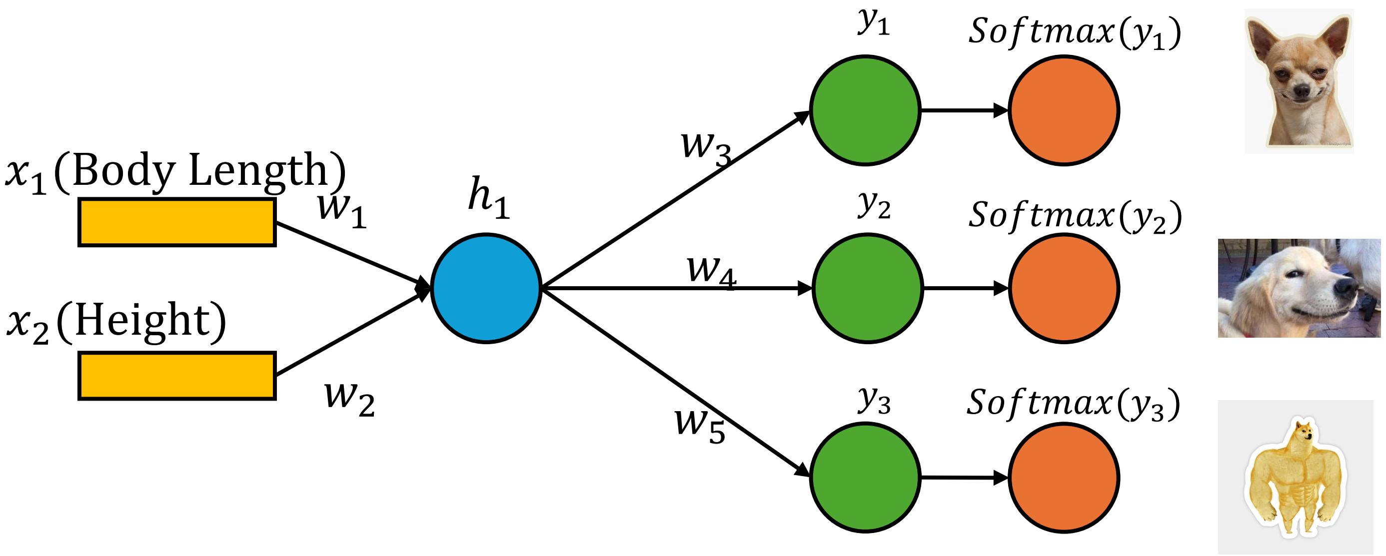
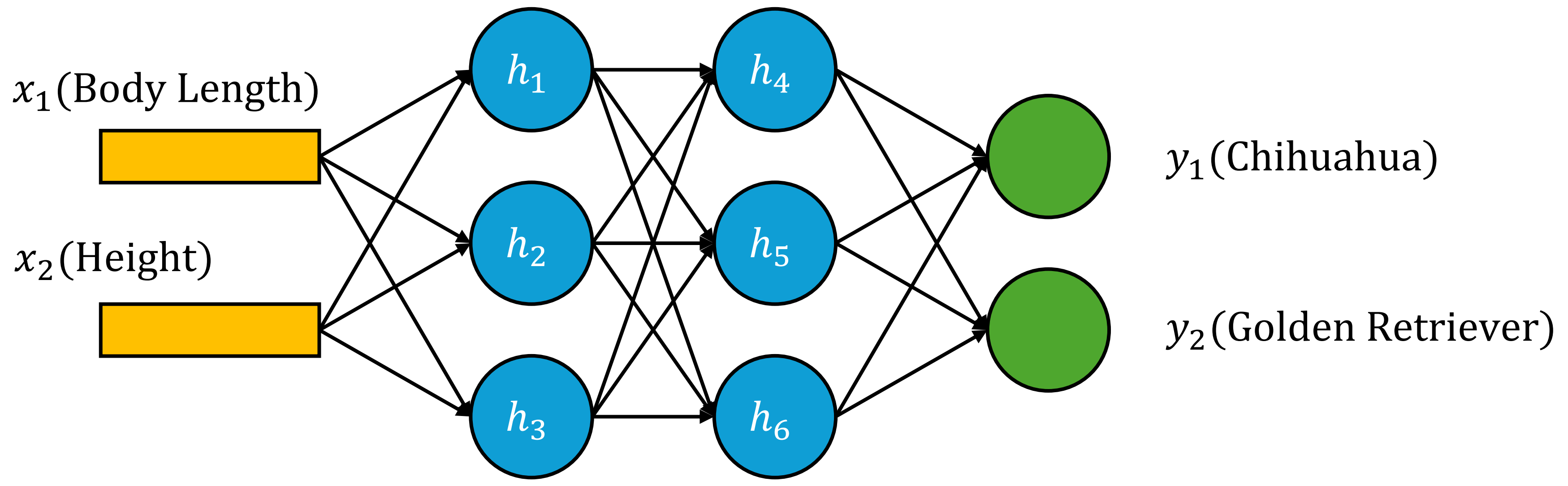
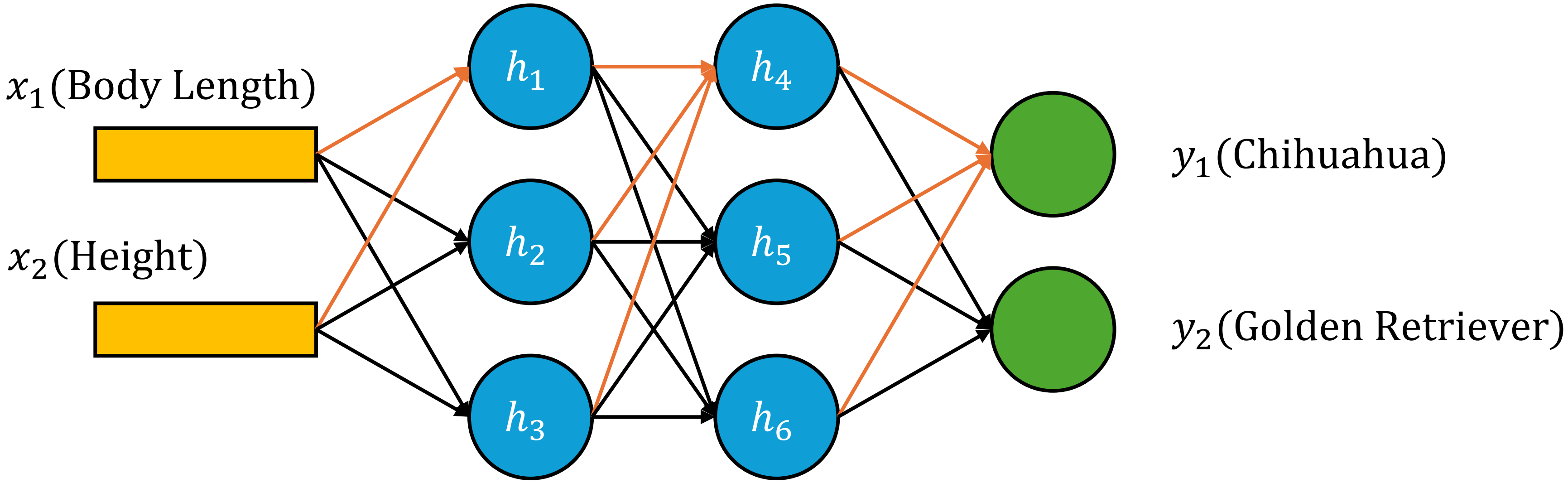
Classification Neural Network




x_1(Body Length)
x_2(Height)
h_1
h_2
h_3
h_4
h_5
h_6
y_1(Chihuahua)
y_2(Golden Retriever)



