

[toc]

# 0 数据集探索

---

<https://zhuanlan.zhihu.com/p/141326006>

<https://www.cnblogs.com/mingke486/p/13814003.html>

[https://blog.csdn.net/m0\\_46275020/article/details/126306732?](https://blog.csdn.net/m0_46275020/article/details/126306732?spm=1001.2101.3001.6650.12&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=13)

[spm=1001.2101.3001.6650.12&utm\\_medium=distribute.pc\\_relevant.none-task-blog-](https://blog.csdn.net/m0_46275020/article/details/126306732?spm=1001.2101.3001.6650.12&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=13)

[2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-](https://blog.csdn.net/m0_46275020/article/details/126306732?spm=1001.2101.3001.6650.12&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=13)

[104659074.235%5Ev38%5Epc\\_relevant\\_sort\\_base3&depth\\_1-utm\\_source=distribute.pc\\_relevant.none-](https://blog.csdn.net/m0_46275020/article/details/126306732?spm=1001.2101.3001.6650.12&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=13)

[task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-](https://blog.csdn.net/m0_46275020/article/details/126306732?spm=1001.2101.3001.6650.12&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-12-126306732-blog-104659074.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=13)

## 1 Regression Shrinkage and Selection via the Lasso

---

<https://www.cnblogs.com/zzqingwenn/p/10874733.html>

<https://www.kaggle.com/code/galibce003/ridge-lasso-elastic-net-diabetes-dataset>

"Regression Shrinkage and Selection via the Lasso" 是由Robert Tibshirani于1996年发表的一篇经典论文，介绍了 Lasso（Least Absolute Shrinkage and Selection Operator）方法。以下是对该论文的概述：

论文背景：

- **问题陈述：** 论文关注的主要问题是，在回归问题中进行变量选择和模型收缩。传统的最小二乘法在高维数据集上容易过拟合，而 Lasso 方法通过对回归系数引入 L1 惩罚，使得一些系数变得精确地等于零，从而实现了变量的稀疏性和模型的简化。

Lasso 方法的引入：

- **L1 惩罚项：** 论文引入了 Lasso 方法，通过在最小二乘法的目标函数中添加 L1 惩罚项，得到了新的优化问题。这个问题的解决方案使得一些系数变为零，实现了变量选择。
- **Geometric Interpretation：** 论文通过几何解释，说明了 L1 惩罚在系数空间中形成了一个菱形，与坐标轴相交在某些系数为零的点，这直观地解释了 Lasso 方法的稀疏性质。

Lasso 方法的性质：

- **Biased Estimation：** 论文讨论了 Lasso 估计的偏倚（bias）性质，即估计值相对于真实值的偏离程度。Lasso 估计具有一定的偏倚，但在高维数据下仍然能够取得较好的性能。
- **Oracle Properties：** 论文介绍了 Lasso 方法的“Oracle Properties”，即在一些理想条件下，Lasso 能够达到最小可能的预测误差。

Lasso 方法的应用：

- **变量选择和模型收缩：** 论文强调了 Lasso 方法在变量选择和模型收缩中的优越性，特别是在高维数据集中。

影响和后续发展：

- 该论文的发表对于推动稀疏学习、统计学习理论和高维数据分析等领域产生了深远的影响。
- 后续有很多相关研究基于 Lasso 方法进行改进和拓展，涵盖了各种统计学和机器学习方法。

总体而言，这篇论文为 Lasso 方法的引入和理论基础奠定了基础，成为高维数据分析领域的经典之作。

## 2 Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties

---

### scad

SCAD是一种基于压缩感知(compressive sensing)理论的信号处理算法，全称为Smoothly Clipped Absolute Deviation。它与LASSO (Least Absolute Shrinkage and Selection Operator) 算法类似，都是用于处理高维数据的稀疏估计方法。SCAD算法能够在一定程度上克服LASSO算法的缺陷，特别是在选择变量时更具有稳定性和准确性。

SCAD算法是一种正则化方法，通过在损失函数中增加一个惩罚项来促进模型的稀疏性。在SCAD算法中，惩罚项是一个基于绝对值的函数，它可以使得系数的某些值变成0，从而实现特征选择。

