Marin Marius

Tema

1) Demonstrați că $P(y=0) = \sigma(-(w\cdot x+b))$

$z = w\cdot x + b$

$y = \sigma(z) = \frac{1}{1+e^{-z}}$

$P(y=1) = \sigma(w\cdot x + b)$

$P(y=0) = 1 - P(y=1) = 1 - \frac{1}{1+e^{-z}} = \frac{1+e^{-z}-1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}} = \frac{\frac{1}{e^{z}}}{1+\frac{1}{e^{z}}}$

$= \frac{1}{\cancel{e^{z}}} \cdot \frac{\cancel{e^{z}}}{e^{z}+1} = \frac{1}{1+e^{z}} = \sigma(-z) = \sigma(-(w\cdot x+b))$

2) a) Demonstrați că $\Delta_b = y_{pred} - y_{true}$

b) Demonstrați că $\Delta_w = (y_{pred} - y_{true})\cdot x$

Notăm $y_{pred} = \hat{y}$, $y_{true} = y$

$L_{CE}(\hat{y}, y) = -[y \log \hat{y} + (1-y)\log(1-\hat{y})]$

$\Delta_b = \frac{\partial L_{CE}(\hat{y}, y)}{\partial b}$, $\Delta w = \frac{\partial L_{CE}(\hat{y}, y)}{\partial w}$

L_{CE} este funcție de $\hat{y}$, $\hat{y}$ este funcție de z, z este fu[nctie] de w

Aplicăm regula lanțului

$$\Delta_b = \frac{\partial L_{CE}(\hat{y}, y)}{\partial b} = \frac{\partial}{\partial \hat{y}} L_{CE}(y, \hat{y}) \cdot \frac{\partial}{\partial z} \cdot \hat{y}(z) \cdot \frac{\partial z}{\partial b} =$$

$$= -\left(y \cdot \frac{1}{\hat{y}} + (1-y) \cdot \frac{-1}{1-\hat{y}}\right) \cdot \frac{\partial}{\partial z} \sigma(z) \cdot \frac{\partial z}{\partial b} =$$

$$= -\left(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}\right) \cdot \frac{-1}{(1+e^{-z})^2} \cdot e^{-z} \cdot (-1) \cdot \frac{\partial(w \cdot x + b)}{\partial b} =$$

$$= \frac{-y(1-\hat{y}) - \hat{y}(1-y)}{\hat{y}(1-\hat{y})} \cdot \frac{e^{-z}}{(1+e^{-z})^2} \cdot 1 =$$

$$= -\frac{y - \cancel{y\hat{y}} - \hat{y} + \cancel{\hat{y}y}}{\hat{y}(1-\hat{y})} \cdot \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} =$$

$$= \frac{\hat{y}-y}{\hat{y}(1-\hat{y})} \cdot \underbrace{\frac{1}{1+e^{-z}}}_{\sigma(z) = \hat{y}} \cdot \left(1 - \frac{1}{1+e^{-z}}\right) = \frac{\hat{y}-y}{\cancel{\hat{y}(1-\hat{y})}} \cdot \cancel{\hat{y} \cdot (1-}$$

$$= \frac{\hat{y}-y}{1} = \hat{y} - y$$

b) $$\Delta w = \frac{\partial L_{CE}(\hat{y}, y)}{\partial w} = \frac{\partial}{\partial \hat{y}} L_{CE}(\hat{y}, y) \cdot \frac{\partial}{\partial z} \hat{y}(z) \cdot \frac{\partial z}{\partial w} =$$

$$= \frac{\hat{y}-y}{\hat{y}(1-\hat{y})} \cdot \hat{y}(1-\hat{y}) \cdot \frac{\partial(w \cdot x + b)}{\partial w} = (\hat{y}-y) \cdot x$$

