牛顿法

张驰

October 6, 2024

$$\theta := \theta - H^{-1}\nabla_\theta \ell(\theta) \tag{0.1}$$

在我们的例子中 θ 为 K 列矩阵，其中 θ_i 为第 i 列。

$$\ell(\theta) = -\sum_{i=1}^{N} \log\left(\frac{e^{\theta^T_{y^{(i)}}\cdot x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_k^T x^{(i)}}}\right) \tag{0.2}$$

$$\frac{\partial \ell(\theta)}{\partial \theta_i} = \sum_{j=1}^{N}\left(\frac{e^{\theta_i^T x^{(j)}}}{\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}}} - 1\{y^{(j)} = i\}\right)\cdot x^{(j)} \tag{0.3}$$

$$\theta_i := \theta_i - H^{-1}\frac{\partial \ell(\theta)}{\partial \theta_i}\text{[1]} \tag{0.4}$$

其中，$H = \frac{\partial^2 \ell(\theta)}{\partial^2 \theta_i}$；$x^{(j)}$ 为第 j 个样例的特征，以列向量表示；$y^{(j)}$ 为第 j 个样例的结果。

$$\begin{aligned}
\frac{\partial^2 \ell(\theta)}{\partial^2 \theta_i} &= \sum_{j=1}^{N} \frac{\partial\left(\frac{e^{\theta_i^T x^{(j)}}}{\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}}} - 1\{y^{(j)} = i\}\right)\cdot x^{(j)}}{\partial \theta_i} \\
&= \sum_{j=1}^{N} x^{(j)} \cdot \frac{\partial\left(\frac{e^{\theta_i^T x^{(j)}}}{\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}}}\right)}{\partial \theta_i} \\
&= \sum_{j=1}^{N} x^{(j)} \cdot \frac{(x^{(j)} e^{\theta_i^T x^{(j)}})\cdot \sum_{k=1}^{K} e^{\theta_k^T x^{(j)}} - e^{\theta_i^T x^{(j)}}\cdot (x^{(j)} e^{\theta_i^T x^{(j)}})}{\left(\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}}\right)^2} \\
&= \sum_{j=1}^{N} x^{(j)} \cdot \frac{x^{(j)} e^{\theta_i^T x^{(j)}}\cdot\left(\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}} - e^{\theta_i^T x^{(j)}}\right)}{\left(\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}}\right)^2} \\
&= \sum_{j=1}^{N} x^{(j)} (x^{(j)})^T \cdot e^{\theta_i^T x^{(j)}} \cdot \frac{\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}} - e^{\theta_i^T x^{(j)}}}{\left(\sum_{k=1}^{K} e^{\theta_k^T x^{(j)}}\right)^2}
\end{aligned} \tag{0.5}$$

[1] $H(\theta_i^{(k+1)} - \theta_i^{(k)}) = \frac{\partial \ell(\theta)}{\partial \theta_i}\big|_{\theta_i = \theta_i^{k+1}} - \frac{\partial \ell(\theta)}{\partial \theta_i}\big|_{\theta_i = \theta_i^{k}}$