$$h_1 = (x_1 w_1 + x_2 w_4) + b_1$$

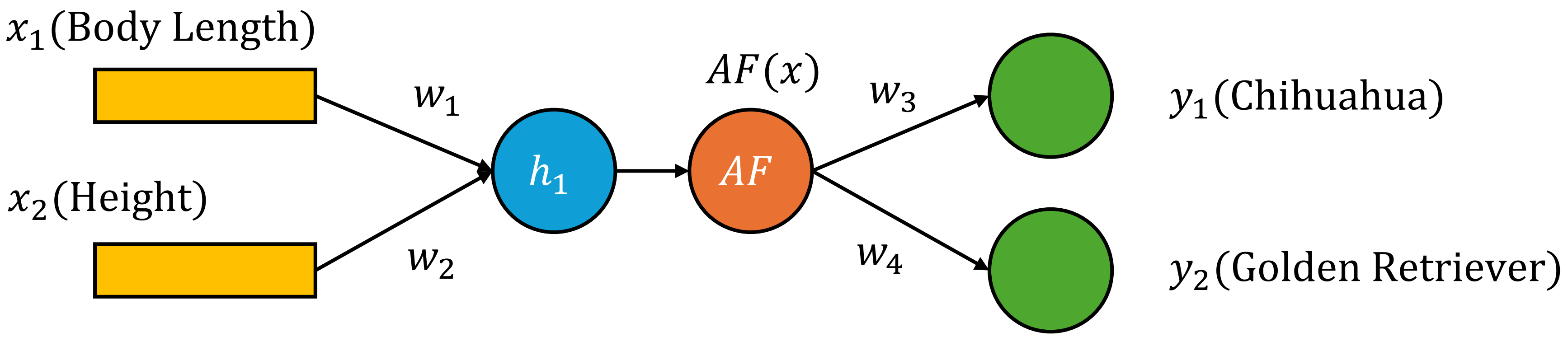
$$h_4 = (h_1 w_7 + h_2 w_{10} + h_3 w_{13}) + b_4$$ → Hey dude, I still linearly.

$$y_1 = (h_4 w_{16} + h_5 w_{18} + h_6 w_{20})$$

Just Apply Activation Function Into the NN

We simply want to hand on the neural network to handle non-linear situations.

Plug the AF in



$$h_1 = (x_1 w_1 + x_2 w_2) + b_1$$

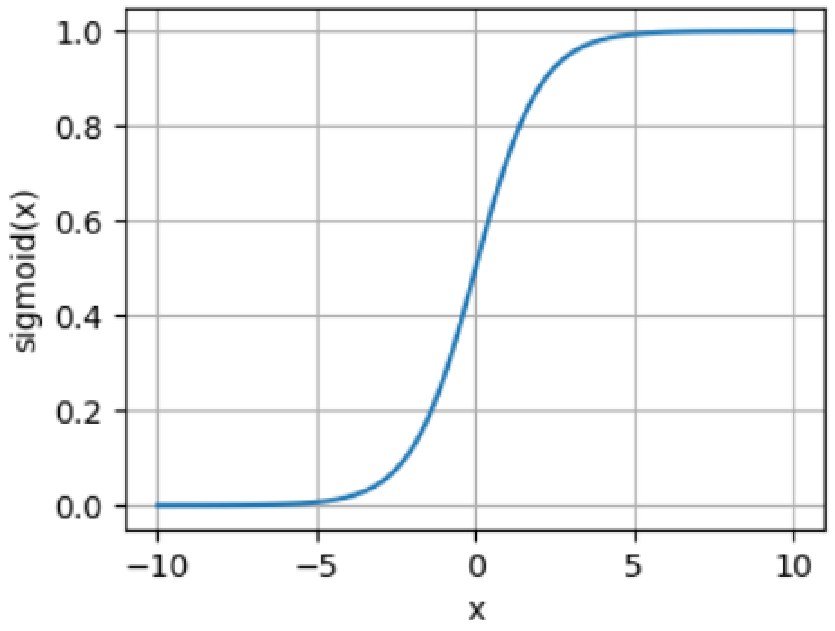
$$\mathrm{AF} = AF(h_1)$$

$$y_1 = AF(h_1) \cdot w_3$$
$$y_2 = AF(h_1) \cdot w_4$$

History of Activation Functions

Logistic Sigmoid

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



This AF squashes the output between [0, 1]

Derivative of Sigmoid

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

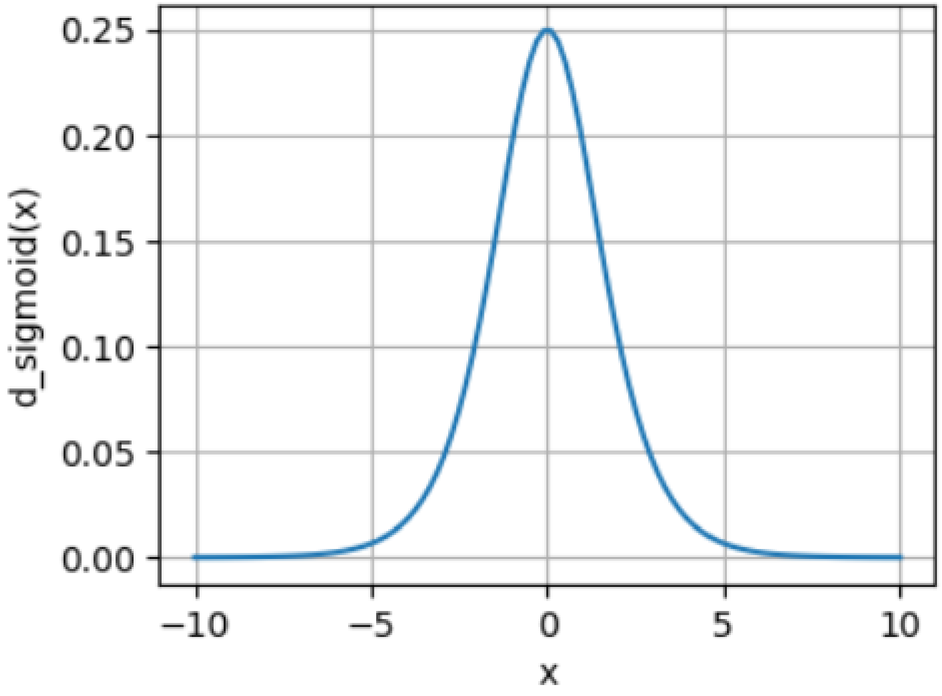
$$\frac{dSigmoid(x)}{dx} = \frac{d(\frac{1}{1+e^{-x}})}{dx}$$

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

$$= -(1+e^{-x})^{-2} \cdot -(e^{-x})$$

$$= \frac{1}{(1+e^{-x})} \cdot \left[1 - \frac{1}{(1+e^{-x})}\right]$$

$$= Sigmoid(x) \cdot (1 - Sigmoid(x))$$



$max = 0.25$

Logistic Sigmoid Problem

Logistic Sigmoid function is saturated for higher and lower inputs, which leads to "**vanishing gradient problem**".

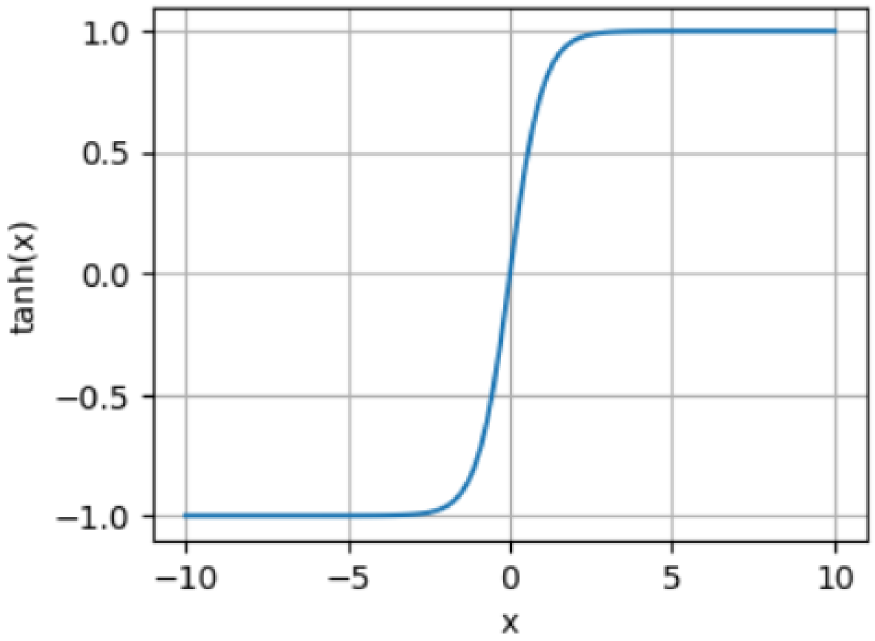
$$w_{new} = w_{current} - \alpha \boxed{\frac{\partial Error}{\partial w}}$$

$$h = x_1 w_n + x_2 w_n + \ldots + x_i w_n$$

$$\frac{\partial Error}{\partial w} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial AF} \cdot \boxed{\frac{\partial AF}{\partial h}} \cdot \boxed{\frac{\partial h}{\partial w}}$$

