导数表达式

计算机视觉第五次作业 | 2101212840 游盈萱

问题描述

$$h = XW_1 + b_1$$

$$h_{sigmoid} = sigmoid(h)$$

$$Ypred = h_{sigmoid}W_2 + b_2$$

$$f = ||Y - Y_{pred}||_F^2$$

给出变量W_1, b_1, W_2, b_2导数表达式

推导过程

令$Z = XW$，$h = max(Z, 0)$，

于是，

$$\begin{aligned} f &= ||Y - Y_{pred}||_F^2 \\ &= tr[(Y - Y_{pred})^T(Y - Y_{pred})] \end{aligned}$$

方程两边取微分得到，

$$\begin{aligned} df &= d\{tr[(Y - Y_{pred})^T(Y - Y_{pred})]\} \\ &= tr\{d[(Y - Y_{pred})^T(Y - Y_{pred})]\} \end{aligned}$$

推导$\frac{\partial f}{\partial Y}$，

$$df = tr\{d[(Y - Y_{pred})^T(Y - Y_{pred})]\}$$

$$= tr[(dY)^T(Y - Y_{pred}) + (Y - Y_{pred})^T dY]\}$$

$$= tr[2(Y - Y_{pred})^T dY]$$

得到$\frac{\partial f}{\partial Y}$，

$$=> \ \frac{\partial f}{\partial Y} = 2(Y - Y_{pred})$$

推导$\frac{\partial f}{\partial Y_{pred}}$，

$$df = tr\{d[(Y - Y_{pred})^T(Y - Y_{pred})]\}$$

$$= tr[(-dY_{pred})^T(Y - Y_{pred}) + (Y - Y_{pred})^T(-dY_{pred})]\}$$

$$= tr[2(Y_{pred} - Y)^T dY_{pred}]$$

得到$\frac{\partial f}{\partial Y_{pred}}$，

$$=> \ \frac{\partial f}{\partial Y_{pred}} = 2(Y_{pred} - Y)$$

推导$\frac{\partial f}{\partial W_2}$

$$df = tr[2(Y_{pred} - Y)^T dY_{pred}]$$

$$= tr[2(Y_{pred} - Y)^T d(h_{sigmoid} W_2 + b_2)]$$

$$= tr[2(Y_{pred} - Y)^T h_{sigmoid} d(W_2)]$$

得到$\frac{\partial f}{\partial W_2}$，

$$=> \ \frac{\partial f}{\partial W_2} = 2h_{sigmoid}^T[(Y_{pred} - Y)]$$

推导$\frac{\partial f}{\partial b_2}$

$$df = tr[2(Y_{pred} - Y)^T dY_{pred}]$$

$$= tr[2(Y_{pred} - Y)^T d(h_{sigmoid} W_2 + b_2)]$$

$$= tr[2(Y_{pred} - Y)^T d(b_2)]$$

得到$\frac{\partial f}{\partial b_2}$，

$$=> \ \frac{\partial f}{\partial b_2} = 2(Y_{pred} - Y)$$

推导$\frac{\partial f}{\partial h_{sigmoid}}$

$$df = tr[2(Y_{pred} - Y)^T dY_{pred}]$$

$$= tr[2(Y_{pred} - Y)^T d(h_{sigmoid}W_2 + b_2)]$$

$$= tr[2(Y_{pred} - Y)^T d(h_{sigmoid})W_2]$$

$$= tr[2W_2(Y_{pred} - Y)^T d(h_{sigmoid})]$$

得到$\frac{\partial f}{\partial h_{sigmoid}}$，

$$=> \quad \frac{\partial f}{\partial h_{sigmoid}} = 2(Y_{pred} - Y)W_2^T$$

推导$\frac{\partial f}{\partial h}$

$$df = tr[\frac{\partial f}{\partial h_{sigmoid}}^T dh_{sigmoid}]$$

$$= tr[\frac{\partial f}{\partial h_{sigmoid}}^T sigmoid'(h) \odot dh]$$

$$= tr[(\frac{\partial f}{\partial h_{sigmoid}} \odot sigmoid'(h))^T dh]$$

$$= tr[(2(Y_{pred} - Y)W_2^T \odot sigmoid'(h))^T dh]$$

$$\frac{\partial f}{\partial h} = 2(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$$

推导$\frac{\partial f}{\partial W_1}$

由$\frac{\partial f}{\partial h} = 2(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$

$$df = tr[2[(Y_{pred} - Y)W_2^T \odot sigmoid'(h)]^T dh]$$

$$= tr[2[(Y_{pred} - Y)W_2^T \odot sigmoid'(h)]^T d(XW_1 + b_1)]$$

$$= tr[2[(Y_{pred} - Y)W_2^T \odot sigmoid'(h)]^T Xd(W_1)]$$

得到$\frac{\partial f}{\partial W_1}$，

$$=> \quad \frac{\partial f}{\partial W_1} = 2X^T(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$$

推导 $\frac{\partial f}{\partial b_1}$

由 $\frac{\partial f}{\partial h} = 2(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$

$$df = tr[2[(Y_{pred} - Y)W_2^T \odot sigmoid'(h)]^T dh]$$

$$= tr[2[(Y_{pred} - Y)W_2^T \odot sigmoid'(h)]^T d(XW_1 + b_1)]$$

$$= tr[2[(Y_{pred} - Y)W_2^T \odot sigmoid'(h)]^T d(b_1)]$$

得到 $\frac{\partial f}{\partial b_1}$，

$$=> \quad \frac{\partial f}{\partial b_1} = 2(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$$

综上所述

$$\frac{\partial f}{\partial W_1} = 2X^T(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$$

$$\frac{\partial f}{\partial b_1} = 2(Y_{pred} - Y)W_2^T \odot sigmoid'(h)$$

$$\frac{\partial f}{\partial W_2} = 2h_{sigmoid}^T[(Y_{pred} - Y)]$$

$$\frac{\partial f}{\partial b_2} = 2(Y_{pred} - Y)$$