Ex 2

$$z = w^T x + b$$

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

$$L(\hat{y}, y) = -(y \log \hat{y} + (-y) \log(1-\hat{y}))$$

$$\nabla_w L = (\hat{y} - y) x$$

$$\nabla_b L = \hat{y} - y$$

$$w \leftarrow w - \eta \nabla_w L$$

$$b \leftarrow b - \eta \nabla_b L$$

$$z = W^T x + b$$

$$\hat{y} = \text{softmax}(z) \qquad y_i = \frac{\exp(z_i)}{\sum_j \exp z_j} \qquad i = 0, \dots n$$

$$\hat{y} = \left(\frac{e^{z_1}}{\sum_j e^{z_j}}, \frac{e^{z_2}}{\sum_j e^{z_j}}, \dots, \frac{e^{z_n}}{\sum e^{z_j}} \right)^T$$

$$x = [1, 3, 0] \qquad b = [0,1 \quad 0,1 \quad 0,1]$$

$$W = \begin{bmatrix} 0,3 & 0,1 & -2 \\ -0,6 & -0,5 & 2 \\ -1 & -0,5 & 0,1 \end{bmatrix} \qquad y = [0, 1, 0]$$

$$z = W^T x + b = \begin{pmatrix} 0,3 & -0,6 & -1 \\ 0,1 & -0,5 & -0,5 \\ -2 & 2 & 0,1 \end{pmatrix}_{3\times3} \cdot \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}_{3\times1} + \begin{pmatrix} 0,1 \\ 0,1 \\ 0,1 \end{pmatrix} =$$

$$= \begin{pmatrix} -1,5 \\ -1,4 \\ 4 \end{pmatrix} + \begin{pmatrix} 0,1 \\ 0,1 \\ 0,1 \end{pmatrix} = \begin{pmatrix} -1,4 \\ -1,3 \\ 4,1 \end{pmatrix}$$

$$\sum_j e^{z_j} = e^{z_1} + e^{z_2} + e^{z_3} = e^{-1,4} + e^{-1,3} + e^{4,1} =$$

$$= 0,2465 + 0,2725 + 60,340 = 60,86$$

$$\hat{y} = \left(\frac{e^{-1,4}}{60,86}, \frac{e^{-1,3}}{60,86}, \frac{e^{4,1}}{60,86} \right)^T = \left(\frac{0,2465}{60,86}, \frac{0,2725}{60,86}, \frac{60,34}{60,86} \right.$$

$$\hat{y} = (0,004;\ 0,0044;\ 0,991)^T$$

$$\nabla_z L = \hat{y} - y = \begin{pmatrix} 0,004 \\ 0,0044 \\ 0,991 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,004 \\ -0,996 \\ 0,991 \end{pmatrix}$$

$$\nabla_W L = \nabla_z L \cdot x^T = \begin{pmatrix} 0,004 \\ -0,996 \\ 0,991 \end{pmatrix} \cdot (1, 3, 0) = \begin{pmatrix} 0,004 & 0,012 & 0 \\ -0,996 & -2,988 & 0 \\ 0,991 & 2,973 & 0 \end{pmatrix}$$

$$\nabla_b L = \nabla_z L = \begin{pmatrix} 0,004 \\ -0,996 \\ 0,991 \end{pmatrix}$$

$$W = W - \eta \nabla_W L = \begin{pmatrix} 0,3 & 0,1 & -2 \\ -0,6 & -0,5 & 2 \\ -1 & -0,5 & 0,1 \end{pmatrix} - \eta \cdot \begin{pmatrix} 0,004 & 0,012 \\ -0,996 & -2,998 \\ 0,991 & 2,973 \end{pmatrix}$$

Ex. pt $\eta = 0,2$, obtinem:

$$W = \begin{pmatrix} 0,3 & 0,1 & -2 \\ -0,6 & -0,5 & 2 \\ -1 & -0,5 & 0,1 \end{pmatrix} - \begin{pmatrix} 0,0008 & 0,0024 & 0 \\ -0,1992 & -0,5996 & 0 \\ 0,1982 & 0,5946 & 0 \end{pmatrix}$$

$$W = \begin{pmatrix} 0,2992 & 0,0976 & -2 \\ -0,4008 & 0,0996 & 2 \\ -1,1982 & -1,0946 & 0,1 \end{pmatrix}$$

$$b = b - \eta \nabla_b L \Rightarrow b = \begin{pmatrix} 0,1 \\ 0,1 \\ 0,1 \end{pmatrix} - 0,2 \cdot \begin{pmatrix} 0,004 \\ -0,996 \\ 0,991 \end{pmatrix} = \begin{pmatrix} 0,1 - 0,0008 \\ 0,1 + 0,1992 \\ 0,1 - 0,1982 \end{pmatrix} =$$

$$\Rightarrow b = \begin{pmatrix} 0,0992 \\ 0,2992 \\ -0,0982 \end{pmatrix}$$