机器学习及其在化学中的应用

(2023 秋)

第 2 次书面作业答案

1. Sigmoid 函数的性质:
$$\sigma(a) := \frac{1}{1+\mathrm{e}^{-a}}. \tag{1}$$
   (a) 对称性.
$$\begin{aligned}\sigma(-a)+\sigma(a) &= \frac{1}{1+\mathrm{e}^{a}}+\frac{1}{1+\mathrm{e}^{-a}}\\ &= \frac{\mathrm{e}^{-a}}{\mathrm{e}^{-a}+1}+\frac{1}{1+\mathrm{e}^{-a}}\\ &= \frac{\mathrm{e}^{-a}+1}{1+\mathrm{e}^{-a}}=1.\end{aligned} \tag{2}$$
   (b) 反函数. 考虑 $y=\sigma(a)$, 由 (1) 解得
$$\mathrm{e}^{-a}=\frac{1-y}{y}, \tag{3}$$
   即:
$$a=\sigma^{-1}(y)=-\ln\frac{1-y}{y}=\ln\frac{y}{1-y}. \tag{4}$$

2. 根据题意, 训练得到的 Logistic 模型为
$$y(\mathrm{x};\mathrm{w})=\sigma(\mathrm{w}^{\mathrm{T}}\mathrm{x})=\sigma(2-2x_1+x_2). \tag{5}$$
   (a) 代入 $x_1=0.5, x_2=-1$, 得到 $y=\sigma(0)=0.5$; 代入 $x_1=1, x_2=-1$, 得到 $y=\sigma(-1)=0.269$.

   (b) 决策线为 $0=\mathrm{w}^{\mathrm{T}}\mathrm{x}$, 也即,
$$0=2-2x_1+x_2. \tag{6}$$

3. 课件已给出一阶导计算公式, 但这里我们仍然从头演示计算过程. 由复合函数链式法则, 我们先计算:
$$\begin{aligned}\frac{\partial y}{\partial \mathrm{w}_i} &= \frac{\partial\sigma(\mathrm{w}^{\mathrm{T}}\mathrm{x})}{\partial(\mathrm{w}^{\mathrm{T}}\mathrm{x})}\cdot\frac{\partial(\mathrm{w}^{\mathrm{T}}\mathrm{x})}{\partial\partial\mathrm{w}_i}\\ &= \frac{\mathrm{e}^{-\mathrm{w}^{\mathrm{T}}\mathrm{x}}}{(1+\mathrm{e}^{-\mathrm{w}^{\mathrm{T}}\mathrm{x}})^2}\mathrm{x}_i\\ &= y(1-y)\mathrm{x}_i.\end{aligned} \tag{7}$$
   于是, 损失函数的一阶梯度为:
$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathrm{w}_i} &= -\frac{\partial}{\partial\mathrm{w}_i}\left(\sum_{n=1}^{N}\left(t^{(n)}\ln\sigma(\mathrm{w}^{\mathrm{T}}\mathrm{x}^{(n)})+(1-t^{(n)})\ln\left(1-\sigma(\mathrm{w}^{\mathrm{T}}\mathrm{x}^{(n)})\right)\right)\right)\\ &= -\sum_{n=1}^{N}\left(t^{(n)}\frac{1}{\sigma(\mathrm{w}^{\mathrm{T}}\mathrm{x})}\frac{\partial y^{(n)}}{\partial\mathrm{w}_i}-(1-t^{(n)})\frac{1}{1-\sigma(\mathrm{w}^{\mathrm{T}}\mathrm{x})}\frac{\partial y^{(n)}}{\partial\mathrm{w}_i}\right)\\ &= -\sum_{n=1}^{N}\left(\frac{t^{(n)}}{y^{(n)}}-\frac{1-t^{(n)}}{1-y^{(n)}}\right)y^{(n)}(1-y^{(n)})\mathrm{x}_i^{(n)}\\ &= \sum_{n=1}^{N}(y^{(n)}-t^{(n)})\mathrm{x}_i^{(n)};\end{aligned} \tag{8}$$

进而可以计算其二阶梯度 (Hessian 矩阵):

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \mathrm{w}_i \partial \mathrm{w}_j} &= \sum_{n=1}^{N} \frac{\partial y^{(n)}}{\partial \mathrm{w}_j} \mathrm{x}_i^{(n)} \\
&= \sum_{n=1}^{N} y^{(n)}(1-y^{(n)}) \mathrm{x}_i^{(n)} \mathrm{x}_j^{(n)},
\end{aligned} \tag{9}
$$

写成矩阵形式即为:

$$
\frac{\partial^2 \mathcal{L}}{\partial \mathrm{w}^2} = \sum_{n=1}^{N} y^{(n)}(1-y^{(n)}) \mathrm{x}^{(n)} \mathrm{x}^{(n)\mathrm{T}}. \tag{10}
$$