Tanh

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



This AF squashes the output between [-1, 1]

Derivative of Tanh

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

Quotient Rule

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{\left(g(x)\right)^2}$$

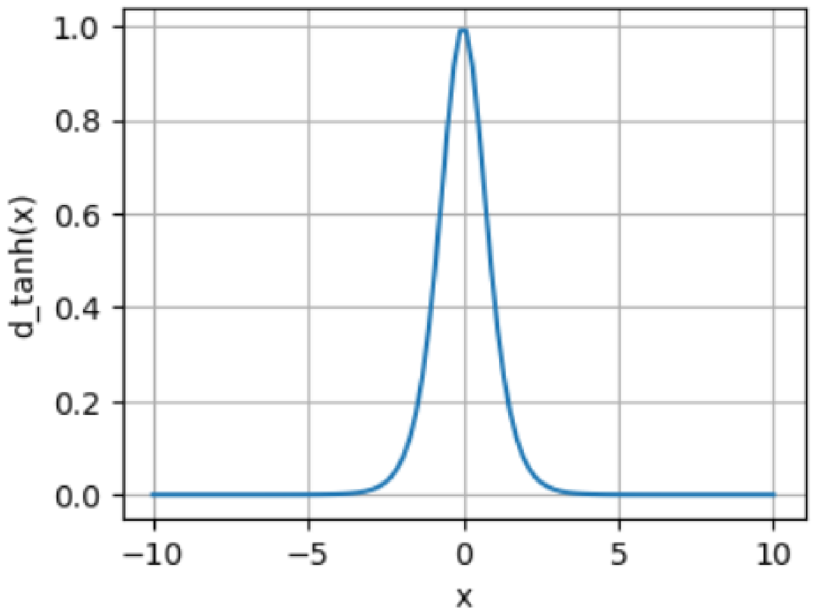
$$\frac{dTanh(x)}{dx} = \frac{d\left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)}{dx}$$

$$= \frac{(e^x + e^{-x})(e^x - (-(e^x))) - (e^x - e^{-x})(e^x + (-(e^x)))}{(e^x + e^{-x})^2}$$

$$= \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2}$$

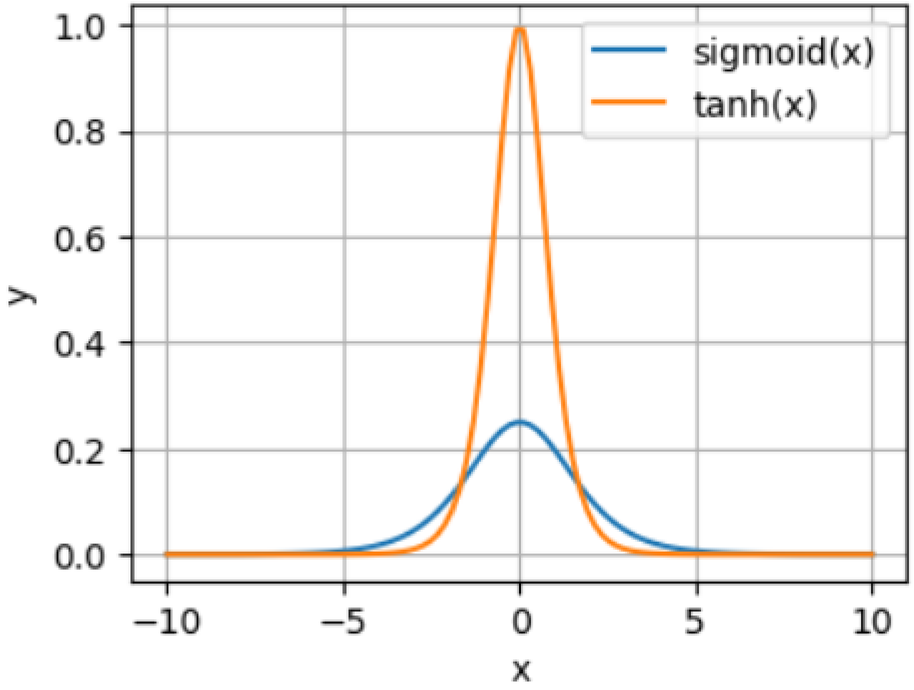
$$= \frac{(e^x + e^{-x})^2 - (e^x - e^{-x})^2}{(e^x + e^{-x})^2}$$

$$= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2 = 1 - Tanh(x)^2$$



$max = 1.0$

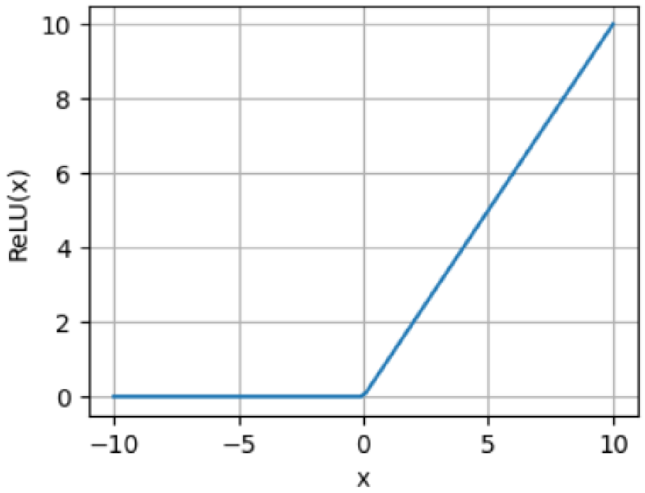
Sigmoid vs Tanh



vanishing gradient problem probably fixed in Tanh but still slow if tanh(x) is closer to 0

ReLU (Rectified Linear Unit)

$$ReLU(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



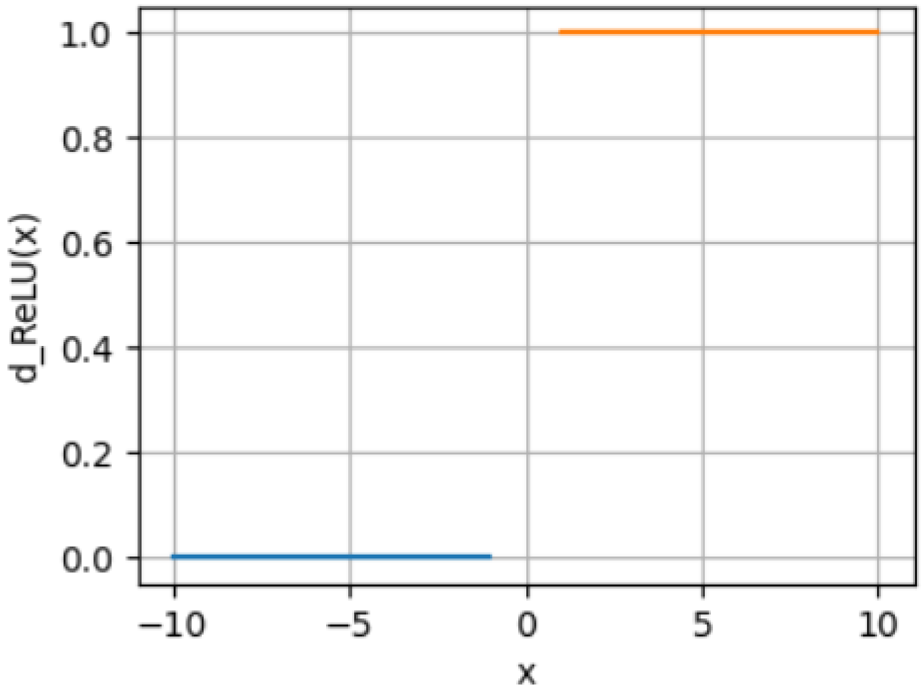
This AF squashes the output between [0, ∞)

