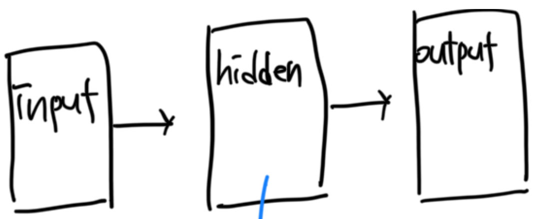

활성화함수: Sigmoid

* loss function
SoftmaxWithLoss
(softmax + CEE

순서: Affine → Sigmoid → Affine → SoftmaxWithLoss

$z_1 = xW_1 + b_1 \quad a_1 = \sigma(z_1) \quad z_2 = xW_2 + b_2 \quad y = \mathrm{Softmax}(z_2)$

$x = [1.0, 2.0]$

$$W_1 = \begin{pmatrix} 0.1 & -0.2 \\ 0.4 & 0.3 \end{pmatrix}, \; b_1 = (0.0, 0.0)$$

$$W_2 = \begin{pmatrix} 0.5 & -0.1 \\ -0.3 & 0.2 \end{pmatrix}, \; b_2 = (0.0, 0.0)$$

① forward

① 입력층 → 은닉층

$$z_1 = xW_1 + b_1 = (1.0, 2.0)\begin{pmatrix} 0.1 & -0.2 \\ 0.4 & 0.3 \end{pmatrix} + (0.0, 0.0) = (0.9 \;\; 0.4)$$

② Sigmoid 적용

$$\sigma(x) = \frac{1}{1+e^{-x}} \Rightarrow \left(\frac{1}{1+e^{-0.9}}, \frac{1}{1+e^{-0.4}}\right) \fallingdotseq (0.710, 0.598)$$

③ 은닉층 → 입력층

$$z_2 = xW_2 + b_2 = (0.710, 0.598)\begin{pmatrix} 0.5 & -0.1 \\ -0.3 & 0.2 \end{pmatrix}(0.0, 0.0) = (0.176, 0.049)$$

정답레이블 (T)

$T = (1 \;\; 0)$

④ Softmax 변환 (출력 확률 계산)

$$y = \mathrm{Softmax}(z_2) \quad y_i = \frac{e^{z_i}}{\sum e^{z_i}} = \frac{e^{(0.176 \;\; 0.049)}}{e^{0.176} + e^{0.049}}$$

$$\Rightarrow y_1 = \frac{1.192}{2.242} \approx 0.532, \quad y_2 = \frac{1.050}{2.242} \approx 0.468$$

· 교차 엔트로피 오차 계산

$\ldots \log y_2) = -(1 \cdot \log 0.532 + 0 \cdot \log 0.468)$

$L = -(t_1 \log y_1 + t_2 \cdots$

$\fallingdotseq 0.631$

② 역전파 : 손실을 기준으로 미분 . SoftmaxWithLoss.backward()로 dout을 전파하고 각 layer에서 backward() 실행해 기울기 갱신

5-1. 출력층 → 은닉층 : Affine.backward()에서 W2. b2 의 기울기계산

5-2. 은닉층 → 출력층 : Sigmoid.backward() 의 미분값 전달 후,

Affine.backward()에서 W1, b1의 기울기 계산

1) Softmax + CrossEntropy 역전파

$$\frac{\partial L}{\partial Z_2} = Y - T = (0.532 \quad 0.468) - (1 \quad 0) = (-0.468 \quad 0.468) \Rightarrow dZ_2$$

2) Affine 역전파 (output)

$$\begin{cases} dW_2 = A_1^T dZ_2 \\ db_2 = dZ_2 \\ dA_1 = dZ_2 W_2^T \end{cases}$$

$$dW_2 = \begin{pmatrix} 0.710 \\ 0.598 \end{pmatrix} (-0.468 \quad 0.468) = \begin{pmatrix} -0.332 & 0.332 \\ -0.280 & 0.280 \end{pmatrix}$$

$$db_2 = (-0.468 \quad 0.468)$$

$$dA_1 = (-0.468 \quad 0.468) \begin{pmatrix} 0.5 & -0.1 \\ -0.3 & 0.2 \end{pmatrix}^T = (-0.374 \quad 0.140)$$

3) Sigmoid 역전파

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) = A_1(1 - A_1)$$: Sigmoid 미분공식

$$= (0.710 \cdot (1 - 0.710) \quad 0.598(1 - 0.598))$$

$$= (0.206 \quad 0.240)$$

$$dZ_1 = dA_1 \cdot \sigma'(Z_1)$$

$$= (-0.374 \quad 0.140) \cdot (0.206 \quad 0.240)$$

$$= (-0.077 \quad 0.034)$$

4) Affine 역전파 (입력층)

[illegible]

$$\begin{cases} dW_1 = X^T dZ_1 \\ db_1 = dZ_1 \\ dX = dZ_1 W_1^T \end{cases}$$

$$dW_1 = \begin{pmatrix} 1.0 \\ 2.0 \end{pmatrix} \begin{pmatrix} -0.077 & 0.034 \end{pmatrix} = \begin{pmatrix} -0.077 & 0.034 \\ -0.154 & 0.068 \end{pmatrix}$$

$$db_1 = \begin{pmatrix} -0.077 & 0.034 \end{pmatrix}$$

$$dX = \begin{pmatrix} -0.077 & 0.034 \end{pmatrix} \begin{pmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{pmatrix}^T = \begin{pmatrix} 0.0059 & 0.0256 \end{pmatrix}$$

dW_2 : 출력층 가중치 기울기

$$\begin{pmatrix} -0.332 & 0.332 \\ -0.280 & 0.280 \end{pmatrix}$$

db_2 : 출력층 편향 기울기

$$\begin{pmatrix} -0.468 & 0.468 \end{pmatrix}$$

dW_1 : 은닉층 가중치 기울기

$$\begin{pmatrix} -0.077 & 0.034 \\ -0.154 & 0.068 \end{pmatrix}$$

db_1 : 은닉층 편향 기울기

$$\begin{pmatrix} -0.077 & 0.034 \end{pmatrix}$$