与LASSO算法不同的是，SCAD算法在惩罚项中添加了一个阈值参数，该参数控制着稀疏性和偏差之间的权衡。当阈值参数取较小值时，SCAD算法退化为LASSO算法；当阈值参数取较大值时，SCAD算法近似于岭回归 (Ridge Regression) 算法。

总的来说，SCAD算法与LASSO算法相比，具有更好的性能和稳定性，但在实践中需要根据具体情况选择合适的惩罚项参数。

### 摘要：

变量选择是高维统计建模的基础，包括非参数回归。许多现有方法采用逐步选择程序，这可能在计算上昂贵，并且忽视变量选择过程中的随机误差。在这篇文章中，提出了一种惩罚似然方法来处理这些问题。所提出的方法同时选择变量并估计系数。因此，它们使我们能够构建估计参数的置信区间。所提出的方法与其他方法不同之处在于惩罚函数是对称的，在 $401^5$ 上是非凹的，并且在原点处具有奇异性，以产生稀疏解。此外，惩罚函数应该由常数限制以减少偏差，并满足某些条件以产生连续解。提出了一种用于优化惩罚似然函数的新算法。所提出的思想具有广泛的适用性。它们可以轻松应用于各种参数模型，如广义线性模型和鲁棒回归模型。它们还可以通过使用小波和样条轻松应用于非参数建模。建立了所提出的惩罚似然估计器的收敛速度。此外，通过适当选择正则化参数，我们表明所提出的估计量在变量选择方面的性能与 Oracle 过程一样好；换句话说，它们的工作效果与正确的子模型已知时一样好。我们的模拟表明，新提出的方法在与其他变量选择技术的比较中表现出色。此外，标准错误公式经过测试，足够准确用于实际应用。

### 结论

我们提出了一种基于惩罚似然法的变量选择方法。引入了一个罚函数族。建立了所提出的惩罚似然估计的收敛速度。在正确选择正则化参数的情况下，我们已经证明了所提出的估计器在可变选择方面的性能与oracle过程一样好。结果表明，该方法是有效的，并以良好的精度估计了标准误差。提出了一种最小化惩罚相似函

数的统一算法，该函数通常是凸函数和凹函数的和。我们的算法得到了统计理论的支持，因此给出了具有良好统计性质的估计量。与非常耗时的最佳子集方法相比，新提出的方法更快、更有效，并且具有较强的理论支持。他们选择变量

<https://blog.csdn.net/fangfanglovezhou/article/details/124859481>

"Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties" 是由Jianqing Fan和Runze Li于2001年发表的一篇论文。以下是对该论文的简要概述：

论文背景：

- **问题陈述：** 论文关注的主要问题是统计回归问题中的变量选择。传统的最小二乘法在高维数据集上容易过拟合，而一些非凹的惩罚项能够在一定条件下实现变量的稀疏性和模型的简化。

方法介绍：

- **非凹惩罚似然：** 论文引入了一种非凹惩罚似然方法，通过对最大似然估计的目标函数引入非凹的惩罚项，从而实现了变量选择。
- **Oracle Properties：** 论文讨论了所提出方法的“Oracle Properties”，即在一些理想条件下，该方法能够达到最小可能的预测误差。

惩罚项的性质：

- **惩罚项的非凹性：** 论文解释了非凹惩罚的性质，说明了它是如何促使模型参数的某些分量趋向于零，从而实现变量选择的。

算法：

- **求解算法：** 论文可能提供了用于求解带有非凹惩罚的优化问题的算法或优化策略。

应用与实验：

- 论文可能通过数值实验或实际应用来验证所提方法的性能，包括模拟数据和真实数据的结果。

影响和后续发展：

- 论文的提出可能对于推动统计学习理论和变量选择方法的发展产生了一定的影响。

总体而言，这篇论文探讨了一种利用非凹惩罚实现变量选择的方法，并考察了该方法的性质和性能。如需详细信息，请查阅原始论文。

## 3 自适应 Lasso | The Adaptive Lasso and Its Oracle Properties

---

[https://blog.csdn.net/Liang\\_DJ/article/details/86775160](https://blog.csdn.net/Liang_DJ/article/details/86775160)

