ML - Homework 5

Tong Thi Van Anh

October 2022

1 Exercise 1

With $t \in \{0, 1\}$, $y(x) = \sigma(w^T x)$ and $\sigma(z) = \frac{1}{1+e^{-z}}$
$\Rightarrow y(x) = \sigma(w^T x) = \frac{1}{1+e^{-w^T x}}$

$$
\begin{aligned}
L(w) &= -\sum_{i=1}^{N} \Big(t_i \log(y_i)) + (1 - t_i) \log(1 - y_i) \Big) \\
\sigma'(z) &= \frac{d}{dz} \frac{1}{1 + e^{-z}} \\
&= \frac{1}{(1 + e^{-z})^2} (e^{-z}) \\
&= \frac{1}{1 + e^{-z}} . \left(1 - \frac{1}{1 + e^{-z}} \right) \\
&= \sigma(z)(1 - \sigma(z)).
\end{aligned}
$$

For each (x_i, y_i), we have the loss function:

$$
l = -(t_i \log(y_i) + (1 - t_i) \log(1 - y_i))
$$

$$
\begin{aligned}
\frac{\partial l}{\partial w_j} &= \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial w_j} \\
&= -\left(t_i \frac{1}{y_i} - (1 - t_i) \frac{1}{1 - y_i} \right) \frac{\partial}{\partial w_j} \sigma(w^T x_i) \\
&= -\left(t_i \frac{1}{\sigma(w^T x_i)} - (1 - t_i) \frac{1}{1 - \sigma(w^T x_i)} \right) \sigma(w^T x_i)(1 - \sigma(w^T x_i)) \frac{\partial}{\partial w_j} w^T x_i \\
&= -(t_i(1 - \sigma(w^T x_i)) - (1 - t_i)\sigma(w^T x_i)) x_{ij} \\
&= -(t_i - y_i) x_{ij}.
\end{aligned}
$$

$$
\Rightarrow \frac{\partial L}{\partial w} = -\sum_{i=1}^{N} (t_i - y_i) x_i = X^T (t - y)
$$

2 Exercise 5

2.1 Binary Cross Entropy: Convex

Using the result from Exercise 1, we find the Hessian of the loss function:

$$
\begin{aligned}
H_{jk} &= \frac{\partial^2 L}{\partial w_j \partial w_k} \\
&= -\sum_{i=1}^{N} -\frac{\partial y_i}{\partial w_k} x_{ij} \\
&= \sum_{i=1}^{N} y_i(1-y_i) x_{ij} x_{ik}
\end{aligned}
$$

$$
\Rightarrow H = \sum_{i=1}^{N} y_i(1-y_i) x_i^2
$$

For $y_i \in [0,1]$, we have: $y_i(1-y_i) \in [0, 1/4]$. Therefore, $H \geq 0 \Rightarrow$ Convex.

2.2 MSE: Non-Convex

The loss function:

$$
L = -\sum_{i=1}^{N} (t_i - y_i)^2
$$

For each (x_i, y_i), we have:

$$
l = -(t_i - y_i)^2
$$

$$
\begin{aligned}
\frac{\partial l}{\partial w_j} &= \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial w_j} \\
&= -(-2)(t_i - y_i) \frac{\partial}{\partial w_j} \sigma(w^T x_i) \\
&= 2(t_i - y_i) \sigma(w^T x_i)(1 - \sigma(w^T x_i)) \frac{\partial}{\partial w_j} w^T x_i \\
&= 2(t_i - y_i) y_i (1 - y_i) x_{ij} \\
\Rightarrow \frac{\partial L}{\partial w} &= 2 \sum_{i=1}^{N} (t_i - y_i) y_i (1 - y_i) x_i \\
&= 2 \sum_{i=1}^{N} (y_i^3 - y_i^2 - t_i y_i^2 + t_i y_i) x_i
\end{aligned}
$$

Calculating the Hessian matrix:

$$\begin{aligned}
H_{jk} &= \frac{\partial^2 L}{\partial w_j \partial w_k} \\
&= 2\sum_{i=1}^{N}(3y_i^2 - 2y_i - 2t_iy_i + t_i)\frac{\partial y_i}{\partial w_k}x_{ij} \\
&= 2\sum_{i=1}^{N}(3y_i^2 - 2y_i - 2t_iy_i + t_i)y_i(1-y_i)x_{ij}x_{ik} \\
\Rightarrow H &= 2\sum_{i=1}^{N}(3y_i^2 - 2y_i - 2t_iy_i + t_i)y_i(1-y_i)x_i^2
\end{aligned}$$

As $y_i(1-y_i) \in [0, 1/4]$ and $x_i^2 \geq 0$, we consider:

$$A = \sum_{i=1}^{N}(3y_i^2 - 2y_i - 2t_iy_i + t_i)$$

Since $t_i \in \{0, 1\}$, we have 2 cases:
When $t_i = 0$:

$$\begin{aligned}
A &= \sum_{i=1}^{N}(3y_i^2 - 2y_i) \\
&= \sum_{i=1}^{N} 3y_i(y_i - 2/3)
\end{aligned}$$

If y_i is in range [0, 2/3], $A \geq 0$ ($H \geq 0$); if y_i is in range [2/3, 1], $A \leq 0$ ($H \leq 0$).
It shows that the function is not convex.
When $t_i = 1$:

$$\begin{aligned}
A &= \sum_{i=1}^{N}(3y_i^2 - 2y_i - 2y_i + 1) \\
&= \sum_{i=1}^{N}(3y_i^2 - 4y_i + 1) \\
&= \sum_{i=1}^{N} 3(y_i - 1/3)(y_i - 1)
\end{aligned}$$

If y_i is in range [0, 1/3], $A \leq 0$ ($H \leq 0$); if y_i is in range [1/3, 1], $A \geq 0$ ($H \geq 0$).
It shows that the function is not convex.

Hence, MSE loss function for logistic regression is non-convex.