Machine Learning 1 - Homework 5

Linh Nguyen

October 2022

1 Problems

1. Calculate vector calculus $\frac{dL}{dw}$
2. Using gradient descent algorithm, implement logistic regression algorithm.
3. Run the logistic regression implementation in exercise 2 for the dataset
4. Draw boundary line for 2 classes
5. Prove that with the logistic model, the loss binary-cross entropy is a convex function with W, the loss mean square error is a non-convex function with W

2 Solutions

2.1 Problem 1: Calculate vector calculus $\frac{dL}{dw}$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_{(2)}^{(1)} & \dots & x_m^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_1^{(n)} & x_2^{(n)} & \dots & x_m^{(n)} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \\ \dots \\ w_m \end{bmatrix}$$

$$\hat{y} = \sigma(Xw)$$

$$L = -\sum_{i=1}^{N}(y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i))$$

Let's calculate derivation of sigmoid function:

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

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

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

Apply Chain Rule:

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

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

2.2 Problem 5

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

2.2.2 MSE - Non Convex

Loss mean squared error function:

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

For each $(x_i; y_i)$:

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

First, Computing first order derivative:

$$
\begin{aligned}
\frac{\partial l}{\partial w} &= \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial w} \\
&= -(-2)(t_i - y_i) \frac{\partial}{\partial w} \sigma(w^t x_i) \\
&= -(-2)(t_i - y_i) \sigma(w^T x_i)[1 - \sigma(w^T x_i)] \frac{\partial}{\partial w} w^T x_i \\
&= 2(t_i - y_i) y_i (1 - y_i) x_i
\end{aligned}
$$

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:

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

As $y_i(1-y_i) \in [0; \frac{1}{4}]$ and ${x_i}^2 \geq 0$

We now consider $H_t = \sum_{i=1}^{N}(t_i - 2t_iy_i - 2y_i + 3{y_i}^2)$ 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