Derivative of ReLU

$$ReLU(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$\frac{dReLU(x > 0)}{dx} = 1$$

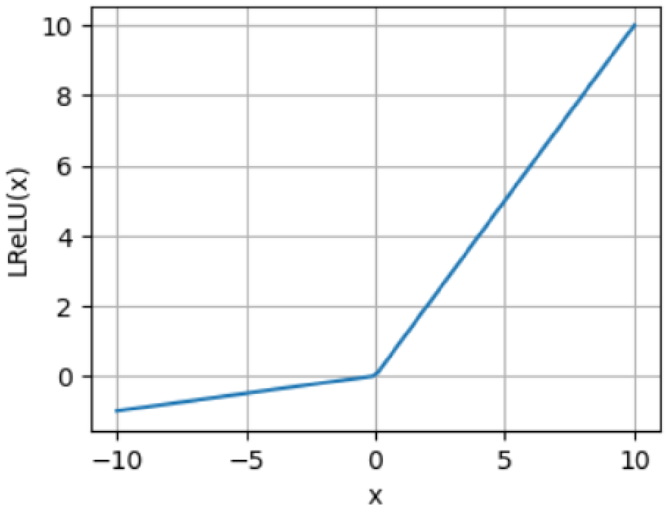
$$\frac{dReLU(x \leq 0)}{dx} = 0$$



$max = 1$

Leaky ReLU

$$LReLU(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$$



This AF squashes the output between (- ∞, +∞)

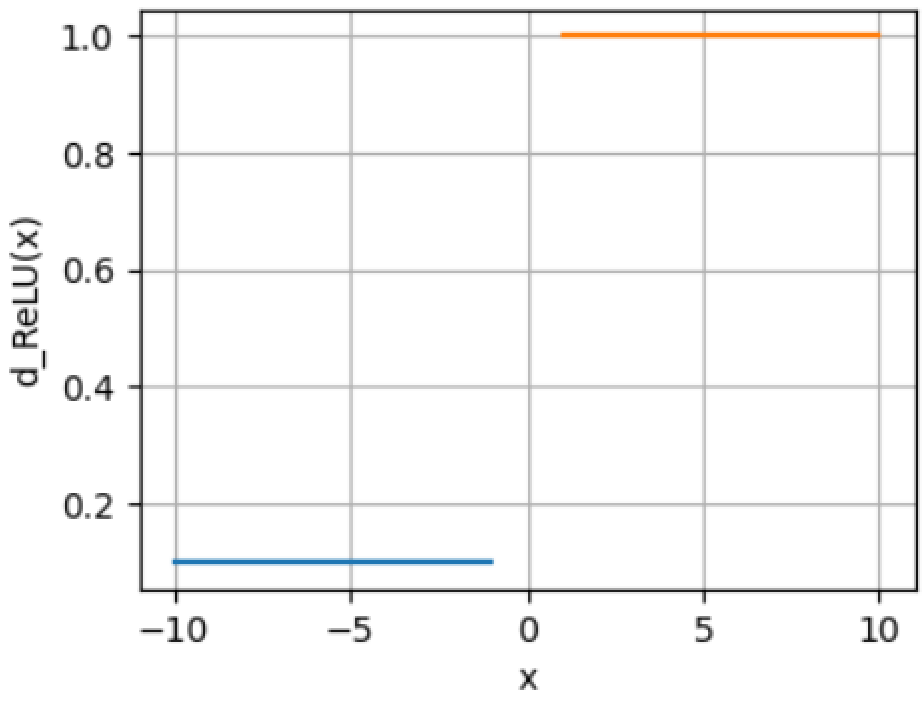
Derivative of LReLU

$$LReLU(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$$

$$\frac{dLReLU(x > 0)}{dx} = 1$$

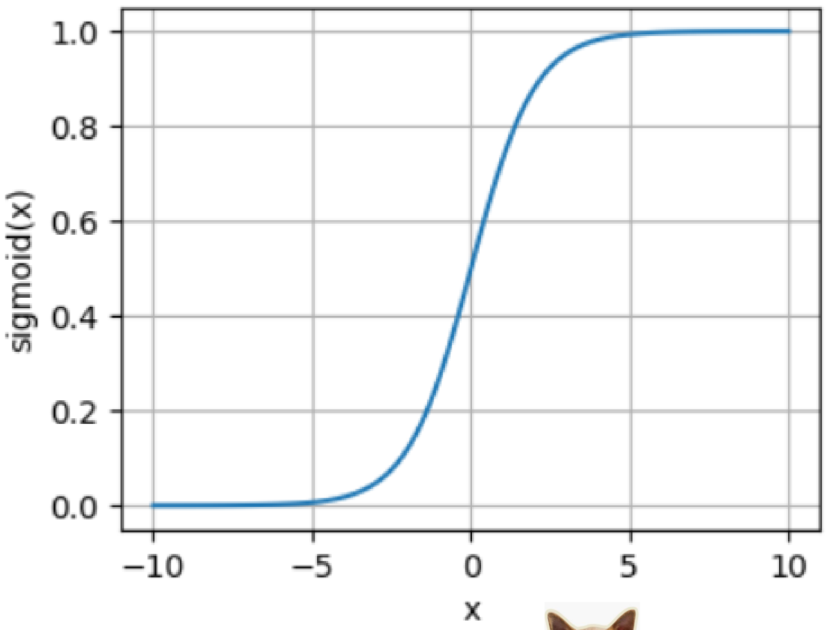
$$\frac{dLReLU(x \leq 0)}{dx} = \alpha$$

