HW5 - Logistic Regression

Phương Quỳnh Chi

October 25, 2022

Exercise 1: Calculate vector calculus $\dfrac{\partial L}{\partial w} = X^T(\hat{y} - y)$

$$L = -log\ p(t|w) = -\sum_{i=1}^{N} y_i log(\hat{y_i}) + (1 - y_i)log(1 - \hat{y_i})$$
$$= -(y\ log\hat{y} + (1 - y)log(1 - \hat{y}))$$
$$\hat{y} = \sigma(X^T w) = \frac{1}{1 + e^{-X^T w}}\ ,\ z = e^{-X^T w}$$

We have:

$\dfrac{\partial L}{\partial w} = \dfrac{\partial L}{\partial \hat{y}}.\dfrac{\partial \hat{y}}{\partial z}.\dfrac{\partial z}{\partial w}$ (chain rule)

- $\dfrac{\partial L}{\partial \hat{y}} = -\Big(y.\dfrac{1}{\hat{y}} - (1 - y).\dfrac{1}{1 - \hat{y}}\Big) = -\Big(\dfrac{y}{\hat{y}} - \dfrac{1 - y}{1 - \hat{y}}\Big)$
- $\dfrac{\partial \hat{y}}{\partial w} = \dfrac{\partial \hat{y}}{\partial z}.\dfrac{\partial z}{\partial w}$
  $= -\dfrac{1}{(1 + z)^2}.(-Xe^{-X^T w}) = X.\dfrac{1}{1 + z}.\dfrac{z}{1 + z} = X\hat{y}(1 - \hat{y})$

$\Rightarrow \dfrac{\partial L}{\partial w} = -\Big(\dfrac{y}{\hat{y}} - \dfrac{1 - y}{1 - \hat{y}}\Big).X\hat{y}(1 - \hat{y})$
$= \dfrac{-y + y\hat{y} + \hat{y} - y\hat{y}}{\hat{y}(1 - \hat{y})}.X\hat{y}(1 - \hat{y})$
$= X(\hat{y} - y)$

Under the matrix form: $\dfrac{\partial L}{\partial w} = X^T(\hat{y} - y)$

Exercise 5:

Loss Mean Square Error: $J(w) = (y - \hat{y})^2$

$\hat{y} = \dfrac{1}{1 + e^-(w^T x + b)} \Rightarrow \dfrac{\partial \hat{y}}{\partial w} = x(1 - \hat{y})\hat{y}$

$\dfrac{\partial J}{\partial w} = \dfrac{\partial J}{\partial \hat{y}}.\dfrac{\partial \hat{y}}{\partial w} = -2(y - \hat{y}.x(1 - \hat{y})\hat{y}$
$= -2x(y - \hat{y})(\hat{y} - \hat{y}^2)$
$= -2x(y\hat{y} - y\hat{y}^2 - \hat{y}^2 + \hat{y}^3)$

$\dfrac{\partial^2 J}{\partial w^2} = -2x[x.y.\hat{y}(1 - \hat{y}) - 2x.y.\hat{y}.\hat{y}(1 - \hat{y}) + 3x.\hat{y}^2.\hat{y}.(1 - \hat{y})]$
$= -2x^2\hat{y}(1 - \hat{y})(y - 2y.\hat{y} - 2\hat{y} + 3\hat{y}^2)$

Having that:$x^2\hat{y}(1-\hat{y}) \geq 0, \hat{y} \in [0,1] \Rightarrow Consider f(\hat{y}) = -2(y-2.y.\hat{y}-2\hat{y}+3\hat{y}^2)$
Because y only takes 2 values 0,1

$$\Rightarrow \begin{cases} f(\hat{y}) & = 4\hat{y} - 6\hat{y}^2 \quad (1) \\ f(\hat{y}) & = -2 + 4\hat{y} + 4\hat{y} - 6\hat{y}^2 = -6\hat{y}^2 + 8\hat{y} - 2 \quad (2) \end{cases}$$

When $y \in \left[0, \frac{1}{3}\right]$, equation (1) ≤ 0
When $y \in \left[\frac{2}{3}, 1\right]$, equation (2) ≤ 0
$\Rightarrow \exists \hat{y} : f(\hat{y}) < 0$
$\Rightarrow \frac{\partial J^2}{\partial w^2}$ non convex

Cross-entropy:$L = -(y\ log\hat{y} + (1-y)log(1-\hat{y}))$
$\frac{\partial L}{\partial w} = (\hat{y} - y)x_i$
$\frac{\partial^2 L}{\partial w^2} = x^2\hat{y}(1-\hat{y})$
Having that $x^2\hat{y}(1-\hat{y}) \geq 0, \hat{y} \in [0,1]$

$\Rightarrow$ L is convex