Homework 5

Nguyễn Hoàng Long - 11202352 - DSEB 62

26/10/2022

1.Tính vector calculus dL/dW ($\mathbf{X}^T(\hat{y}-y)$)

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

$$\hat{y} = \sigma(X^T w) = \frac{1}{1+e^{-X^T w}}\ ,\ z = e^{-X^T w}$$

We have:

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

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

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

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

5. Chứng minh với model logistic thì loss binary-crossentropy là convex function với W, loss mean square error không là convex function với W

$$
\begin{aligned}
J(W) = (y-\hat{y})^2 . \hat{y} = \frac{1}{1+e^{-(w^T x+b)}} &=> \frac{\partial \hat{y}}{\partial w} = x.\hat{y}(1-\hat{y}) \\
+\frac{\partial J}{\partial w} &= \frac{\partial J}{\partial \hat{y}}.\frac{\partial \hat{y}}{\partial w} = -2(y-\hat{y}).x(1-\hat{y})\hat{y} \\
&= -2x(y-\hat{y})(\hat{y}-\hat{y}^2) \\
&= -2x(y\hat{y}-y\hat{y}^2-\hat{y}^2+\hat{y}^3) \\
+\frac{\partial^2 J}{\partial w^2} &= -2x[x.y.\hat{y}(1-\hat{y}) - 2.x.y\hat{y}.\hat{y}(1-\hat{y}) - 2x\hat{y}\hat{y}(1-\hat{y}) + 3x\hat{y}^2.\hat{y}(1-\hat{y})] \\
&= -2x^2.\hat{y}(1-\hat{y})(y-2y.\hat{y}-2\hat{y}+3\hat{y}^2)
\end{aligned}
$$

We have: $x^2\hat{y}(1-\hat{y}) \geq 0, \hat{y} \in [0,1]$ => Consider $f(\hat{y}) = -2(y-2y\hat{y}-2\hat{y}+3\hat{y}^2)$

Because y only takes 2 values $0, 1$

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

When $y \in [0; \frac{1}{3}]$, equation (1) ≤ 0
When $y \in [\frac{2}{3}; 1]$, equation (2) ≤ 0

=> $\exists \hat{y} : f(\hat{y}) < 0$

=> $\frac{\partial^2 J}{\partial w^2}$ non-convex

Cross-entropy:

$$
\begin{aligned}
L &= -ylog(\hat{y}) - (1-y)log(1-\hat{y}) \\
&=> \frac{\partial J}{\partial w} = (\hat{y}-y)x_i \\
&=> \frac{\partial^2 J}{\partial w^2} = x^2.\hat{y}(1-\hat{y})
\end{aligned}
$$

We have: $x^2\hat{y}(1-\hat{y}) \geq 0, \hat{y} \in [0,1]$
=> Convex