2019-03-21

- Recoding Bonte's Simplified Fixed Hessian Newton Method.

  整理思路：Matlab代码>>Python代码

  发现的问题:

  - 设置初始权重为1会使得Sigmoid函数的输入区间很大,
    以至于很难使用多项式来替代Sigmoid函数. (SET weight = 0)
  - 使用似然函数估计模型时：$Y \in \{0,\ 1\}$ 和 $Y \in \{-1, +1\}$ 似然函数计算公式不同
  - 使用似然函数衡量模型时可能出现的问题，替代方案是ROC曲线和AUC值
- 研究了一下ROC曲线和AUC值, 并用Python实现
- 使用多项式替换Sigmoid函数, 并没有很大的改变:
  - 可能是Sigmoid函数的输入值区间很小，多项式在这个小区间内拟合的很好

$Y \in \{0, 1\}$ 和 $Y \in \{-1, +1\}$ 似然函数计算公式不同

$Y = 2Y - 1$ 数据预处理时容易变换类别，MLE也应该跟着变换

$sigm(z) = \frac{1}{1+e^{-z}}$ $\quad sigm'(z) = sigm(z) \cdot (1 - sigm(z))$ $\quad h_\theta(x) = sigm(\theta^T x)$	
$Y \in \{0, 1\}$	$Y \in \{-1, +1\}$
$P(y = 1 \mid x; \theta) = h_\theta(x)$	$P(y = +1 \mid x; \theta) = h_\theta(x) = sigm(\theta^T x)$
$P(y = 0 \mid x; \theta) = 1 - h_\theta(x)$	$P(y = -1 \mid x; \theta) = 1 - h_\theta(x) = sigm(-\theta^T x)$
$P(y \mid x; \theta) = (h_\theta(x))^y (1 - h_\theta(x))^{1-y}$	$P(y \mid x; \theta) = sigm(y\theta^T x)$
$L(\theta) = \prod_{i=1}^{m} (h_\theta(x_i))^{y_i} (1 - h_\theta(x_i))^{1-y_i}$	$L(\theta) = \prod_{i=1}^{m} \frac{1}{1 + e^{-y_i \theta^T x_i}}$
$\ell(\theta) = \log L(\theta)$ $= \sum_{i=1}^{m} y_i \log h_\theta(x_i) + (1 - y_i) \log 1 - h_\theta(x_i)$	$\ell(\theta) = \log L(\theta)$ $= -\sum_{i=1}^{m} \log\left(1 + e^{-y_i \theta^T x_i}\right)$
$grad = X^T (Y - h_\theta(X))$	$\partial \ell / \partial \theta_i = \sum_i (1 - sigm(y_i \theta^T x_i))\, y_i x_i$

Software Technology Lab

Table 1. Data Exhibiting Complete Separation.

x	y
-5	0
-4	0
-3	0
-2	0
-1	0
1	1
2	1
3	1
4	1
5	1

For these data, it can be shown that the ML estimate of the intercept is 0. Figure 1 shows a graph of the log-likelihood as a function of the slope "beta".

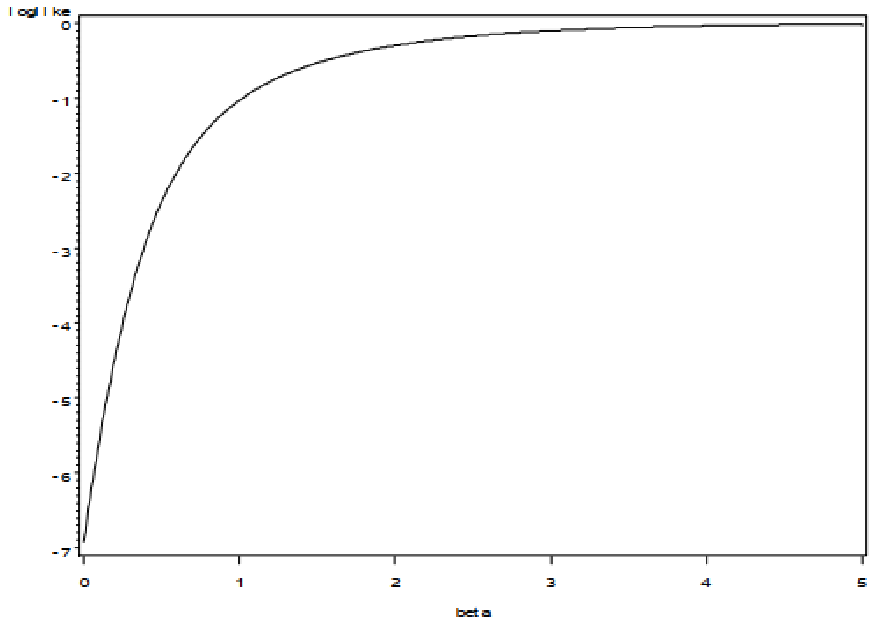

Figure 1. Log-likelihood as a function of the slope under complete separation

It is apparent that, although the log-likelihood is bounded above by 0, it does not reach a maximum as beta increases.

Table 2. Data Exhibiting Quasi-Complete Separation.

x	*y*
-5	0
-4	0
-3	0
-2	0
-1	0
0	0
0	1
1	1
2	1
3	1
4	1
5	1

What distinguishes this data set from the previous one is that there are two additional observations, each with *x* values of 0 but having different values of *y*.

The log-likelihood function for these data, shown in Figure 2, is similar in shape to that in Figure 1. However, the asymptote for the curve is not 0, but a number that is approximately -1.39. In general, the log-likelihood function for quasi-complete separation will not approach 0, but some number lower than that. In any case, the curve has no maximum so, again, the maximum likelihood estimate does not exist.

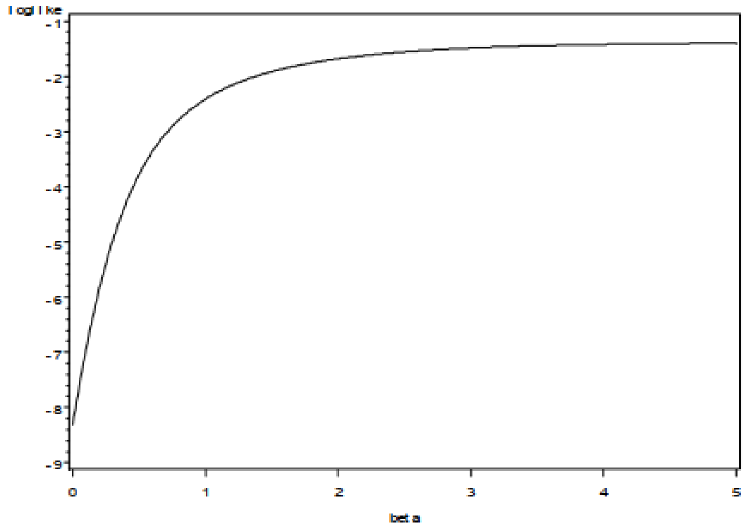

Figure 1. Log-likelihood as a function of the slope, quasi-complete separation.

➢使用多项式替换Sigmoid函数, 并没有很大的改变:

- 可能是Sigmoid函数的输入值区间很小, 多项式在这个小区间内拟合的很好