"The Adaptive Lasso and Its Oracle Properties" 是由Zou和Hastie于2006年发表的论文。该论文主要介绍了自适应 Lasso 方法及其 Oracle 属性。下面是对该论文的简要概述：

## 摘要

套索是一种用于同时估计和变量选择的流行技术。已经证明，在某些条件下，套索变量选择是一致的。在这项工作中，我们推导了套索变量选择一致性的一个必要条件。因此，在某些情况下，套索对于变量选择是不一致的。然后，我们提出了套索的一种新版本，称为自适应套索，其中使用自适应权重对 L1 惩罚中的不同系数进行惩罚。我们表明自适应套索具有 Oracle 特性；也就是说，它的表现与真实的底层模型一样好。与套索类似，自适应套索被证明是近似极小化最大风险的最优选择。此外，通过相同的高效算法解决套索问题，自适应套索也可以得到解决。我们还讨论了在广义线性模型中自适应套索的扩展，并展示了在温和的正则条件下，Oracle 特性仍然成立。作为我们理论的副产品，非负 Garotte 被证明对于变量选择是一致的。

## 结论

在这篇文章中，我们提出了自适应套索用于同时估计和变量选择。我们已经表明，尽管套索变量选择在某些情况下可能不一致，但自适应套索通过利用自适应加权的  $\ell_1$  惩罚具有 Oracle 特性。自适应套索收缩还导致了近乎极小-最大化的估计器。由于高效的路径算法，自适应套索具有与套索相同的计算优势。我们的模拟结果显示，自适应套索在与其他稀疏建模技术的比较中表现出色。值得强调的是，Oracle 特性并不自动导致最佳的预测性能。在困难的预测问题中，套索可能具有优势。我们的结果为  $\ell_1$  相关的方法提供了新的见解，并支持在统计建模中使用  $\ell_1$  惩罚。

### 背景和动机：

- **Lasso 方法：** Lasso (Least Absolute Shrinkage and Selection Operator) 是一种用于线性回归和变量选择的方法，它通过对系数添加 L1 惩罚实现了参数的稀疏性，进而进行变量选择。
- **自适应 Lasso：** 自适应 Lasso 是对传统 Lasso 的改进，它引入了自适应权重，允许不同的特征具有不同的惩罚项。

### 主要内容：

1. **自适应权重：** 论文首先介绍了在 Lasso 惩罚项中引入自适应权重的思想。这些权重是通过估计每个系数的绝对值的倒数来获得的，旨在提高对较小系数的估计的权重，从而更好地适应数据。
2. **Oracle 性质：** 论文讨论了 Adaptive Lasso 的 Oracle 性质。Oracle 性质是指模型能够以近乎最佳的方式进行估计，即在具有相应结构的真实模型下，模型的性能能够接近最优。
3. **交叉验证：** 论文中提到了使用交叉验证来选择自适应 Lasso 惩罚参数的方法。通过交叉验证，可以选择最优的自适应权重，使得模型在测试数据上表现最好。
4. **数值实验：** 论文进行了一些数值实验，通过模拟和实际数据集验证了自适应 Lasso 方法的性能。这些实验有助于展示该方法在不同情况下的鲁棒性和效果。

### 结论：

- 论文总结了自适应 Lasso 的提出、Oracle 性质和使用交叉验证选择参数的方法。
- 强调了 Adaptive Lasso 在处理高维数据中的变量选择问题时的优势，特别是在具有较小效应的变量的情况下。

### 应用和影响：

- Adaptive Lasso 方法的提出为处理高维数据集中的变量选择问题提供了一种更灵活和适应性更强的工具。
- 该方法在统计学和机器学习领域中引起了广泛关注，成为处理高维数据的常见工具之一。

## 4 弹性网 | Regularization and variable selection via the elastic net

---

弹性网络回归（Elastic Net Regression）：弹性网络是岭回归和 Lasso 回归的组合，综合了它们的正则化效果。

<https://github.com/search?q=Regularization+and+variable+selection+via+the+elastic+net&type=code>

