Exercise 5

1. Compute gradient of the regularized logistic loss ($L_2$)

$$l(w) = \underbrace{\sum_{n=0}^{N-1} \ln(1 + \exp(-y_n w^T x_n))}_{l_1(w)} + \underbrace{C w^T w}_{l_2(w)}$$

$C \geq 0$

---

$$l_1(w) = \sum_{n=0}^{N-1} \ln(1+\exp(-y_n w^T x_n))$$

$$\frac{\partial}{\partial w} l_1(w) = \frac{\partial}{\partial w} \sum_{n=0}^{N-1} \ln(1+\exp(-y_n w^T x_n)) =$$

$$= \sum_{n=0}^{N-1} \frac{\partial}{\partial w} \left[\ln(1+\exp(-y_n w^T x_n))\right] =$$

$$= \sum_{n=0}^{N-1} \frac{1}{1+\exp(-y_n w^T x_n)} \frac{\partial}{\partial w}(1+\exp(-y_n w^T x_n)) =$$

$$= \sum_{n=0}^{N-1} \frac{\exp(-y_n w^T x_n) \cdot \overbrace{\frac{\partial}{\partial w}(-y_n w^T x_n)}^{-y_n x_n}}{1+\exp(-y_n w^T x_n)} =$$

$$= -\sum_{n=0}^{N-1} \frac{\exp(-y_n w^T x_n) y_n x_n}{1+\exp(-y_n w^T x_n)}$$

$$l_2(w) = C w^T w$$

$$\frac{\partial}{\partial w} l_2(w) = \frac{\partial}{\partial w} C w^T w = C \frac{\partial}{\partial w} w^T w = C(2w^T) = 2Cw^T$$

[Scientific paper: $\alpha = x^T A x$ $\quad \frac{\partial \alpha}{\partial x} = 2x^T A \rightarrow A = I$
$A$: symmetric

---

$$\frac{\partial}{\partial w} l(w) = -\sum_{n=0}^{N-1} \frac{\exp(-y_n w^T x_n) y_n x_n}{1+\exp(-y_n w^T x_n)} + 2Cw^T$$

2. Manually compute the update step

Stochastic gradient $\rightarrow$ compute the gradient for one sample and apply gradient descent rule

$w^T = [2\ 1]$ $\quad x_n^T = [-1\ 1]$ $\quad y[n] = -1$ $\quad N = 1$

$$\frac{\partial}{\partial w} l(w) = -\frac{\exp\left(-(-1)[2\ 1]\begin{bmatrix}-1\\1\end{bmatrix}\right)(-1)\begin{bmatrix}-1\\1\end{bmatrix}}{1+\exp\left(-(-1)[2\ 1]\begin{bmatrix}-1\\1\end{bmatrix}\right)} + 2C\underset{1\times2}{[2\ 1]}\, I_{2\times2}$$

$$= -\frac{\exp\left(\overset{-2}{2(-1)} + 1.1\right)\begin{bmatrix}1\\-1\end{bmatrix}}{1+\exp(-1)} + 2C[2\ 1]\, I$$

$$= -\frac{e^{-1}}{1+e^{-1}}\begin{bmatrix}1\\-1\end{bmatrix} + 2C[2\ 1]$$

2.

$$w_{n+1} = w_n - \eta \frac{\partial \ell(w)}{\partial w} = w_n - \eta\left[(y_n \sigma(-y_n w_n^T x_n) - y_n\right] x_n + 2C w_n$$

$$\sigma = \frac{1}{1+\exp(-y_n w_n^T x_n)}$$

$$w = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad x_n = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad y_n = -1 \quad C = 1$$

$$-y_n w_n^T x_n = -(-1)\begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = -2+1 = -1$$

$$\sigma = \frac{1}{1+e^{-1}} = 0.731$$

$$w_{n+1} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \eta\left[(\underbrace{-0.731+1}_{0.2689})\right]\begin{bmatrix} -1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 2 \\ 1 \end{bmatrix} - 0.01\begin{bmatrix} -0.2689 \\ 0.2689 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$