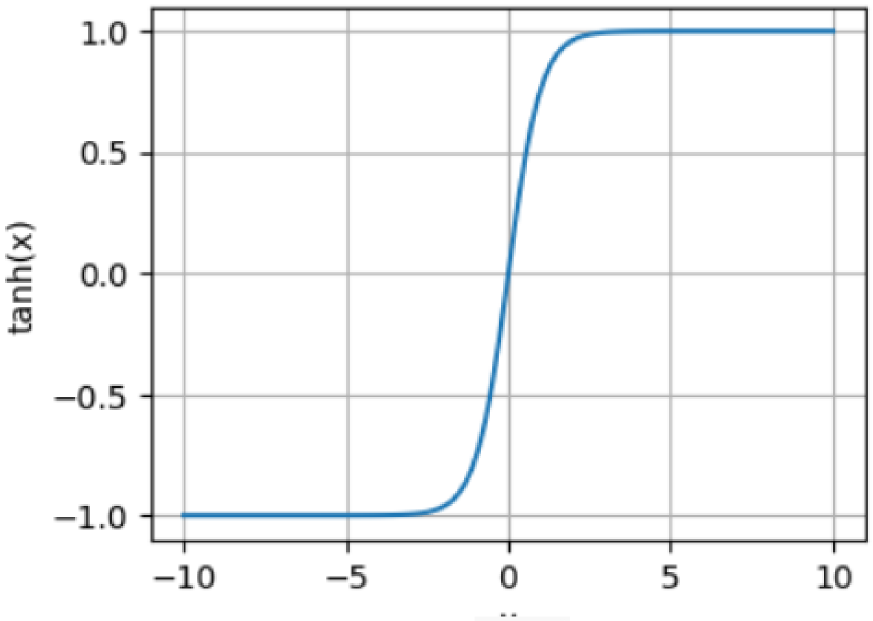
$Assume$: $\alpha = 0.1$



$max = 1$

Sigmoid and Tanh





$Sigmoid(x) > 0.5$

$Sigmoid(x) \leq 0.5$

Only fit with binary output

$Tanh(x) > 0$

$Tanh(x) \leq 0$

Chihuahua

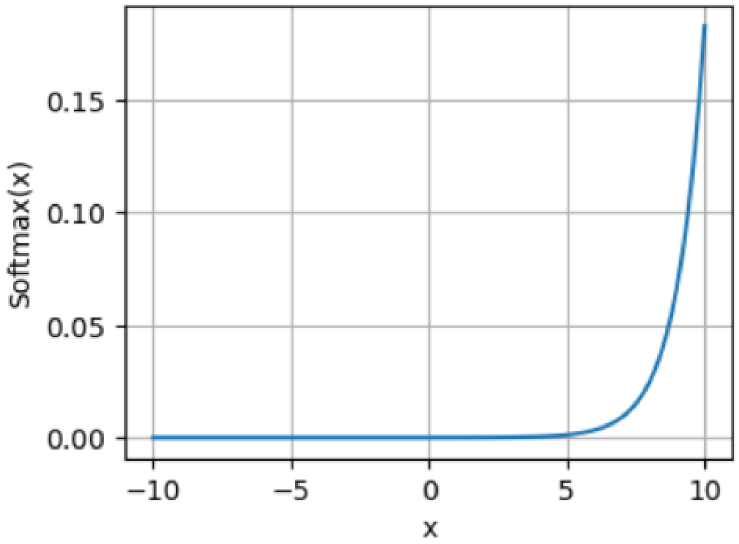
Golden Retriever

Chiba Inu

How to predict more than just binary classes ???

Softmax

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$



This AF squashes the output between (0, 1]

Derivative of Softmax

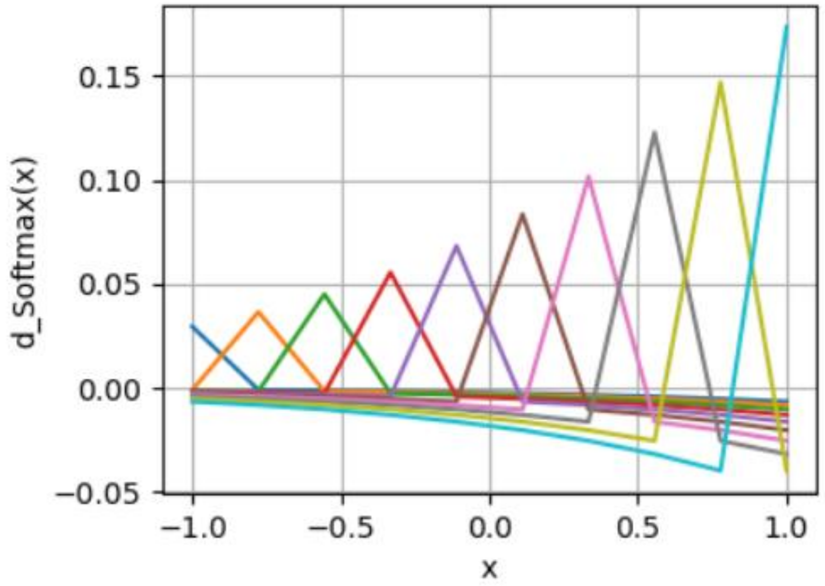
$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

$$\frac{\partial Softmax(x_i)}{\partial x_i} = \frac{\partial \left(\frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \right)}{\partial x_i}$$

