Homework 2: Solutions
600.482/682 Deep Learning
Fall 2018

Kaushik Srinivasan

October 12, 2018

With collaboration with Nathan Vallapureddy

1.

$$f(x_1, x_2, w_1, w_2) = (1 + e^{-(w_1x_1 - w_2x_2)})^{-1}$$

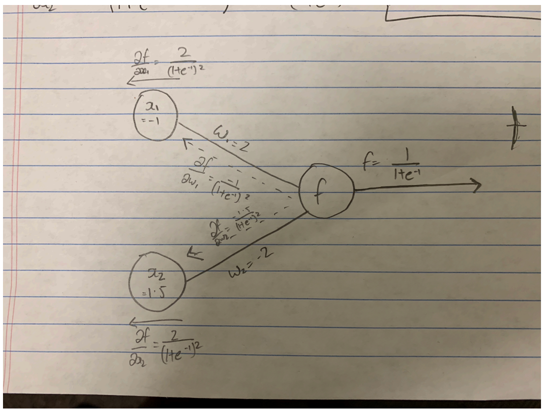
$$f(-1, 1.5, 2, -2) = \frac{1}{(1 + e^{-1})}$$

$$\frac{\partial f}{\partial w_1} = \frac{x_1}{(1 + e^{-(w_1x_1 - w_2x_2)})^2} = \frac{-1}{(1 + e^{-1})^2}$$

$$\frac{\partial f}{\partial w_w} = \frac{-x_2}{(1 + e^{-(w_1x_1 - w_2x_2)})^2} = \frac{-1.5}{(1 + e^{-1})^2}$$

$$\frac{\partial f}{\partial x_1} = \frac{w_1}{(1 + e^{-(w_1x_1 - w_2x_2)})^2} = \frac{2}{(1 + e^{-1})^2}$$

$$\frac{\partial f}{\partial x_2} = \frac{-w_2}{(1 + e^{-(w_1x_1 - w_2x_2)})^2} = \frac{2}{(1 + e^{-1})^2}$$



2. (a)

$$
\begin{aligned}
O(D,\theta) &= -\sum_i \log[z_i^{y_i}(1-z_i)^{(1-yi)}] \\
&= -\sum_i y_i \log z_i + (1-y_i)\log(1-z_i) \\
&= -\sum_i y_i \log \sigma(\theta^T x_i) + (1-y_i)\log(1-\sigma(\theta^T x_i)) \\
\nabla O &= \left(\frac{\partial O}{\partial \theta_1}, \frac{\partial O}{\partial \theta_2}, \cdots, \frac{\partial O}{\partial \theta_n}\right) \\
\frac{\partial O}{\partial \theta_j} &= \frac{\partial O}{\partial \theta_1}\left[-\sum_i y_i \log \sigma(\theta^T x_i) + (1-y_i)\log(1-\sigma(\theta^T x_i))\right] \\
&= -\sum_i \left[y_i \frac{\sigma'(\theta^T x_i)}{\sigma(\theta^T x_i)} - (1-y_i)\frac{-\sigma'(\theta^T x_i)}{1-\sigma(\theta^T x_i)}\right] \qquad \boxed{\sigma'(\theta^T x_i) = \sigma(\theta^T x_i)\cdot(1-\sigma(\theta^T x_i))} \\
&= -\sum_i y_i(1-\sigma(\theta^T x_i))x_i^{(j)} - (1-y_i)\sigma(\theta^T x_i)x_i^{(j)}
\end{aligned}
$$

$$
\boxed{\frac{\partial O}{\partial \theta_j} = -\sum_i (y_i - \sigma(\theta^T x_i))x_i^{(j)}}
$$

$$
\boxed{\frac{\partial^2 O}{\partial \theta_j \theta_k} = \sum_i \sigma(\theta^T x_i)\cdot(1-\sigma(\theta^T x_i))x_i^{(j)}x_i^{(k)}}
$$

$$
g + H\Delta\theta = 0 \Longrightarrow \boxed{\Delta\theta = H^{-1}g}
$$

(b) Check Code

(c) Code submitted

3. (a)

$$
P(y=k|x,\theta) = \frac{exp\{\theta_k^T x\}}{\sum_j exp\{\theta_j^T x\}}
$$

$$
\begin{aligned}
O(D;\theta) &= \prod_i \prod_k \left(\frac{exp\{\theta_k^T x_i\}}{\sum_j exp\{\theta_j^T x_i\}}\right)^{I_{\{y_i=k\}}} \\
-\log(O(D;\theta)) &= -log\left(\prod_i \prod_k \left(\frac{exp\{\theta_k^T x_i\}}{\sum_j exp\{\theta_j^T x_i\}}\right)^{I_{\{y_i=k\}}}\right) \\
&= -\left(\sum_i \sum_k I_{\{y_i=k\}}\theta_k^T x_i - \log(\sum_j exp\{\theta_j^T x_i\})\right)
\end{aligned}
$$

(b)

$$
\nabla - \log(O(D;\theta))
$$

$$
\begin{aligned}
&= \sum_i \frac{\partial}{\partial \theta_k^{(n)}} (\theta_k^T x_i - \log(\sum_j \exp\{\theta_j^T x_i)\}) \\
&= \left(\sum_i \frac{\partial}{\partial \theta_k^{(n)}} \theta_k^T x_i\right) - \left(\sum_i \frac{\partial}{\partial \theta_k^{(n)}} \log\left(\sum_j exp\{\theta_j^T x_i\}\right)\right) \\
&= \sum_i x_i^{(n)} - \sum_i \frac{\frac{\partial}{\partial \theta_k^{(n)}} \sum_j exp\{\theta_j^T x_i\}}{\sum_j exp\{\theta_j^T x_i\}} \\
&= \sum_i x_i^{(n)} - \sum_i \frac{exp\{\theta_j^T x_i\} \frac{\partial}{\partial \theta_k^{(n)}} \theta_j^T x_i}{\sum_j exp\{\theta_j^T x_i\}} \\
&= \sum_i x_i^{(n)} - \sum_i \frac{exp\{\theta_k^T x_i\} x_i^{(n)}}{\sum_j exp\{\theta_j^T x_i\}} \\
&= \sum_i x_i^{(n)} \left(1 - \frac{exp\{\theta_k^T x_i\}}{\sum_j exp\{\theta_j^T x_i\}}\right)
\end{aligned}
$$

(c) check code