DDA3020 Machine Learning
Gradient of Softmax Regression

10/14/2022

Recall the objective function of softmax regression:

$$\begin{aligned} J(\mathbf{W}) &= -\frac{1}{m}\sum_{i}^{m}\sum_{j}^{C}\left[\mathbb{I}(y_i = j)\log(f_{\mathbf{w}_j,b_j}(\mathbf{x}_i))\right] \\ &= -\frac{1}{m}\sum_{i}^{m}\sum_{j}^{C}\left[\mathbb{I}(y_i = j)\log\frac{\exp(\mathbf{w}_j^\top\mathbf{x}+b_j)}{\sum_{c=1}^{C}\exp(\mathbf{w}_c^\top\mathbf{x}+b_c)}\right. \end{aligned} \tag{1}$$

Let $z_j = \mathbf{w}_j^\top\mathbf{x}_i + b_j$ and $f_j = f_{\mathbf{w}_j,b_j}(\mathbf{x}_i) = \frac{\exp(z_j)}{\sum_{c=1}^{C}\exp(z_c)}$. Consider one sample for convenience.

$$\begin{aligned} \frac{\partial J_i}{\partial z_k} &= -\sum_{j=1}^{C}\left[\mathbb{I}(y_i=j)\frac{1}{f_j}\frac{\partial f_j}{\partial z_k}\right] \\ &= -\left(\sum_{j=k}^{C}\left[\mathbb{I}(y_i=j)\frac{1}{f_j}\frac{\partial f_j}{\partial z_k}\right] + \sum_{j\neq k}^{C}\left[\mathbb{I}(y_i=j)\frac{1}{f_j}\frac{\partial f_j}{\partial z_k}\right]\right) \\ &= -\left(\left[\mathbb{I}(y_i=k)\frac{1}{f_k}\frac{\partial f_k}{\partial z_k}\right] + \sum_{j\neq k}^{C}\left[\mathbb{I}(y_i=j)\frac{1}{f_j}\frac{\partial f_j}{\partial z_k}\right]\right) \\ &= -\left[\mathbb{I}(y_i=k)\frac{1}{f_k}\left(\frac{\exp(z_k)}{\sum_{c=1}^{C}\exp(z_c)} - \frac{\exp(z_k)\exp(z_k)}{\left(\sum_{c=1}^{C}\exp(z_c)\right)^2}\right)\right] \\ &\quad -\sum_{j\neq k}^{C}\left[-\mathbb{I}(y_i=j)\frac{1}{f_j}\frac{\exp(z_j)\exp(z_k)}{\left(\sum_{c=1}^{C}\exp(z_c)\right)^2}\right] \\ &= -\left(\mathbb{I}(y_i=k)\frac{1}{f_k}\times\left(f_k - f_k^2\right) - \sum_{j\neq k}^{C}\mathbb{I}(y_i=j)\frac{1}{f_j}\times f_j\times f_k\right) \\ &= -\left(\mathbb{I}(y_i=k) - \mathbb{I}(y_i=k)f_k - \sum_{j\neq k}^{C}\mathbb{I}(y_i=j)f_k\right) \\ &= -\left(\mathbb{I}(y_i=k) - \sum_{j=1}^{C}\mathbb{I}(y_i=j)f_k\right) \qquad \text{Note that } \sum_{j=1}^{C}\mathbb{I}(y_i=j)=1 \\ &= f_k - \mathbb{I}(y_i=k) \end{aligned} \tag{2}$$

Now we obtain

$$\frac{\partial J}{\partial \mathbf{w}_k} = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial J_i}{\partial z_k}\frac{\partial z_k}{\partial \mathbf{w}_k} = \frac{1}{m}\sum_{i=1}^{m}\left(f_k - \mathbb{I}(y_i=k)\right)\mathbf{x}_i \tag{3}$$