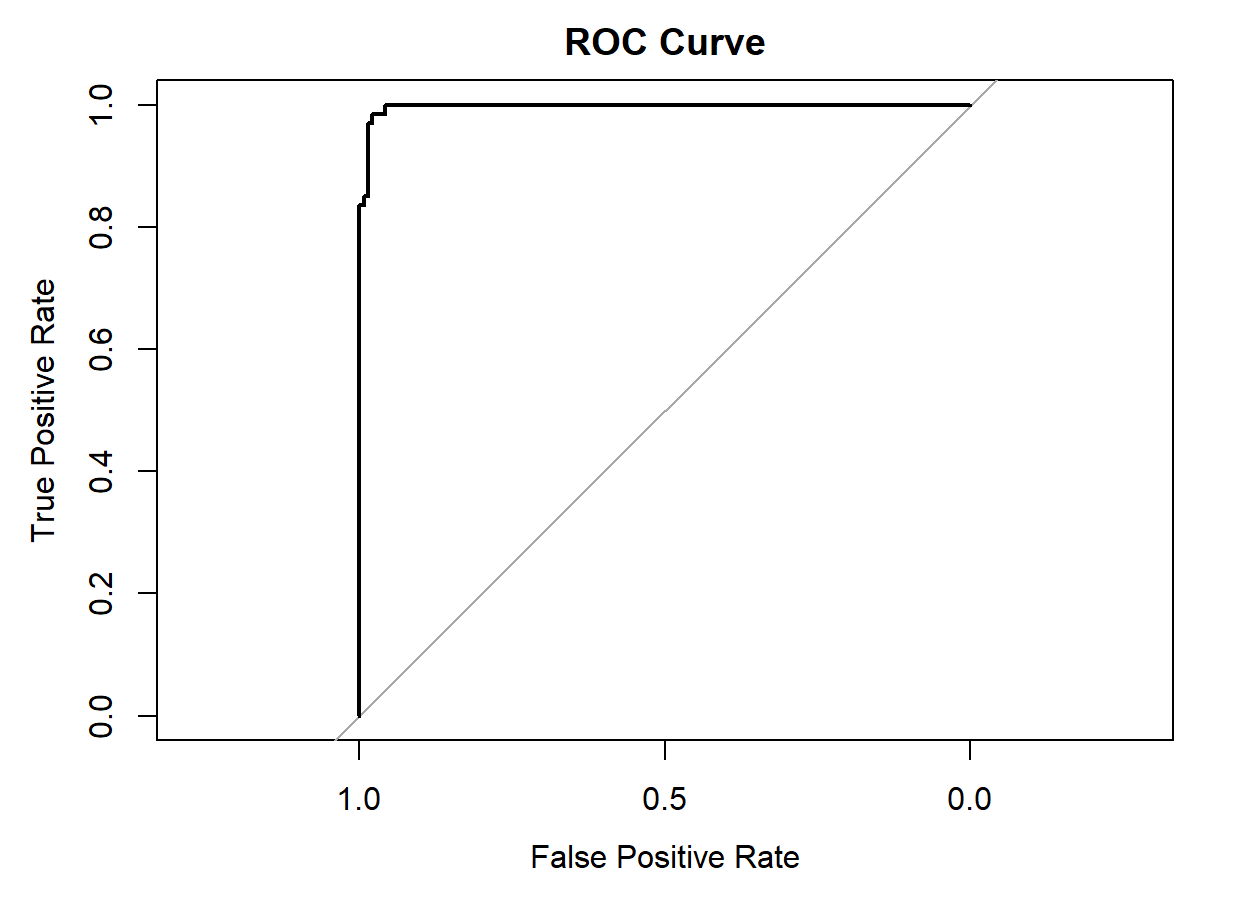
模型的ROC图如图3，由图可见曲线下方面积接近1，即AUC值接近1，故模型性能较好。

图 3 ROC曲线图

## 四、课程设计总结

本次课程设计题目是对课本的参考文献中的证明方法或者数值实验进行复现，我选择的是回归方法中的岭回归。主要用到了机器学习以及高等代数的相关知识，这加深了我对于课堂学习的印象、加强了我对本方法的掌握以及熟练程度。复现证明的过程中，在理解证明上遇到了一些困难，通过查找资料很好的解决了。在应用实例中，由于对R语言编程的不熟悉，编写出正确的代码也遇到了些困难，通过网上搜寻资料以及询问同学最终得到解决。

