

统计数据分析方法个人作业 Part 1

冯超

2023-06-20

改正后的表达式

sgl1 中正确的梯度表达式

$$\frac{1}{n} X^{(k)\top} (X^{(k)} \beta - r_{(-k)})$$

sgl3 中正确的损失函数表达式

$$l(\beta^{(k,l)}) = \frac{1}{n} \sum_{i=1}^n [\log(1 + e^{x_i^\top \beta}) - y_i x_i^\top \beta]$$

sgl3 中正确的梯度表达式

$$\frac{1}{n} \sum_{i=1}^n \left[x_i \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}} - x_i y_i \right]$$

Sparse Group Lasso 的解的特点，及与 Group Lasso、Lasso 解的异同

从解的稀疏性来看，Lambda 越大，越可能得到稀疏解。其中，Group Lasso 和 Sparse Group Lasso 都能得到整组稀疏的系数，即同一组内的系数均为 0，但 Lasso 并没有考虑整组的稀疏性，只是将单个变量的系数置为 0。考察组内的稀疏性，Group Lasso 无法在组内进一步将某个变量的系数置为 0，而 Sparse Group Lasso 仍会将组内的某些变量置为 0。

从分类准确率的结果来看，Sparse Group Lasso 的分类效果较好，最高的分类准确率最高能达到近 80%。但我进行多次实验后发现，三种模型的拟合结果受随机种子的影响较大，因此我认为该结果并不十分可靠。此外，Lambda 取值大小也对分类效果有影响。根据原始论文提供的代码，Sparse Group Lasso 选用的 Lambda 范围大致为 0.003 到 0.03 之间。越小的 Lambda 会使得运行速度越慢，即训练耗时越长，但我手动实现的 sgl3 函数运行 100 个 lambda 耗时 4 个小时，因此我没有对大量的 lambda 参数进行实验，而是借用论文调包时产生的 lambda 列表。

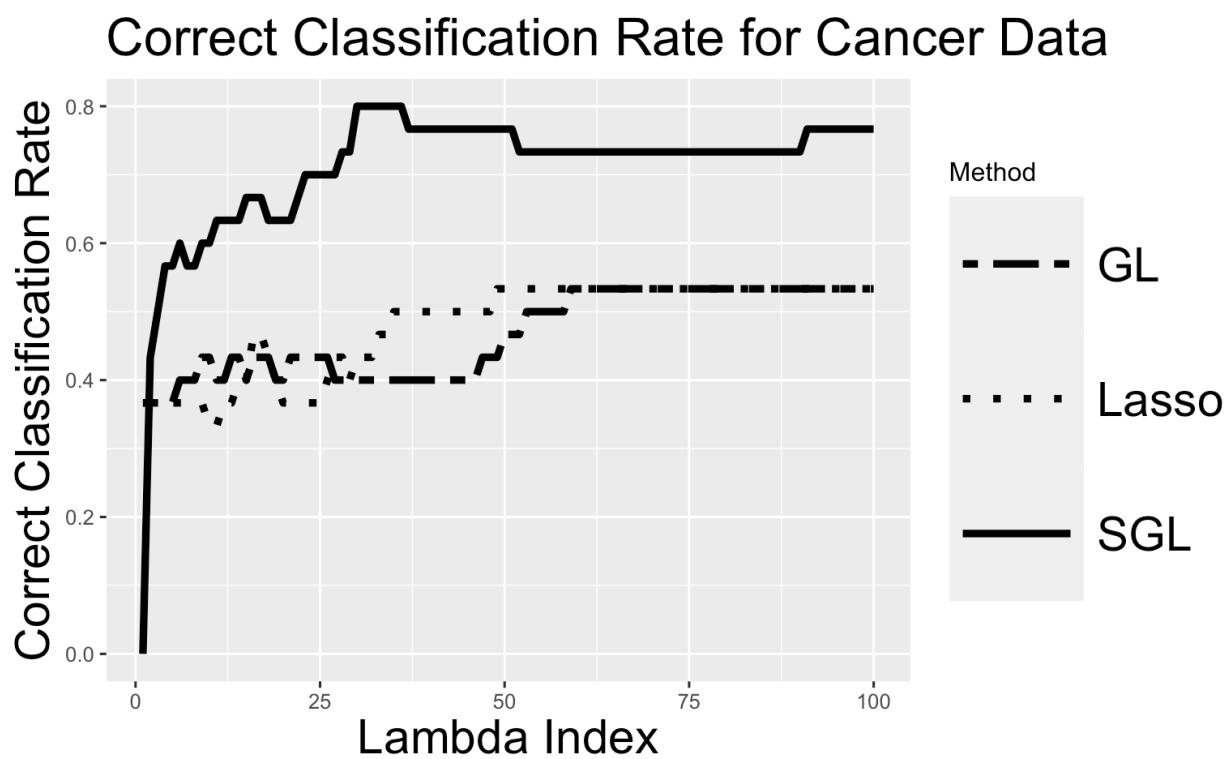


图 1: 分类结果