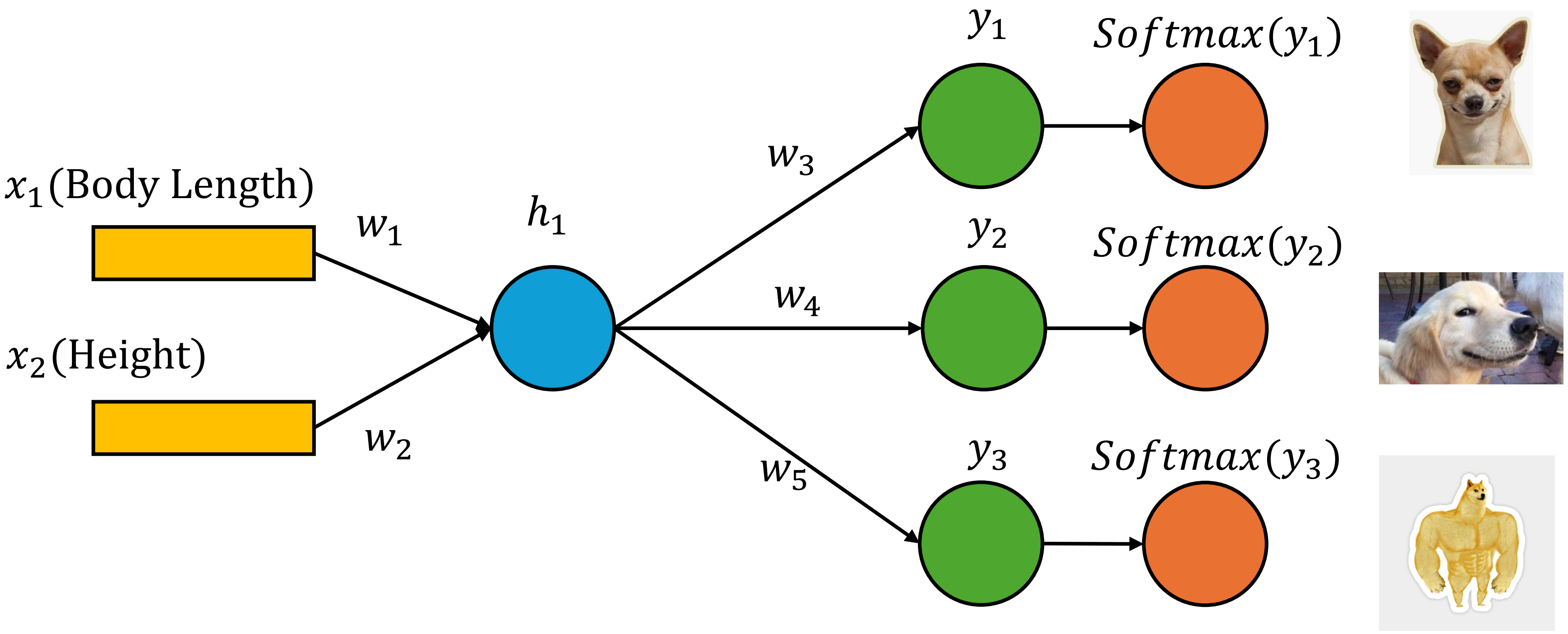
Quotient Rule

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{\left(g(x)\right)^2}$$

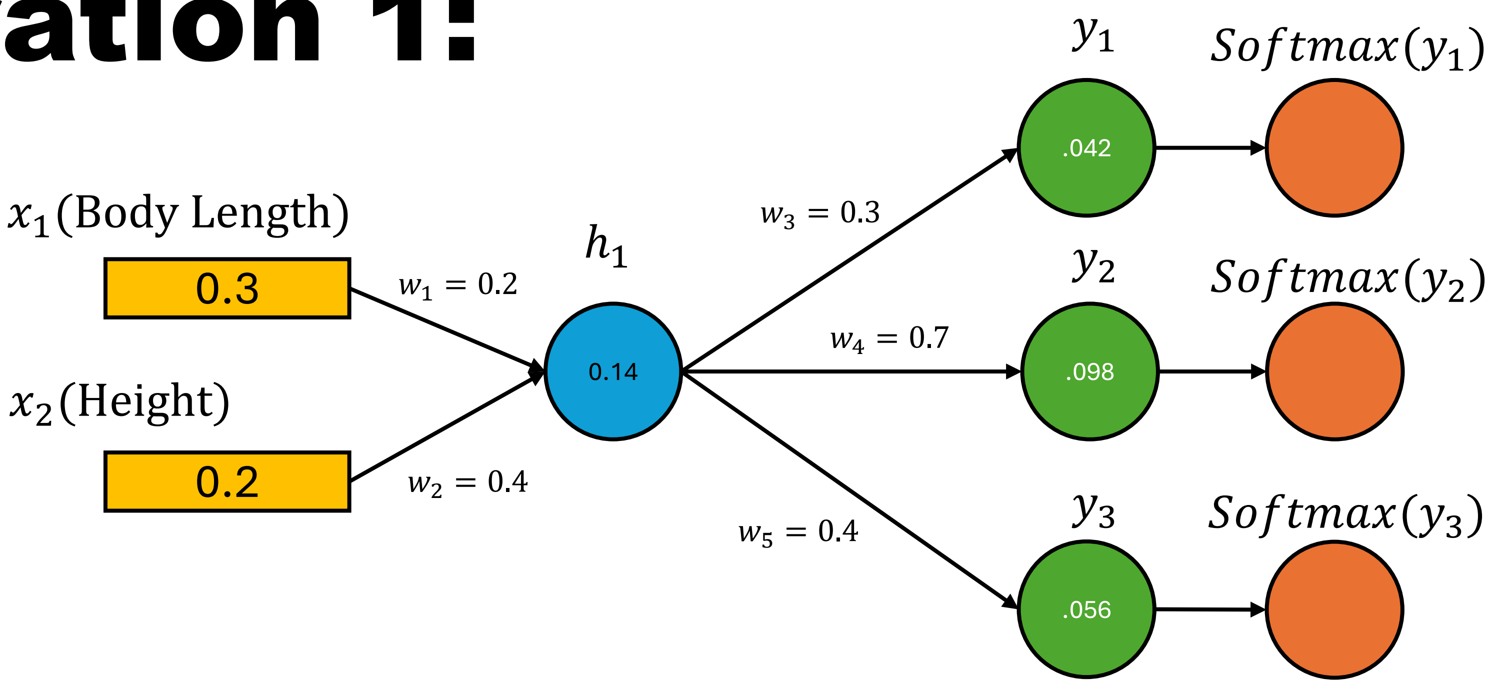
$$\frac{\partial Softmax(x_i)}{\partial x_i} = \begin{cases} Softmax(x_i) \cdot \left(1 - Softmax(x_i)\right), & i = j \\ -Softmax(x_i) \cdot Softmax(x_j), & i \neq j \end{cases}$$



X1 (Body Length)	X2 (Height)	Y (Breed)
0.3	0.2	[1, 0, 0]
0.6	0.5	[0, 1, 0]
0.4	0.3	[0, 0, 1]



Iteration 1:



$y_1 = 1$

$y_2 = 0$

$y_3 = 0$

$$h_1 = (x_1 w_1 + x_2 w_2) = \big((0.3)(0.2) + (0.2)(0.4)\big) = 0.14$$

$$y_1 = h_1 w_3 = (0.14)(0.3) = 0.042$$
$$y_2 = h_1 w_4 = (0.14)(0.7) = 0.098$$
$$y_3 = h_1 w_5 = (0.14)(0.4) = 0.056$$

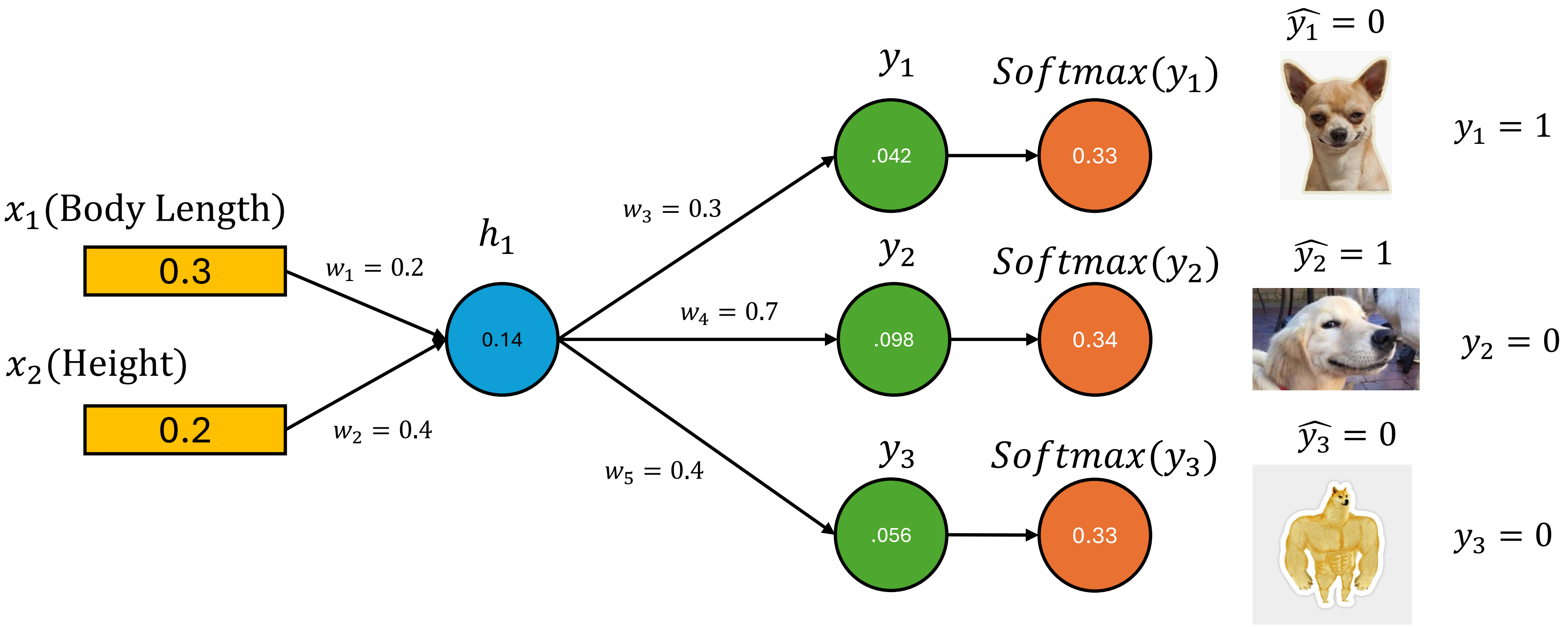
$$Softmax(y_i) = \frac{e^{y_i}}{e^{y_1} + e^{y_2} + e^{y_3}}$$

$$e^{y_1} + e^{y_2} + e^{y_3} = e^{0.042} + e^{0.098} + e^{0.056} = 3.20$$

$$Softmax(y_1) = \frac{e^{y_1}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.042}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.04}{3.20} = 0.33$$

$$Softmax(y_2) = \frac{e^{y_2}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.098}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.03}{3.20} = 0.34$$

$$Softmax(y_3) = \frac{e^{y_3}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.056}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.06}{3.20} = 0.33$$

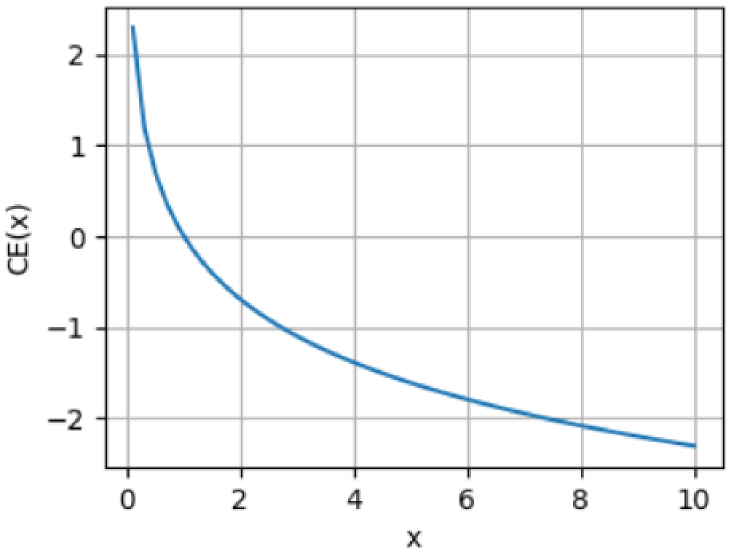


$$Softmax(y_1) = 0.33 \qquad ArgMax(S_1) = \widehat{y_1} = 0$$
$$Softmax(y_2) = 0.34 \longrightarrow ArgMax(S_2) = \widehat{y_2} = 1$$
$$Softmax(y_3) = 0.33 \qquad ArgMax(S_3) = \widehat{y_3} = 0$$

How to do error validation ???

