机器学习

第二次作业

周帆

统计与管理学院
上海财经大学

Outline

作业

背景介绍：Lasso 问题求解 I

- 给定一组样本 $\{(x_i, y_i)\}_{i=1}^{n}$，考虑如下优化问题：

$$\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T \boldsymbol{\beta})^2 + \lambda_0 \sum_{j=1}^{p} |\beta_j|$$

- 或者等价地，令 $\lambda = n\lambda_0$，求解如下优化

$$\boldsymbol{\beta}^{t+1} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|y - x^T \boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1,$$

背景介绍：基于 Proximal 算法 I

- 记 $f(\boldsymbol{\beta}) = \|y - x^T\boldsymbol{\beta}\|_2^2$ and $g(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$, 其中 $f(\boldsymbol{\beta})$ 可导但 $g(\boldsymbol{\beta})$ 不可导。
- 给定当前迭代值 $\boldsymbol{\beta}^t$

$$\boldsymbol{\beta}^{t+1} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ f(\boldsymbol{\beta}^t) + \langle \nabla f(\boldsymbol{\beta}^t), \boldsymbol{\beta} - \boldsymbol{\beta}^t \rangle + g(\boldsymbol{\beta}) + \frac{1}{2\mu_t} \|\boldsymbol{\beta} - \boldsymbol{\beta}^t\|_2^2 \right\}$$

- 最优化条件为

$$0 \in \nabla f(\boldsymbol{\beta}^t) + \partial g(\boldsymbol{\beta}^{t+1}) + \frac{1}{\mu_t}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^t\right).$$

- 等价为

$$\begin{aligned}\boldsymbol{\beta}^{t+1} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \{ g(\boldsymbol{\beta}) + \frac{1}{2\mu_t} \|\boldsymbol{\beta} - (\boldsymbol{\beta}^t - \mu_t \nabla f(\boldsymbol{\beta}^t))\|_2^2 \\ &= \operatorname{Prox}_{\mu_t g}(\boldsymbol{\beta}^t - \mu_t \nabla f(\boldsymbol{\beta}^t)).\end{aligned}$$

背景介绍：基于 Proximal 算法 II

- 上述问题有显式解为:

$$(\beta^{t+1})_j = \left(1 - \frac{\lambda\mu}{|\tilde{\beta}^t_j|}\right)_+ \tilde{\beta}^t_j, \quad j = 1, ..., p$$

- 其中 $\tilde{\boldsymbol{\beta}}^t = \boldsymbol{\beta}^t - \mu_t \nabla f(\boldsymbol{\beta}^t)$.
- 对于 p 维向量 $u = (u_1, ..., u_p)^T \in \mathcal{R}^p$,

$$(u)_+ = \left(\max\{u_1\ 0\}, ..., \max\{u_p, 0\}\right)^T.$$

波士顿房屋价格数据 I

波士顿房屋数据

该数据集包含美国人口普查局收集的美国马萨诸塞州波士顿住房价格的有关信息，共有 506 个样本以及 13 个变量。

- CRIM–城镇人均犯罪率
- ZN - 占地面积超过 25,000 平方英尺的住宅用地比例。
- INDUS - 每个城镇非零售业务的比例。
- CHAS - Charles River 虚拟变量（如果是河道，则为 1; 否则为 0）
- NOX - 一氧化氮浓度（每千万份）
- RM - 每间住宅的平均房间数
- AGE - 1940 年以前建造的自住单位比例
- DIS- 加权距离波士顿的五个就业中心
- RAD - 径向高速公路的可达性指数
- TAX - 每 10,000 美元的全额物业税率

波士顿房屋价格数据 II

- PTRATIO - 城镇的学生与教师比例
- B - $1000(\mathrm{B_k} - 0.63)^2$ 其中 Bk 是城镇黑人的比例
- LSTAT - 人口状况下降

以及因变量为 MEDV(自有住房的中位数报价, 单位 1000 美元)。该数据集可通过 R 中 MASS 包或 Python 中 sklearn 包中直接获取。

作业

运用所学 LASSO 方法以及如下的 Proximal 算法，在 13 个变量中筛选出对 MEDV 有作用的变量，具体算法及相关设定如下。

基于 Proximal 算法的 Lasso 问题求解 I

- 迭代算法
  - 给定初值 $\boldsymbol{\beta}^0$ (可均设为 0 向量) 以及设定 μ(可设为 1 或根据数据作出相应调整), 以及事先给定收敛阈值 ϵ.
  - 循环运算如下迭代算法

$$(\beta^{t+1})_j = \left(1 - \frac{\lambda\mu}{|\tilde{\beta}_j^t|}\right)_+ \tilde{\beta}_j^t, \quad j = 1, ..., p$$

  - 其中 $\tilde{\boldsymbol{\beta}}^t = \boldsymbol{\beta}^t - \mu\nabla f(\boldsymbol{\beta}^t)$.
  - 重复上述步骤，直到收敛到指定误差为止。

基于 Proximal 算法的 Lasso 问题求解 I

- 收敛标准:
  1. 绝对收敛标准: $\|\boldsymbol{\beta}^{\mathrm{t+1}} - \boldsymbol{\beta}^{\mathrm{t}}\| < \epsilon$
  2. 相对收敛标准: $\frac{\|\boldsymbol{\beta}^{\mathrm{t+1}} - \boldsymbol{\beta}^{\mathrm{t}}\|}{\|\boldsymbol{\beta}^{\mathrm{t}}\|} < \epsilon.$
  3. 调整的相对收敛标准: $\frac{\|\boldsymbol{\beta}^{\mathrm{t+1}} - \boldsymbol{\beta}^{\mathrm{t}}\|}{\|\boldsymbol{\beta}^{\mathrm{t}}\| + \epsilon} < \epsilon.$
- 有关 λ 的设定: 通常我们可以采取网格筛选的思路: 即从 $10^{-3+0.1\mathrm{s}},\ \mathrm{s} = 0, ..., 60$ 中选取。
- 有关 λ 的选取: 5 交叉验证。

作业要求及标准 I

- 独自完成以上算法的编程，并且提供一段对结果的描述分析;
- 评价标准：不可调用现有函数包;
- 交作业日期：10 月 12 日晚 6 点或之前将代码和结果描述发送至助教邮箱;