本次课程设计让我对于机器学习应用于数据挖掘产生了兴趣，这加深了我对于课本上知识的理解。也在此感谢钟老师和张老师关于课本上知识的讲解，帮助我顺利地完成了本次课程设计。

## 附录

|  |
| --- |
| **附录1：**参考文献 |
| Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970. |

|  |
| --- |
| **附录2：**对数据集构建多元线性模型 |
| 介绍：用lm（）构建多元线性模型并检验多重共线性 |
| library(MASS)  library(openxlsx)  # 加载数据集  data(biopsy)  # 重命名变量  names(biopsy) <- c("thick", "u.size", "u.shape", "adhsn", "s.size", "nucl", "chrom", "n.nuc", "mit", "class")  # 构建多元线性模型  model <- lm(class ~ ., data = biopsy)  # 检查多重共线性  vif\_values <- car::vif(model)  vif\_value <- data.frame(vif\_values)  write.xlsx(vif\_value, "data.xlsx", sheetName = "Sheet1")  print(vif\_values) |

|  |
| --- |
| **附录3：**对数据集构建岭回归模型 |
| 介绍：用glmnet（）构建岭回归模型并对模型进行检验 |
| library(glmnet)  library(MASS)  library(openxlsx)  biopsy$ID <- NULL  names(biopsy) <- c("thick", "u.size", "u.shape", "adhsn", "s.size", "nucl", "chrom", "n.nuc", "mit", "class")  biopsy.v2 <- na.omit(biopsy)  set.seed(123) #random number generator  ind <- sample(2, nrow(biopsy.v2), replace = TRUE, prob = c(0.7, 0.3))  train <- biopsy.v2[ind == 1, ] #the training data set  test <- biopsy.v2[ind == 2, ] #the test data set  x <- as.matrix(train[, 1:9])  y <- train[, 10]  ridge <- glmnet(x, y, family = "binomial", alpha = 0)  # 将结果转换为普通矩阵  ridge\_coef <- as.matrix(coef(ridge))  # 将结果保存到Excel文件  write.xlsx(data.frame(ridge\_coef), "ridge\_results.xlsx")  plot(ridge, xvar = "lambda", label = TRUE)  ridge.coef <- predict(ridge, s = 0.05, type = "coefficients")  newx <- as.matrix(test[, 1:9])  ridge.y <- predict(ridge, newx = newx, type = "response", s = 0.05)  # 计算实际值的向量  actuals <- ifelse(test$class == "malignant", 1, 0)  # 计算误分类误差  misClassError <- function(actual, predicted) {  # 计算真正例率（TPR）和假正例率（FPR）  truePositives <- sum(actual & (predicted > 0.5))  falsePositives <- sum((!actual) & (predicted > 0.5))  trueNegatives <- sum((!actual) & (predicted <= 0.5))  falseNegatives <- sum(actual & (predicted <= 0.5))    # 计算TPR和FPR  tpr <- truePositives / (truePositives + falseNegatives)  fpr <- falsePositives / (falsePositives + trueNegatives)    # 计算误分类误差  misClassError <- 1 - tpr + fpr    return(misClassError)  }  # 计算误分类误差  misClassErrorValue <- misClassError(actuals, ridge.y)  library(pROC)  # 计算ROC曲线数据  roc\_data <- roc(actuals, ridge.y)  # 绘制ROC曲线  plot(roc\_data, main = "ROC Curve", xlab = "False Positive Rate", ylab = "True Positive Rate") |

