

1、先把数据的统计结果算出来（已完成）

- (1) 男女分开 构造两个数据集 把gender去掉
- (2) 对数据进行总结，男女数居各多少，特征数量，
- (3) 600×646的数据，计算每个病人的每个特征的均值，标准差。根据标准差是否为零判断特征的取留。

2、验算数学推导

SLEP包: $\min_x * 1/2 \|Ax - y\|_2^2 + \lambda \|x\|_1$

QP问题: $\min_{\beta^{(n)}} 1/2 (\beta^{(n)'} Q^{(n)} \beta^{(n)} - C^{(n)'} \beta^{(n)} + \sum_{ij} w_{ij}^{-r} |\beta_{ij}^{(n)}|$

步骤:

1、 $\beta^{(n)}$ 、 $Q^{(n)}$ 、 $C^{(n)}$ 、 w 都是 $p \times p$ 的矩阵， $\beta_v^{(n)} = \beta^{(n)}(:, v)$ 、 $C_v^{(n)} = C^{(n)}(:, v)$ 、 $w_v = w(:, v)$

$QQ = \text{diag}(Q^{(n)}, Q^{(n)} \dots Q^{(n)})$

$\beta_v^{(n)}$ 、 $C_v^{(n)}$ 、 w_v 都是 $p^2 \times 1$ 的向量， QQ 是 $p^2 \times p^2$ 的矩阵

2、对应的 $\sum_{ij} w_{ij}^{-r} |\beta_{ij}^{(n)}| = \|x\|_1$ ，其中 x 是 $p^2 \times 1$ 的向量。

$\sum_{ij} w_{ij}^{-r} |\beta_{ij}^{(n)}| = \sum_{ij} |w_{ij}^{-r} \beta_{ij}^{(n)}|$

假设 $x = Z^* \beta_v^{(n)}$ ，这里 Z 与 w 有对应关系

$\beta_v^{(n)} = Z^{-1} * X$ Z 是 $p^2 \times p^2$ 的矩阵， Z^{-1} 也是 $p^2 \times p^2$ 的矩阵

$H = Z^{-1} * QQ * Z = h * h$

$C_v^{(n)'} Z^{-1} = h' * e$

3、QP等价: $\min_{\beta_v^{(n)}} 1/2 (\beta_v^{(n)'} QQ \beta_v^{(n)} - C_v^{(n)'} \beta_v^{(n)} + \|x\|_1$ 其中 $x = Z^* \beta_v^{(n)}$

$\min_{\beta_v^{(n)}} 1/2 (X' Z^{-1} * QQ * Z^{-1} X) - e' h X + \|x\|_1$

$\min_{\beta_v^{(n)}} 1/2 (X' h' h X) - e' h X + \|x\|_1$

4、等价于 $\min_x 1/2 \|hx - e\|_2^2 + \|x\|_1$ ，可以代入SLEP包中求解。

问题

1、数据量太大了， $p^2 \times p^2$ 矩阵过于庞大，导致程序计算非常缓慢， $p=50$ 时就非常慢了。

2、 Z 的代入那里，和 w 的换算关系

3、调和平均数中，有 β 的元素是0。导致 w 会无解

4、对比算法选哪几个？

3、 调用程序库计算

4、 调用分类器

5、 分类结果评估
