Homework 1

520030910246 薛家奇

Problem 1

$$
\begin{aligned}
\beta^{\text{new}} &= \beta^{\text{old}} + \eta \frac{\partial \log \Pr(\beta)}{\partial \beta} \\
&= \beta^{\text{old}} + \eta \frac{\partial \log \prod_{i=1}^{n} \frac{e^{y_i^* X_i^\top \beta}}{1+e^{X_i^\top \beta}}}{\partial \beta} \\
&= \beta^{\text{old}} + \eta \frac{\partial \sum_{i=1}^{n} \left(y_i^* X_i^\top \beta - \log\left(1 + e^{X_i^\top \beta}\right)\right)}{\partial \beta} \\
&= \beta^{\text{old}} + \eta \sum_{i=1}^{n} \frac{\partial \left(y_i^* X_i^\top \beta - \log\left(1 + e^{X_i^\top \beta}\right)\right)}{\partial \beta} \\
&= \beta^{\text{old}} + \eta \sum_{i=1}^{n} \left(y_i^* X_i - \frac{e^{X_i^\top \beta}}{1 + e^{X_i^\top \beta}} X_i \right) \\
&= \beta^{\text{old}} + \eta \sum_{i=1}^{n} \left(y_i^* - \frac{e^{X_i^\top \beta}}{1 + e^{X_i^\top \beta}} \right) X_i \\
&= \beta^{\text{old}} + \eta \sum_{i=1}^{n} \left(y_i^* - p_i \right) X_i
\end{aligned}
$$

where

$$
p_i = \frac{e^{X_i^\top \beta}}{1 + e^{X_i^\top \beta}} = \frac{1}{1 + e^{-X_i^\top \beta}}
$$

The reasons of gradient should be computed on $\log \Pr(\beta)$, not on $\Pr(\beta)$:

1. Converting cumulative multiplication into cumulative addition($\prod \to \sum$) makes calculating gradients easier.
2. Cumulative multiplication($\prod$) of many decimals between zero and one will cause the result to converge to zero and `float` type with finite precision will cause loss of precision problem, while cumulative addition($\sum$) will not.