

Final Project (Part 1, 20 分)

6 月 20 日 11:59pm 前提交至 elearning

请提交 **R codes**、产生的图以及要求的表达式和文字描述（可以附加在代码或者 notebook 中）。（建议用 R 实现，但若对 python 更熟悉，提交 python codes 亦可。）

请先仔细阅读参考文献“A Sparse-Group Lasso”（阅读算法部分即可，不需要看证明）以及附件的“sgl_assign.R”代码。然后根据以下要求完成相应的内容（需要填充的函数在“sgl_assign.R”的第 222 行之后）。

1. 根据代码中各个函数的注释，完成 sgl1 函数中的函数 **threshOperator**、**calculateL1**、**calculateGradL1** 以及 **updateU1**，使得完成后的 sgl1 函数可以按照注释中的 inputs 和 outputs 实现文献 Section 3.2 第一部分的简单 sparse group lasso 算法；并根据注意 1.1 写出正确表达式。

注意 1.1：文献第 236 页第二行的 **gradient** 计算表达式有误，请写出正确表达式，并在代码中使用正确的式子计算。

注意 1.2：不得修改 sgl_assign.R 中已经给出的代码部分，只需要完成各个函数对应的代码。可以另外添加一些自己需要的辅助函数。

注意 1.3：如果选择使用 python，则自行按照 sgl_assign.R 写出各个函数对应的函数头和函数体。sgl1、sgl2 和 sgl3 的 inputs、outputs 和代码运行逻辑不得进行修改。

注意 1.4：为了简化问题，所有的 y 和 X 均已经过中心化标准化处理。

2. 完成 sgl2 函数中的函数 **calculatePenalty**、**updateU2** 以及 **updateNesterov**，使得完成后的 sgl2 函数可以按照注释中的 inputs 和 outputs 实现 Section 3.2 第二部分中使用了 Nesterov 方法加速的 sparse group lasso 算法。
3. 完成 sgl3 中的函数 **calculateGradLogistic** 和 **updateNesterovLogistic**，使得完成后的 sgl3 函数可以按照注释中的 inputs 和 outputs 实现 Section 4 中针对 logistic regression 的使用了 Nesterov 方法加速的 sparse group lasso 算法。
4. 请利用前几问得到的“sgl_assign.R”中合适的函数对应的方法，复现文献中的 Figure 2。复现结果与原文不需完全一致，lambda 的个数可以少取一些，结果呈现一定趋势即可。根据 sparse group lasso 和 group lasso、lasso 的目标函数和所得解的差异，简要描述 sparse group lasso 的解的特点，及与 group lasso、lasso 解的异同。

注意 4.1：若前几问中的函数没有成功写出，可调用现有 sparse group lasso 的 package 做第四问（会酌情扣分）。

注意 4.2：数据预处理可以参照文献附录提供的代码。如果遇到解无法收敛，或者迭代中某些值趋向于无穷的问题，可对 default 参数值进行调整（如 maxIterO, thrO 等）。lasso 与 group lasso 的实现可直接调用现有 packages。

其他注意事项：

1. 提交的代码应为.R 或.Rmd 或.py 或.ipynb 格式文件（请对代码进行适当的注释）。提交其他格式文件的酌情扣分。
2. 提交的代码应可以直接运行得到结果。如果代码内容有所缺失，则缺失的部分一律按照完全错误处理。
3. 为公平起见，若无特殊原因，截止时间后的一周内提交的 proj 满分为 15 分，逾期一周以上零分处理。
4. 若发现抄袭(包括网络抄袭)，抄袭和被抄袭的均按零分处理。