年 月 日 第 次作业

May 8, 2020.

Guorui Shen

guorui233@outlook.com

The difference between Cross Entropy & Logistic Regression Loss

Cross Entropy loss function

(eg. three-class)

for ~~multiclass~~ classification problem.

Data points $\{x_i, y_i\}_{i=1}^{m}$, $x_i \in R^n$, $\cancel{y_i = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}}$ ~~or~~ $\cancel{\begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix}}$ ~~or .. or~~ $\cancel{\begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix}}$ $y_i = \begin{bmatrix}1\\0\\0\end{bmatrix}$ or $\begin{bmatrix}0\\1\\0\end{bmatrix}$ or $\begin{bmatrix}0\\0\\1\end{bmatrix}$

$\cancel{\hat{y}_{i1}, \hat{y}_{i1} = \text{softmax}(z_1, z_2, z_3)}$

$(\cancel{\hat{y}_1, \hat{y}_2, \hat{y}_3}) = \text{softmax}(z_1, z_2, z_3)$

$$\hat{y}_i = (\hat{y}_{i1}, \hat{y}_{i2}, \hat{y}_{i3}) = \left(\frac{e^{z_1}, e^{z_2}, e^{z_3}}{e^{z_1} + e^{z_2} + e^{z_3}}\right)$$

$$\hat{y}_{i1} + \hat{y}_{i2} + \hat{y}_{i3} = 1$$

Loss:

$$\frac{1}{m}\sum_{i=1}^{m} \underbrace{y_{i1}\cdot\log\frac{1}{\hat{y}_{i1}} + y_{i2}\cdot\log\frac{1}{\hat{y}_{i2}} + y_{i3}\cdot\log\frac{1}{\hat{y}_{i3}}}_{\text{only one term is non-zero}}$$

Logistic Loss function

for binary classification.

$\{x_i, y_i\}_{i=1}^{m}$, $x_i \in R^n$,

$y_i = \begin{bmatrix}1\\0\end{bmatrix}$ or $\begin{bmatrix}0\\1\end{bmatrix}$

Notice: It's not $y_i = 0$ or 1.

$$\hat{y}_i = (\hat{y}_{i1}, \hat{y}_{i2})$$

$$= (\text{Sigmoid}(\theta^T x_i), 1 - \text{Sigmoid}(\theta^T x_i))$$

$$= \left(\frac{1}{1+e^{-\theta^T x_i}}, \frac{e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right)$$

$$\hat{y}_{i1} + \hat{y}_{i2} = 1$$

loss:

$$\frac{1}{m}\sum_{i=1}^{m} \overbrace{y_{i1}\cdot\log\frac{1}{\hat{y}_{i1}} + (1-y_{i1})\log\frac{1}{1-\hat{y}_{i1}}}^{\text{only one term is non-zero.}}$$

怎么画出这部分神经网络图??