Cross Entropy Loss Function

$$CE = -\sum_{i=1}^{n} Observed \cdot \log(P_i)$$



$$\begin{aligned} Softmax(y_1) &= 0.33 \\ Softmax(y_2) &= 0.34 \\ Softmax(y_3) &= 0.33 \end{aligned} \longrightarrow CE = -\sum_{i=1}^{n} Observed \cdot \log(P_i) = -\sum_{i=1}^{n} Observed \cdot \log(Softmax_i)$$

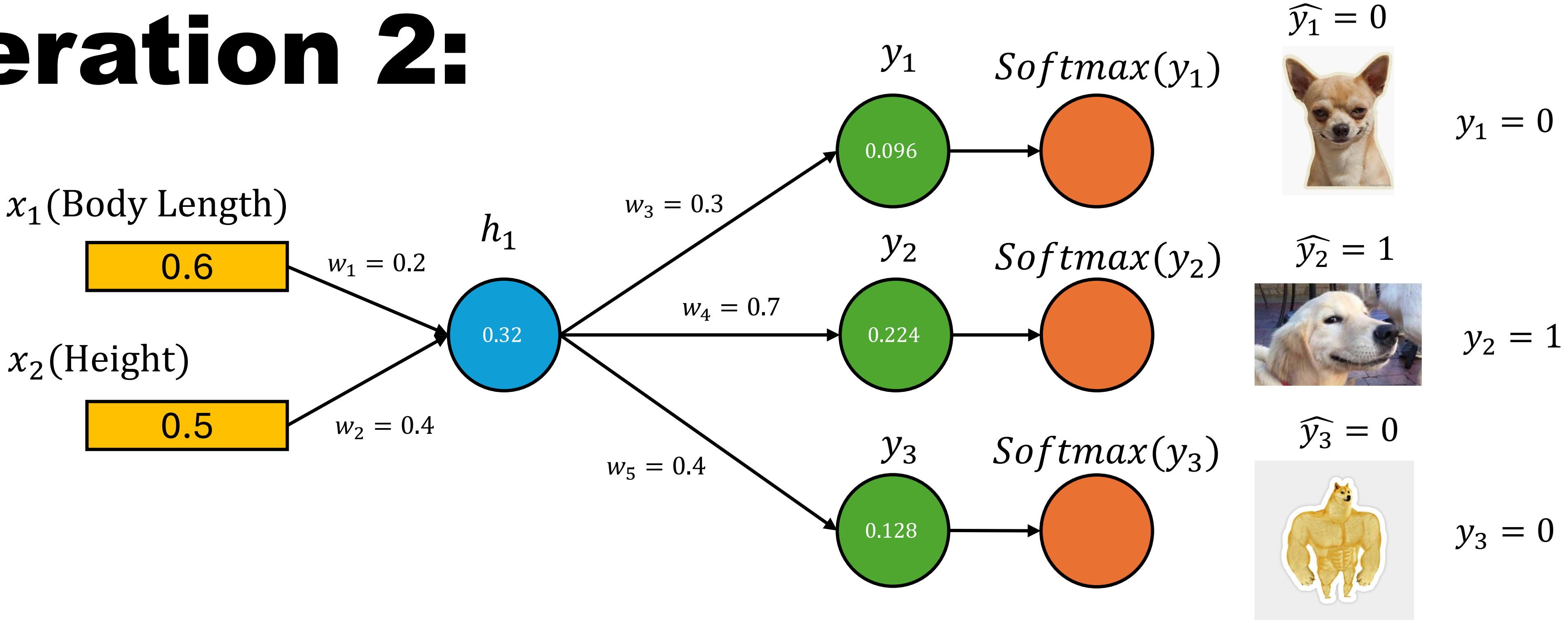
$$CE_1 = -1 \cdot \log(0.33) = 0.48$$
$$CE_2 = -0 \cdot \log(0.34) = 0$$
$$CE_3 = -0 \cdot \log(0.33) = 0$$

$$-\sum_{i=1}^{n} Observed \cdot \log(Softmax_i) = 0.48$$

X1 (Body Length)	X2 (Height)	Y (Breed)
0.3	0.2	[1, 0, 0]
0.6	0.5	[0, 1, 0]
0.4	0.3	[0, 0, 1]

Round	CE
1	0.48
2	???
3	???

Iteration 2:



$$h_1 = (x_1 w_1 + x_2 w_2) = \big((0.6)(0.2) + (0.5)(0.4)\big) = 0.32$$

$$y_1 = h_1 w_3 = (0.32)(0.3) = 0.096$$
$$y_2 = h_1 w_4 = (0.32)(0.7) = 0.224$$
$$y_3 = h_1 w_5 = (0.32)(0.4) = 0.128$$

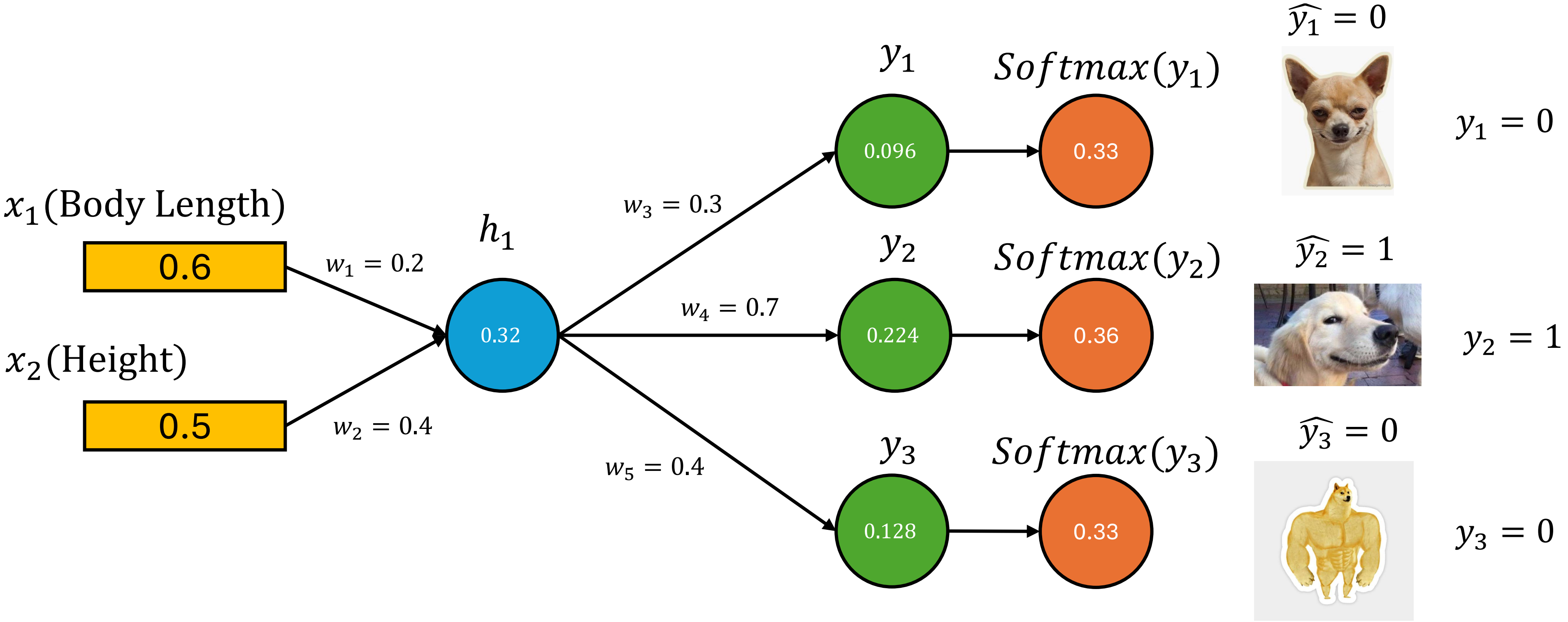
$$Softmax(y_i) = \frac{e^{y_i}}{e^{y_1} + e^{y_2} + e^{y_3}}$$