### 摘要

我们提出了一种新的正则化和变量选择方法，称为弹性网络。实际数据和模拟研究表明，弹性网络通常优于套索，同时具有类似的稀疏表示。此外，弹性网络促使产生一种分组效应，即强相关的预测因子往往共同进入或退出模型。当预测因子的数量 ( $p$ ) 远远大于观测值的数量 ( $n$ ) 时，弹性网络尤其有用。相比之下，套索在  $p \ll n$  的情况下并不是一个非常令人满意的变量选择方法。我们提出了一种名为 LARS-EN 的算法，用于高效计算弹性网络正则化路径，类似于套索中的 LARS 算法。

### 结论

Elastic Net 是一种用于回归和变量选择的正则化方法，结合了 L1 (Lasso) 和 L2 (Ridge) 惩罚。

"Regularization and variable selection via the elastic net" 这篇论文由 Hui Zou 和 Trevor Hastie 于 2005 年发表，介绍了 Elastic Net 方法及其在回归和变量选择中的应用。以下是对该论文的简要概述：

#### 背景和动机：

- **L1 和 L2 惩罚：** Lasso (L1) 和 Ridge (L2) 是两种常用的正则化方法。L1 惩罚倾向于生成稀疏模型，而 L2 惩罚有助于应对共线性问题。
- **Elastic Net 的引入：** Elastic Net 结合了 L1 和 L2 惩罚，通过引入两个调整参数 ( $\alpha$  和  $\lambda$ ) 来平衡它们的影响。这样，Elastic Net 既能保留 L1 的变量选择性质，又能克服 L2 在高度相关变量上的劣势。

#### 主要内容：

1. **Elastic Net 的优化问题：** 论文给出了 Elastic Net 的优化问题，该问题是通过目标函数引入 L1 和 L2 惩罚项来实现的。
2. **Elastic Net 的解：** 论文讨论了 Elastic Net 的解法，特别是介绍了使用坐标下降法 (coordinate descent) 来高效地求解 Elastic Net 优化问题。
3. **Elastic Net 的路径算法：** 论文介绍了 Elastic Net 路径算法，该算法在多个  $\lambda$  值上计算 Elastic Net 的解，形成一条解路径。这对于选择最优的正则化参数很有帮助。
4. **性质与实验：** 论文探讨了 Elastic Net 的性质，包括其稀疏性和一致性。此外，通过实验验证了 Elastic Net 在变量选择和预测性能上的优越性。

## 结论：

- 论文总结了 Elastic Net 的提出、其解的性质以及在实际问题中的应用。
- Elastic Net 作为一种灵活的正则化方法，能够在回归问题中同时实现变量选择和克服共线性问题。

## 应用和影响：

- Elastic Net 在统计学、机器学习和数据科学领域得到了广泛的应用，特别是在高维数据集上的回归问题。
- 论文的方法和理论为理解 Elastic Net 的性质以及有效地实现相关算法提供了基础，成为该领域的经典文献之一。

# 5 分组Lasso | Model selection and estimation in regression with grouped variables

---

## 摘要

我们考虑在回归中选择分组变量（因子）以进行准确预测的问题。这样的问题在许多实际情况下都会自然而然地出现，多因素方差分析问题是其中最重要且最为人熟知的例子。我们不是通过逐步向后消除来选择因子，而是专注于估计的准确性，并考虑了对套索、LARS算法和非负 Garrotte 进行因子选择的扩展。套索、LARS 算法和非负 Garrotte 是最近提出的可用于选择个别变量的回归方法。我们研究并提出了这些方法用于因子选择的扩展的高效算法，并展示这些扩展在因子选择问题中相较于传统的逐步向后消除方法具有更好的性能。我们研究了这些方法之间的相似性和差异。通过模拟和实际示例来说明这些方法。

## 结论

- Lasso变体：Group Lasso, Sparse Group Lasso

[https://blog.csdn.net/qq\\_40924873/article/details/128014355](https://blog.csdn.net/qq_40924873/article/details/128014355)

