Explanation of Project 2D

convert(X) ~ one-hot Encoding

$x_1 = [11, \text{"warm-blooded"}, 2]$

$x_2 = [13, \text{"cold-blooded"}, 3]$

"Body Temperature" only has two values.

$$\begin{cases} \text{warm-blooded}: [1, 0] \\ \text{cold-blooded}: [0, 1] \end{cases}$$

Thus:

$x_1 = [11, 1, 0, 2]$

$x_2 = [13, 0, 1, 3]$

Example: size = {small, medium, Large}

three values!

$$\begin{cases} \text{small}: [1, 0, 0] \\ \text{medium}: [0, 1, 0] \\ \text{large}: [0, 0, 1] \end{cases}$$

Explanation of Adaboost.M1 algorithm

Say $N = 100$

Step-1 : $w_1 = w_2 = \cdots = w_{100} = \frac{1}{100}$

Step-2 : at m-th iteration. $(m=1)$

Fit $G_m(x)$ to 100 training data using w_i

Say we have 80 easier sample, which are classified correctly by $G_m(x)$.

$$Err_m = \frac{100-80}{100} = 0.2$$

$$\alpha_m = \log\left((1-0.2)/0.2\right) = \log 4$$

For 80 easier samples: $y_i = G_m(x_i)$

$$w_i \leftarrow \frac{1}{100} \cdot \exp[\log 4 \cdot 0] = \underline{\underline{\frac{1}{100} \cdot 1}} \text{ (still same)}$$

For 20 harder samples : $y_i \neq G_m(x_i)$

$$w_i \leftarrow \frac{1}{100} \cdot \exp[\log 4 \cdot 1] = \underline{\underline{\frac{1}{100} \cdot \exp(\log 4)}}$$

(increasing)

Thus, we have:

$$w_i(\text{easier}) = \frac{1}{100} < w_i(\text{harder}) = \frac{1}{100}\exp(\log 4)$$

As it can be seen, after 1st iteration, easier samples has relatively lower weights while harder samples has relatively larger weights.

Derivation of the derivative of sigmoid activation function

sigmoid function: $\sigma(z) = \frac{1}{1+e^{-z}} = (1+e^{-z})^{-1}$

its derivative is:

$$\frac{d\sigma(z)}{dz} = -(1+e^{-z})^{-2} \cdot (e^{-z}) \cdot (-1)$$

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

$$= \frac{1+e^{-z}-1}{1+e^{-z}} \cdot \frac{1}{1+e^{-z}}$$

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

thus: $\frac{d\sigma(z)}{dz} = (1-\sigma(z)) \cdot \sigma(z)$