$$e^{y_1} + e^{y_2} + e^{y_3} = e^{0.096} + e^{0.224} + e^{0.128} = 3.49$$

$$Softmax(y_1) = \frac{e^{y_1}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.096}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.10}{3.49} = 0.32$$

$$Softmax(y_2) = \frac{e^{y_2}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.224}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.25}{3.49} = 0.36$$

$$Softmax(y_3) = \frac{e^{y_3}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.128}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.14}{3.49} = 0.33$$



$$
\begin{aligned}
Softmax(y_1) &= 0.32 \\
Softmax(y_2) &= 0.36 \\
Softmax(y_3) &= 0.33
\end{aligned}
\longrightarrow
\begin{aligned}
ArgMax(S_1) &= \widehat{y_1} = 0 \\
ArgMax(S_2) &= \widehat{y_2} = 1 \\
ArgMax(S_3) &= \widehat{y_3} = 0
\end{aligned}
$$

$$Softmax(y_1) = 0.32$$
$$Softmax(y_2) = 0.36$$
$$Softmax(y_3) = 0.33$$

$$\longrightarrow CE = -\sum_{i=1}^{n} Observed \cdot \log(P_i) = -\sum_{i=1}^{n} Observed \cdot \log(Softmax_i)$$

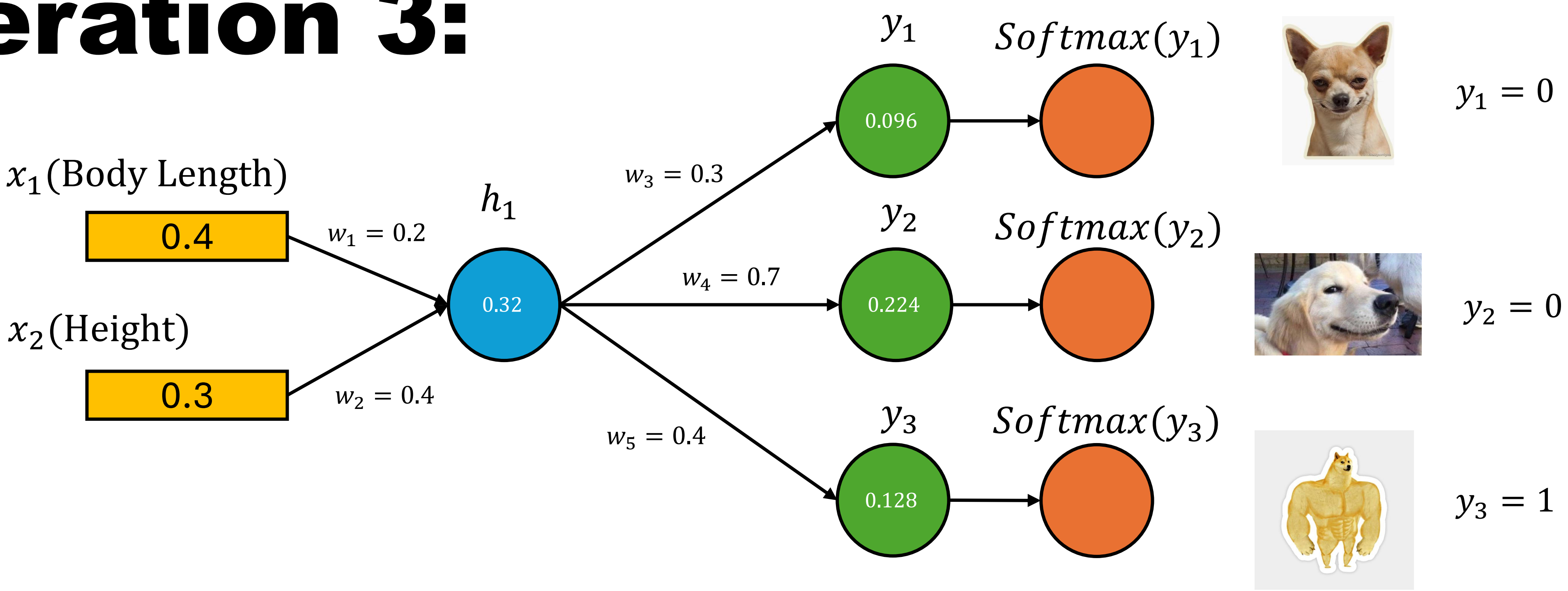
$$CE_1 = -0 \cdot \log(0.32) = 0$$
$$CE_2 = -1 \cdot \log(0.36) = 0.44$$
$$CE_3 = -0 \cdot \log(0.33) = 0$$

$$-\sum_{i=1}^{n} Observed \cdot \log(Softmax_i) = 0.44$$

X1 (Body Length)	X2 (Height)	Y (Breed)
0.3	0.2	[1, 0, 0]
0.6	0.5	[0, 1, 0]
0.4	0.3	[0, 0, 1]

Round	CE
1	0.48
2	0.44
3	???

Iteration 3:



$$h_1 = (x_1 w_1 + x_2 w_2) = \big((0.4)(0.2) + (0.3)(0.4)\big) = 0.2$$

$$y_1 = h_1 w_3 = (0.2)(0.3) = 0.06$$
$$y_2 = h_1 w_4 = (0.2)(0.7) = 0.14$$
$$y_3 = h_1 w_5 = (0.2)(0.4) = 0.08$$

$$Softmax(y_i) = \frac{e^{y_i}}{e^{y_1} + e^{y_2} + e^{y_3}}$$

$$e^{y_1} + e^{y_2} + e^{y_3} = e^{0.06} + e^{0.14} + e^{0.08} = 3.29$$

$$Softmax(y_1) = \frac{e^{y_1}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.06}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.06}{3.29} = 0.33$$

$$Softmax(y_2) = \frac{e^{y_2}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.14}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.15}{3.29} = 0.35$$

$$Softmax(y_3) = \frac{e^{y_3}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{e^{0.08}}{e^{y_1} + e^{y_2} + e^{y_3}} = \frac{1.08}{3.29} = 0.33$$

$$Softmax(y_1) = 0.33$$
$$Softmax(y_2) = 0.35$$
$$Softmax(y_3) = 0.33$$

$$\longrightarrow CE = -\sum_{i=1}^{n} Observed \cdot \log(P_i) = -\sum_{i=1}^{n} Observed \cdot \log(Softmax_i)$$

$$CE_1 = -0 \cdot \log(0.33) = 0$$
$$CE_2 = -0 \cdot \log(0.35) = 0$$
$$CE_3 = -1 \cdot \log(0.33) = 0.48$$

$$-\sum_{i=1}^{n} Observed \cdot \log(Softmax_i) = 0.48$$

X1 (Body Length)	X2 (Height)	Y1 (Breed)
0.3	0.2	[1, 0, 0]
0.6	0.5	[0, 1, 0]
0.4	0.3	[0, 0, 1]

Round	CE
1	0.48
2	0.44
3	0.48

$$\sum_{i=1}^{n} CE_i = CE_1 + CE_2 + CE_3$$

$$= 0.48 + 0.44 + 0.48$$

$$= 1.4$$

This is total error

Back propagation

$$w_{new} = w_{current} - \alpha \frac{\partial \sum_{i=1}^{n} CE_i}{\partial w}$$

Chain Rule Example

$$\frac{\partial \sum_{i=1}^{n} CE_i}{\partial w_3}$$

$$= \frac{\partial \sum_{i=1}^{n} CE_i}{\partial Softmax} \cdot \frac{\partial Softmax}{\partial y} \cdot \frac{\partial y}{\partial w_3}$$

$$\frac{\partial \sum_{i=1}^{n} CE_i}{\partial w_1}$$

$$= \frac{\partial \sum_{i=1}^{n} CE_i}{\partial Softmax} \cdot \frac{\partial Softmax}{\partial y} \cdot \frac{\partial y}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$