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

$$\begin{aligned}
\frac{d\,\sigma(x)}{dx} &= \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) \\
&= \frac{d}{dx}\left(1+e^{-x}\right)^{-1} \\
&= -\left(1+e^{-x}\right)^{-2} \times \frac{d}{dx}\left(1+e^{-x}\right) \\
&= -\left(1+e^{-x}\right)^{-2} \times \left(0 + \frac{d}{dx} e^{(-x)}\right) \\
&= -\left(1+e^{-x}\right)^{-2} \times \left(e^{-x} \times \frac{d}{dx}(-x)\right) \\
&= -\left(1+e^{-x}\right)^{-2} \times e^{-x} \times -1 \\
&= e^{-x}\left(1+e^{-x}\right)^{-2} \\
&= \frac{e^{-x}}{\left(1+e^{-x}\right)^{2}} \\
&= \frac{(e^{-x}+1) - 1}{(1+e^{-x})(1+e^{-x})} \\
&= \frac{\cancel{(e^{-x}+1)}}{\cancel{(1+e^{-x})}(1+e^{-x})} - \frac{1}{(1+e^{-x})(1+e^{-x})} \\
&= \frac{1}{(1+e^{-x})}\left[1 - \frac{1}{(1+e^{-x})}\right] = \sigma(x)\left(1-\sigma(x)\right)
\end{aligned}$$

$$J(\hat{y}, y) = -\left(y \log \hat{y} + (1-y)\log(1-\hat{y})\right)$$

where $\hat{y} = \sigma(\theta^T x)$

$$\frac{\partial J(\hat{y}, y)}{\partial \hat{y}} = -\left[\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right]$$

Now, we need

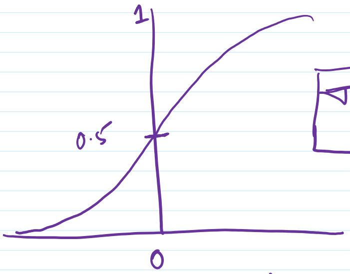
$$\begin{aligned}
\frac{\partial J(\hat{y}, y)}{\partial \theta} &= \frac{\partial J(\hat{y}, y)}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta} \qquad \boxed{\hat{y} = \sigma(\underbrace{\theta^T x}_{z})} \\
&= \underbrace{\frac{\partial J(\hat{y}, y)}{\partial \hat{y}}}_{\checkmark} \cdot \underbrace{\frac{\partial \hat{y}}{\partial z}}_{\checkmark} \cdot \underbrace{\frac{\partial z}{\partial \theta}}_{\frac{\partial \theta^T x}{\partial \theta} = x} \\
&= -\left[\frac{y}{\hat{y}} + \frac{(1-y)}{(1-\hat{y})}\right] \cdot \hat{y}(1-\hat{y}) \cdot x \\
&= -\left[y(1-\hat{y}) - \hat{y}(1-y)\right] \cdot x \\
&= -\left(y - \cancel{y\hat{y}} - \hat{y} + \cancel{\hat{y}y}\right) \cdot x \\
&= -(y-\hat{y}) \cdot x \\
&= -\left[y - \sigma(\theta^T x)\right] \cdot x
\end{aligned}$$

$$\theta_{new} = \theta_{old} - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\theta_{new} = \theta_{old} - \alpha \left[ y - \sigma(\theta^T x) \right] . x$$

## Multi-class Cross-Entropy Loss :-

Sigmoid / Logistic Function :-




$$\boxed{\sigma(x) = \frac{1}{1+e^{-x}}}$$

## "SOFTMAX FUNCTION"

function that squashes a vector in the range (0, 1) and the sum of resulting values is 1.

| Sigmoid (2 classes) | Softmax ($k > 2$ classes) |
| --- | --- |

Let $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$

$$S = \frac{e^{x_j}}{\sum_{j=1}^{k} e^{x_j}}$$

Example :-

| | | |
| --- | --- | --- |
| class 1 | 1.25 | $e^{1.25} = 3.49$ |
| class 2 | 2.44 | $e^{2.44} = 11.47$ |
| class 3 | 0.78 | $e^{0.78} = 2.18$ |
| class 4 | 0.12 | $e^{0.12} = 1.27$ |

$$\sum_{j=1}^{k} e^{x_j} = 3.49 + 11.47 + 2.18 + 1.24 = 18.41$$

$$S(\text{class } 1) = \frac{e^{1.125}}{\sum_{j=1}^{k} e^{x_j}} = \frac{3.49}{18.41}$$

Categorical cross-entropy loss :-

$$CE = - \sum_{i=1}^{n \to \#\text{ class}} y_i \log(\hat{y}_i)$$

ground

ground truth

predicted score

Let $n=2$,

binary cross-entropy loss :-

$$BCE = -\left(\sum_{i=1}^{2} y_i \log \hat{y}_i\right)$$

$$= -y_1 \log \hat{y}_1 - (y_2 \log \hat{y}_2)$$

$$y_2 = (1-y_1)$$

$$\hat{y}_2 = (1-\hat{y}_2)$$

$$\Rightarrow BCE = -y_1 \log \hat{y}_1 - \left((1-y_2) \log(1-\hat{y}_2)\right)$$