ps4-4-sol

virusdoll

June 17, 2021

(a)

$$\begin{aligned}
\ell(W) &= \sum_{i=1}^{m} \big(\log|W| + \sum_{j=1}^{d} \log g'(w_j^T x^{(i)}) \big) \\
&= \sum_{i=1}^{m} \big(\log|W| + \sum_{j=1}^{d} \log (2\pi)^{\frac{1}{2}} \exp(-\frac{1}{2}(w_j^T x^{(i)})^2) \big) \\
&= \sum_{i=1}^{m} \big(\log|W| + \sum_{j=1}^{d} -\frac{1}{2}\log 2\pi - \frac{1}{2}(w_j^T x^{(i)})^2 \big) \\
&= \sum_{i=1}^{m} \big(\log|W| - \frac{1}{2}d\log 2\pi - \frac{1}{2}{x^{(i)}}^T W^T W x^{(i)} \big)
\end{aligned}$$

$$\begin{aligned}
\nabla_W \ell(W) &= \nabla_W \sum_{i=1}^{m} \big(\log|W| - \frac{1}{2}d\log 2\pi - \frac{1}{2}{x^{(i)}}^T W^T W x^{(i)} \big) \\
&= \sum_{i=1}^{m} \big(W^{-T} - \frac{1}{2}\nabla_W \text{ tr } {x^{(i)}}^T W^T W x^{(i)} \big) \\
&= \sum_{i=1}^{m} \big(W^{-T} - \frac{1}{2}\nabla_W \text{ tr } W^T W x^{(i)} {x^{(i)}}^T \big) \\
&= \sum_{i=1}^{m} \big(W^{-T} - W x^{(i)} {x^{(i)}}^T \big) \\
&= nW^{-T} - WX^T X
\end{aligned}$$

Let $\nabla_W \ell(W) = 0$, then

$$\begin{aligned}
nW^{-T} &= WX^T X \\
W^T W &= n(X^T X)^{-1}
\end{aligned}$$

Let R be an arbitrary orthogonal matrix, then

$$\begin{aligned}(RW)^T RW &= W^T R^T RW \\ &= W^T I W \\ &= W^T W \qquad \text{for any } R\end{aligned}$$

So the result of W can be any of the RW.

(b)

For $x^{(i)}$, we have

$$\begin{aligned}\nabla_W \ell(W) &= \nabla_W \big(\log|W| + \sum_{j=1}^{d} \log g'(w_j^T x^{(i)}) \big) \\ &= \nabla_W \big(\log|W| + \sum_{j=1}^{d} \log \frac{1}{2} \exp(-|w_j^T x^{(i)}|) \big) \\ &= \nabla_W \big(\log|W| + \sum_{j=1}^{d} \log \frac{1}{2} - |w_j^T x^{(i)}| \big) \\ &= \nabla_W \big(\log|W| + d \log \frac{1}{2} - \sum_{j=1}^{d} |w_j^T x^{(i)}| \big) \\ &= W^{-T} - \nabla_W \sum_{j=1}^{d} |w_j^T x^{(i)}| \\ &= W^{-T} - \begin{bmatrix} \frac{\partial}{\partial w_1^T} |w_1^T x^{(i)}| \\ \cdots \\ \frac{\partial}{\partial w_d^T} |w_d^T x^{(i)}| \end{bmatrix} \\ &= W^{-T} - \begin{bmatrix} \operatorname{sign}(w_1^T x^{(i)}) {x^{(i)}}^T \\ \cdots \\ \operatorname{sign}(w_d^T x^{(i)}) {x^{(i)}}^T \end{bmatrix} \\ &= W^{-T} - \operatorname{sign}(W^T x^{(i)}) {x^{(i)}}^T\end{aligned}$$

Then,

$$W := W + \alpha \big(W^{-T} - \operatorname{sign}(W^T x^{(i)}) {x^{(i)}}^T \big)$$