- Group Lasso 分组最小角回归算法 Yuan在2006年将lasso方法推广到group上面，诞生了group lasso。我们可以将所有变量分组，然后在目标函数中惩罚每一组的L2范数，这样达到的效果就是可以将一整组的系数同时消成零，即抹掉一整组的变量，这种手法叫做Group Lasso 分组最小角回归算法。其目标函数为：<https://blog.csdn.net/asd136912/article/details/78401915>
- Lasso回归的坐标下降法 [https://shenweichen.blog.csdn.net/article/details/77487729?](https://shenweichen.blog.csdn.net/article/details/77487729)  
spm=1001.2101.3001.6650.3&utm\_medium=distribute.pc\_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-3-77487729-blog-80344815.235%5Ev38%5Epc\_relevant\_sort\_base3&depth\_1-utm\_source=distribute.pc\_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromBaidu%7ERate-3-77487729-blog-80344815.235%5Ev38%5Epc\_relevant\_sort\_base3&utm\_relevant\_index=4

Ming Yuan Georgia Institute of Technology, Atlanta, USA and Yi Lin University of Wisconsin—Madison, USA  
[Received November 2004. Revised August 2005]

很抱歉，由于我无法直接查阅特定论文的内容，我无法提供 "Model selection and estimation in regression with grouped variables" 这篇由 Ming Yuan 和 Yi Lin 于2004年11月收到、2005年8月修订的论文的详细概

述。然而，我可以根据一般的模型选择和估计的方法给出一些可能的主题和内容：

## 用python代码分别实现Lasso回归问题的坐标下降算法和弹性网回归算法，并做测试

论文主题：

1. **回归模型中的变量选择：** 论文可能涉及在回归模型中选择哪些变量用于建模，以提高模型的解释性和泛化性能。
2. **分组变量：** 论文可能关注具有分组结构的变量，例如在分子生物学中的基因表达数据，以及如何在某种情况下进行回归建模。
3. **组稀疏性：** 论文可能介绍了对整个变量组引入稀疏性，以便一次性选择组和变量。
4. **Group Lasso 惩罚：** 可能讨论了 Group Lasso 惩罚的引入，以解决变量选择和组选择的问题。这是 L1 惩罚的一种扩展。
5. **块坐标下降法：** 论文可能提到了块坐标下降法，这是一种有效的方法，用于在 Group Lasso 中处理块状结构。
6. **交叉验证：** 论文可能涉及使用交叉验证来选择最优的正则化参数，以平衡拟合和泛化性能。

方法和算法：

1. **Group Lasso 优化问题的解决：** 论文可能介绍了如何有效地解决 Group Lasso 优化问题，以获得稀疏解。
2. **块坐标下降算法：** 论文可能讨论了块坐标下降算法的实现，以加速对 Group Lasso 问题的求解。
3. **交叉验证的应用：** 论文可能说明了如何使用交叉验证来选择合适的超参数，以获得最佳的模型性能。

结论和影响：

1. **总结方法的有效性：** 论文可能总结了所提出的方法在变量选择和回归建模中的有效性。
2. **对领域的影响：** 论文可能强调了该方法对于处理具有分组结构的数据集的重要性，并可能探讨了该方法在实际领域中的应用和影响。

由于我无法访问具体的论文内容，以上内容仅是一些基于题目和作者的合理猜测。若有具体的问题或需要详细了解论文的内容，请查阅论文原文。

## 6 NEARLY UNBIASED VARIABLE SELECTION UNDER MINIMAX CONCAVE PENALTY Author(s): Cun-Hui Zhang

---

我很抱歉，由于我无法直接访问特定的论文内容，我无法提供 "NEARLY UNBIASED VARIABLE SELECTION UNDER MINIMAX CONCAVE PENALTY" 这篇论文的详细概述。然而，我可以基于题目和通常的知识为您提供一些可能的主题和内容的猜测：

