Machine Learning 1 - Week 5

Nguyễn Thị Kiều Nhung - 11203041

October 26, 2022

Problem 1.

We have Loss Function (log form)

$$L = -\sum_{i=1}^{N} \Big(t_i \log(y_i) + (1 - t_i) \log(1 - y_i) \Big) \tag{1}$$

where:

$$\begin{aligned} t &\in \{0; 1\} \\ y(x) &= \sigma(w^T x) \\ \sigma(z) &= \frac{1}{1 + e^{-z}} \end{aligned}$$

Sigmoid function and is a special one:

$$\begin{aligned} \sigma'(z) &= \frac{d}{dz} \frac{1}{1 + e^{-z}} && (2) \\ &= \frac{e^{-z}}{(1 + e^{-z})^2} && (3) \\ &= \frac{1}{1 + e^{-z}} \frac{e^{-z}}{1 + e^{-z}} && (4) \\ &= \sigma(z)\sigma(1 - \sigma(z)) && (5) \end{aligned}$$

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

$$l = -\Big(t_i \log(y_i) + (1 - t_i) \log(1 - y_i) \Big) \tag{6}$$

Apply Chain Rule:

$$\begin{aligned}
\frac{\partial l}{\partial w} &= \frac{\partial l}{\partial y_i}\frac{y_i}{\partial w} \\
&= -\Big(\frac{t_i}{y_i} - \frac{1-t_i}{1-y_i}\Big)\frac{\partial y_i}{\partial w} \\
&= -\Big(\frac{t_i}{y_i} - \frac{1-t_i}{1-y_i}\Big)\frac{\partial}{\partial w}\sigma(w^T x_i) \\
&= -\Big[\frac{t_i}{\sigma(w^T x_i)} - \frac{1-t_i}{1-\sigma(w^T x_i)}\Big]\sigma(w^T x_i)[1-\sigma(w^T x_i)] \times x_i \\
&= \Big[t_i(1-\sigma(w^T x_i)) - (1-t_i)\sigma(w^T x_i)\Big]x_i \qquad \text{(common denominator)} \\
&= \Bigg[t_i - t_i\sigma(w^T x_i) - \sigma(w^T x_i) + t_i\sigma(w^T x_i)]x_i \\
&= -\Bigg[t_i - \sigma(w^T x_i)]x_i \\
&= -(t_i - y_i)x_i
\end{aligned}$$

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

Problem 2,3,4: Coding

Problem 5.

5.1. Binary Cross Entropy - Convex

To prove a function is convex, we should show that its second-order derivatives is positive.

Using result from Exercise 1, computing $Hessian$ of loss function:

$$\begin{aligned}
H(w) &= \frac{\partial^2 L}{\partial w \partial w^t} \\
&= -\sum_{i=1}^{N} -\frac{\partial y_i}{\partial w}{x_i}^T \\
&= -\sum_{i=1}^{N} -y_i(1-y_i)x_i{x_i}^T \\
&= \sum_{i=1}^{N} y_i(1-y_i){x_i}^2
\end{aligned}$$

For

$$y \in [0;1] \implies y(1-y) \in [0;\frac{1}{4}] \implies H \geq 0 \implies \textbf{convex}$$

5.2. MSE - Non Convex

Loss mean squared error function:

$$L = -\sum_{i=1}^{N}(t_i - y_i)^2 \tag{7}$$

For each $(x_i; y_i)$:

$$l = -(t_i - y_i)^2 \tag{8}$$

First, Computing first order derivative:

(9)

Sum up,

$$\begin{aligned}
\frac{\partial L}{\partial w} &= 2\sum_{i=1}^{N}(t_i - y_i)y_i(1-y_i)x_i \\
&= 2\sum_{i=1}^{N}(t_i y_i - t_i {y_i}^2 - {y_i}^2 + {y_i}^3)x_i
\end{aligned}$$

Next, Computing Hessian:

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

As $y_i(1-y_i) \in [0; \frac{1}{4}]$ and ${x_i}^2 \geq 0$
We now consider $H_t = \sum_{i=1}^{N}(t_i - 2t_i y_i - 2y_i + 3{y_i}^2)$ where $t_i \in \{0; 1\}$

When $t_i = 0$:

$$\begin{aligned}
H_{t=0} &= \sum_{i=1}^{N}(-2y_i + 3{y_i}^2) \\
&= \sum_{i=1}^{N} 3y_i(y_i - \frac{2}{3})
\end{aligned}$$

If $y \in [\frac{2}{3}; 1]$, $H \geq 0$
If $y \in [0, \frac{2}{3}]$, $H \leq 0$

It implies that the function is non convex

When $t_i = 1$:

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

If $y \in [0; \frac{1}{3}]$, $H \geq 0$
If $y \in [\frac{1}{3}; 1]$, $H \leq 0$

It also implies that the function is non convex