05 - Logistic Regression

Mai Tra My

October 26, 2022

1 Exercise 1

$$\sigma(x) = \frac{1}{1+e^{-x}}$$
$$\Rightarrow \sigma'(x) = \sigma(x)\sigma(1-x)$$

Loss function:

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

Let $X \in R^{n*d}$, $W \in R^{d*1}$, $y \in R^{n*1}$
At each point:

$$\begin{aligned}
\frac{\partial L_i}{\partial w} &= \frac{\partial L}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial w} \\
&= -\left(y_i\frac{1}{\hat{y}} + (1-y_i)\frac{-1}{1-\hat{y}_i}\right)\frac{\partial \hat{y}_i}{\partial w} \\
&= -\frac{y_i-\hat{y}_i}{\hat{y}(1-\hat{y}_i)}\frac{\partial \hat{y}_i}{\partial w}
\end{aligned}$$

And,

$$\begin{aligned}
\frac{\partial \hat{y}_i}{\partial w} &= \frac{\partial \hat{y}_i}{\partial (x_i w)}\frac{\partial (x_i w)}{\partial w} \\
&= \hat{y}_i(1-\hat{y}_i)x_i^T
\end{aligned}$$

$$\Rightarrow \frac{\partial L_i}{\partial w} = x_i^T(\hat{y}_i - y_i)$$

In general,

$$\frac{\partial L}{\partial w} = X^T(\hat{y} - y)$$

2 Section 5

1) Cross-entropy

$$H = \frac{\partial^2 L}{\partial W^2} = X^T\hat{y}(1-\hat{y})X$$

H is symmetric.
For $a \in R^d$

$$\begin{aligned} a^T H a &= a^T (X^T \hat{y}(1-\hat{y})X)a \\ &= (a^T X^T)\hat{y}(1-\hat{y})(Xa) \\ &= \sum_{i=1}^{N} \hat{y}_i(1-\hat{y}_i)(a^T X_i^T)^2 \end{aligned}$$

$$\Rightarrow a^T H a \geqslant 0$$

Loss binary cross-entropy is convex.

2) MSE

$$L = ||y - \hat{y}||^2$$