## 摘要

我们提出了MC+, 这是一种快速、连续、几乎无偏和准确的高维线性回归中惩罚变量选择的方法。LASSO快速而连续, 但存在偏差。LASSO的偏差可能阻止一致的变量选择。子集选择无偏但计算成本高昂。MC+有两个元素: 极小极大凹函数惩罚 (MCP) 和带惩罚的线性无偏选择 (PLUS) 算法。MCP在稀疏区域中给予惩罚损失的凸性, 具体来说, 在对变量选择和无偏性设定一定阈值的情况下, MCP在稀疏区域中提供了惩罚损失的凸性到最大程度。PLUS计算可能是非凸的惩罚损失函数在惩罚损失的临界点图的某个主要分支上的多个精确局部最小值。其输出是一个从无穷大惩罚到零惩罚的最小二乘解的连续分段线性路径。我们证明在通用惩罚水平下, MC+具有匹配未知数的符号的高概率, 因此正确选择, 而不需要假设LASSO所需的强不可表示条件。这种选择一致性适用于 $p \gg n$ 的情况, 并且被证明对于可能有多个局部最小值的MC+解是成立的。我们证明MC+在球中对回归系数的估计具有概率上的一致极小极大收敛率。我们使用SURE方法推导出一般惩罚LSE (包括LASSO和MC+估计量) 的自由度和 $C^\alpha$ -类型风险估计, 并证明它们的无偏性。基于估计的自由度, 我们提出了一个用于选择适当的惩罚水平的噪声水平的估计量。对于全秩设计和一般二次惩罚, 我们提供了惩罚LSE连续性的必要和充分条件。模拟结果极大地支持我们对卓越变量选择性能的断言, 并展示了所提方法的计算效率。

## 论文主题:

1. **变量选择:** 论文可能关注如何在回归分析中选择最重要的变量, 以建立更简单、更可解释的模型。
2. **Minimax Concave Penalty:** 论文的核心可能是介绍了 Minimax Concave Penalty (MCP) 这种方法, 作为一种稀疏学习的技术。MCP 是一种惩罚项, 用于在拟合模型时产生具有稀疏性的解。
3. **Nearly Unbiased Variable Selection:** 论文可能讨论了如何通过使用 Minimax Concave Penalty 实现接近无偏的变量选择, 即在变量选择中减小偏差。
4. **统计性质:** 论文可能讨论了所提出的方法在统计学上的性质, 包括一致性、渐近正态性等。

## 方法和算法:

1. **Minimax Concave Penalty 的形式:** 论文可能介绍了 MCP 的具体形式, 包括其数学表达式和如何与回归模型结合使用。
2. **Nearly Unbiased Variable Selection 方法:** 论文可能阐述了如何通过 MCP 实现接近无偏的变量选择, 可能包括一些具体的算法或优化方法。
3. **数值实验:** 论文可能通过数值实验来验证所提出方法的性能, 例如在模拟数据或实际数据集上的应用。

## 结论和影响:

1. **方法的有效性:** 论文可能总结了提出的方法在变量选择任务中的有效性, 并强调其在不同情境下的应用价值。
2. **对领域的影响:** 论文可能探讨了该方法对于相关领域、如统计学、机器学习等的影响。

请注意, 以上内容仅是基于题目的猜测, 具体的论文内容可能会有所不同。如需了解详细内容, 请查阅原始论文。



# 7 Regression shrinkage and selection via the lasso: a retrospective

---

Robert Tibshirani Stanford University, USA

我很抱歉，目前我无法提供 "Regression shrinkage and selection via the lasso: a retrospective" 这篇论文的详细概述，因为我无法直接访问特定的论文内容。然而，我可以基于题目和一般的知识为您提供关于 Lasso 方法的一般信息。

论文主题：

1. **Lasso 方法介绍：** 论文可能回顾了 Lasso 方法的基本原理和概念。Lasso (Least Absolute Shrinkage and Selection Operator) 是一种正则化方法，通过对回归系数引入 L1 惩罚来实现变量选择和模型收缩。
2. **回顾发展历程：** 论文可能回顾了 Lasso 方法的发展历程，包括它的起源、提出的动机以及它在统计学和机器学习中的演进。
3. **应用领域：** 论文可能讨论了 Lasso 方法在不同领域中的应用，例如在统计建模、机器学习、生物信息学等方面的成功案例。

方法和算法：

1. **Lasso 惩罚：** 论文可能详细介绍了 Lasso 惩罚的形式，以及如何通过对回归问题的目标函数添加 L1 惩罚来实现参数的稀疏性。
2. **坐标下降法等算法：** 论文可能探讨了用于求解 Lasso 问题的一些优化算法，如坐标下降法等。

