DD2424 - Assignment 1 (Bonus)

Summary

I have completed both exercises for the bonus points. For the first part, I have ...

Oskar Stigland
DD2424
Spring 2023

Part I - Improving Performance of the Network

Part II - Multiple Binary Cross-Entropy Losses

Obtaining the gradient

In order to obtain the expression for $\partial\ell/\partial\boldsymbol{s}$, we consider the chain rule, where

$$\frac{\partial \ell}{\partial \boldsymbol{s}} = \frac{\partial \ell}{\partial \boldsymbol{p}}\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{s}}$$

First, we considering differentiating the class-specific loss w.r.t. the correspoding probability, s.t. $\ell = -1/K\sum_{k=1}^{K}\ell_k$:

$$\frac{\partial \ell_k}{\partial p_k} = \frac{\partial}{\partial p_k}\left[(1-y_k)\log(1-p_k) + y_k\log(p_k)\right] = \frac{y_k - 1}{1-p_k} + \frac{y_k}{p_k}$$

Second, we recognize that $p_k = \sigma(s_k)$, wherer $\sigma : \mathbb{R} \to [0,1]$ is the logistic sigmoid function, for which we also have that

$$\sigma(s) = \frac{1}{1+e^{-s}} = \frac{e^s}{e^s+1}$$

as well as $1-\sigma(s) = \sigma(-s)$ and $\sigma'(s) = \sigma(s)(1-\sigma(s))$. Hence, we get that

$$\begin{aligned}
\frac{\partial \ell_k}{\partial s_k} &= \frac{\partial}{\partial s_k}\left[(1-y_k)\log(1-\sigma(s_k)) + y_k\log(\sigma(s_k))\right] \\
&= \frac{y_k-1}{1-\sigma(s_k)}\sigma(s_k)(1-\sigma(s_k)) + \frac{y_k}{\sigma(s_k)}\sigma(s_k)(1-\sigma(s_k)) \\
&= (y_k-1)\sigma(s_k) + y_k(1-\sigma(s_k) \\
&= y_k - \sigma(s_k)
\end{aligned}$$

Hence, the Jacobians we're looking for are given by

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{p}} &= -\frac{1}{K}\left[\boldsymbol{y}^T\mathrm{diag}(\boldsymbol{p})^{-1} - (\boldsymbol{1}-\boldsymbol{y})^T\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p})^{-1}\right] \\
\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{s}} &= \mathrm{diag}(\boldsymbol{p})\,\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p})
\end{aligned}$$

such that

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{s}} &= -\frac{1}{K}\left[\boldsymbol{y}^T\underbrace{\mathrm{diag}(\boldsymbol{p})^{-1}\mathrm{diag}(\boldsymbol{p})\,\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p})}_{=\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p})} - (\boldsymbol{1}-\boldsymbol{y})^T\underbrace{\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p})^{-1}\mathrm{diag}(\boldsymbol{p})\,\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p})}_{=\mathrm{diag}(\boldsymbol{p})}\right] \\
&= -\frac{1}{K}\left[\boldsymbol{y}^T\mathrm{diag}(\boldsymbol{1}-\boldsymbol{p}) - (\boldsymbol{1}-\boldsymbol{y})^T\mathrm{diag}(\boldsymbol{p})\right] \\
&= -\frac{1}{K}\left[\boldsymbol{y} - \boldsymbol{y}^T\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p} + \boldsymbol{y}^T\mathrm{diag}(\boldsymbol{p})\right] \\
&= -\frac{1}{K}(\boldsymbol{y}-\boldsymbol{p})
\end{aligned}$$

Hence, using the result from the previous loss, we should have that

$$\frac{\partial \ell}{\partial \boldsymbol{W}} = -\frac{1}{K}(\boldsymbol{y}-\boldsymbol{p})\,\boldsymbol{x}^T, \quad \frac{\partial \ell}{\partial \boldsymbol{b}} = -\frac{1}{K}(\boldsymbol{y}-\boldsymbol{p})$$

Results