|  |
| --- |
| **附录4：**迭代结果 |
| 介绍：岭回归模型迭代次数及λ取值 |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 迭代次数 | Df | %Dev | Lambda | 迭代次数 | Df | %Dev | Lambda | | 1 | 9 | 0 | 395.1 | 51 | 9 | 28.57 | 3.771 | | 2 | 9 | 0.47 | 360 | 52 | 9 | 30.25 | 3.436 | | 3 | 9 | 0.52 | 328 | 53 | 9 | 31.95 | 3.131 | | 4 | 9 | 0.57 | 298.9 | 54 | 9 | 33.69 | 2.853 | | 5 | 9 | 0.62 | 272.3 | 55 | 9 | 35.44 | 2.599 | | 6 | 9 | 0.68 | 248.1 | 56 | 9 | 37.21 | 2.368 | | 7 | 9 | 0.75 | 226.1 | 57 | 9 | 38.98 | 2.158 | | 8 | 9 | 0.82 | 206 | 58 | 9 | 40.76 | 1.966 | | 9 | 9 | 0.9 | 187.7 | 59 | 9 | 42.53 | 1.792 | | 10 | 9 | 0.99 | 171 | 60 | 9 | 44.29 | 1.632 | | 11 | 9 | 1.08 | 155.8 | 61 | 9 | 46.04 | 1.487 | | 12 | 9 | 1.19 | 142 | 62 | 9 | 47.77 | 1.355 | | 13 | 9 | 1.3 | 129.4 | 63 | 9 | 49.47 | 1.235 | | 14 | 9 | 1.42 | 117.9 | 64 | 9 | 51.14 | 1.125 | | 15 | 9 | 1.56 | 107.4 | 65 | 9 | 52.79 | 1.025 | | 16 | 9 | 1.71 | 97.86 | 66 | 9 | 54.39 | 0.9341 | | 17 | 9 | 1.87 | 89.17 | 67 | 9 | 55.96 | 0.8511 | | 18 | 9 | 2.05 | 81.25 | 68 | 9 | 57.49 | 0.7755 | | 19 | 9 | 2.24 | 74.03 | 69 | 9 | 58.98 | 0.7066 | | 20 | 9 | 2.46 | 67.45 | 70 | 9 | 60.42 | 0.6439 | | 21 | 9 | 2.69 | 61.46 | 71 | 9 | 61.81 | 0.5867 | | 22 | 9 | 2.94 | 56 | 72 | 9 | 63.16 | 0.5345 | | 23 | 9 | 3.21 | 51.02 | 73 | 9 | 64.46 | 0.4871 | | 24 | 9 | 3.51 | 46.49 | 74 | 9 | 65.72 | 0.4438 | | 25 | 9 | 3.84 | 42.36 | 75 | 9 | 66.93 | 0.4044 | | 26 | 9 | 4.2 | 38.6 | 76 | 9 | 68.09 | 0.3684 | | 27 | 9 | 4.58 | 35.17 | 77 | 9 | 69.2 | 0.3357 | | 28 | 9 | 5 | 32.05 | 78 | 9 | 70.27 | 0.3059 | | 29 | 9 | 5.45 | 29.2 | 79 | 9 | 71.29 | 0.2787 | | 30 | 9 | 5.95 | 26.6 | 80 | 9 | 72.26 | 0.254 | | 31 | 9 | 6.48 | 24.24 | 81 | 9 | 73.19 | 0.2314 | | 32 | 9 | 7.06 | 22.09 | 82 | 9 | 74.08 | 0.2108 | | 33 | 9 | 7.68 | 20.13 | 83 | 9 | 74.92 | 0.1921 | | 34 | 9 | 8.35 | 18.34 | 84 | 9 | 75.72 | 0.175 | | 35 | 9 | 9.07 | 16.71 | 85 | 9 | 76.49 | 0.1595 | | 36 | 9 | 9.84 | 15.22 | 86 | 9 | 77.21 | 0.1453 | | 37 | 9 | 10.67 | 13.87 | 87 | 9 | 77.89 | 0.1324 | | 38 | 9 | 11.56 | 12.64 | 88 | 9 | 78.54 | 0.1206 | | 39 | 9 | 12.5 | 11.52 | 89 | 9 | 79.15 | 0.1099 | | 40 | 9 | 13.51 | 10.49 | 90 | 9 | 79.73 | 0.1002 | | 41 | 9 | 14.58 | 9.561 | 91 | 9 | 80.27 | 0.09127 | | 42 | 9 | 15.71 | 8.712 | 92 | 9 | 80.79 | 0.08316 | | 43 | 9 | 16.91 | 7.938 | 93 | 9 | 81.27 | 0.07577 | | 44 | 9 | 18.16 | 7.233 | 94 | 9 | 81.72 | 0.06904 | | 45 | 9 | 19.48 | 6.59 | 95 | 9 | 82.14 | 0.06291 | | 46 | 9 | 20.86 | 6.005 | 96 | 9 | 82.54 | 0.05732 | | 47 | 9 | 22.3 | 5.471 | 97 | 9 | 82.91 | 0.05223 | | 48 | 9 | 23.8 | 4.985 | 98 | 9 | 83.26 | 0.04759 | | 49 | 9 | 25.34 | 4.542 | 99 | 9 | 83.59 | 0.04336 | | 50 | 9 | 26.94 | 4.139 | 100 | 9 | 83.89 | 0.03951 | |