结论和影响：

1. **方法的影响：** 论文可能总结了 Lasso 方法在统计学和机器学习领域的影响，以及它如何改变了变量选择和模型拟合的方式。
2. **对未来的展望：** 论文可能讨论了 Lasso 方法未来的发展方向，以及如何进一步改进或扩展这一方法。

请注意，以上内容是基于题目和我对 Lasso 方法的一般了解所提供的猜测，实际论文内容可能有所不同。如需了解详细信息，请查阅原始论文。

## 杂

---

- Lasso (最小绝对收缩和选择算子)、弹性网及组套索  
[https://blog.csdn.net/qq\\_38334677/article/details/131756217?spm=1001.2101.3001.6650.10&utm\\_medium=distribute.pc\\_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-10-131756217-blog-128014355.235%5Ev38%5Epc\\_relevant\\_sort\\_base3&depth\\_1-utm\\_source=distribute.pc\\_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-10-131756217-blog-128014355.235%5Ev38%5Epc\\_relevant\\_sort\\_base3&utm\\_relevant\\_index=17](https://blog.csdn.net/qq_38334677/article/details/131756217?spm=1001.2101.3001.6650.10&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-10-131756217-blog-128014355.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-10-131756217-blog-128014355.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=17)

- 数据集还蛮经典 [https://blog.csdn.net/qq\\_42871249/article/details/104659074?spm=1001.2101.3001.6650.1&utm\\_medium=distribute.pc\\_relevant.none-task-blog-2%7Edefault%7EOPENSEARCH%7ERate-1-104659074-blog-132959366.235%5Ev38%5Epc\\_relevant\\_sort\\_base3&depth\\_1-utm\\_source=distribute.pc\\_relevant.none-task-blog-2%7Edefault%7EOPENSEARCH%7ERate-1-104659074-blog-132959366.235%5Ev38%5Epc\\_relevant\\_sort\\_base3&utm\\_relevant\\_index=2](https://blog.csdn.net/qq_42871249/article/details/104659074?spm=1001.2101.3001.6650.1&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EOPENSEARCH%7ERate-1-104659074-blog-132959366.235%5Ev38%5Epc_relevant_sort_base3&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EOPENSEARCH%7ERate-1-104659074-blog-132959366.235%5Ev38%5Epc_relevant_sort_base3&utm_relevant_index=2)

## 正则化线性模型+岭回归+lasso回归+弹性网络+糖尿病数据案例分析

原创

VIP文章

@Irene

于 2020-03-05 10:39:27 发布

阅读量2k

收藏 19

点赞数 1

版权

分类专栏：

R语言

案例分享

### 1、岭回归+lasso回归+弹性网络

group lasso

10-17

lasso regression classification and bayesian network

【线性回归、岭回归、Lasso回归分别预测患者糖尿病病情】数... 最新发布 Lenhart001的博客 1371  
(详细描述数据集：如特征属性名称及意义、记录数等) (写出项目中涉及的主要算法原理及模型评价方法)

1 条评论 &gt;



应统难民研究生一枚

热评

可以求一份数据吗

写评论

对sklearn中的糖尿病数据集进行线性回归，岭回归，套索回归模型分... Superman980527的博客 4561  
def z3(): # 导入糖尿病数据集 from sklearn.datasets import load\_diabetes #打印数据集里面的Keys print(load\_dia...

正规方程求解法、岭回归、线性回归解决糖尿病预测问题 m0\_63845261的博客 2943  
一、问题描述 糖尿病预测数据集是Sklearn 提供的数据集。它从442例糖尿病患者的资料中取10个特征：年...

监督学习-线性模型-2. 岭回归&Lasso回归 04-09  
监督学习-线性模型-2. 岭回归&Lasso回归

python 数据集 多元线性回归模型预测房价 03-13  
python 源码集锦-多元线性回归模型预测房价

ML之LiR&Lasso：基于datasets糖尿病数据... 心比天高，仗剑走天涯，保持热爱，奔赴向梦想！低... 6463  
ML之LiR&Lasso：基于datasets糖尿病数据集利用LiR和Lasso算法进行(9→1)回归预测(三维图散点图可视化) 目...