Exercise 1: Medusa-Boost

We start by introducing the indicator-function $1_{\{t_i \neq \phi(x_i)\}}$ of the set where the weak classifier fails and define

$$a = \sum_{i=1}^{N} (1_{\{t_i \neq \phi(x_i)\}} c_i)$$

$$b = \sum_{i=1}^{N} (1 - 1_{\{t_i \neq \phi(x_i)\}}) c_i$$

$$c = \sum_{i=1}^{N} c_i = a + b$$

$$E(\phi, w) = \sum_{i=1}^{N} \frac{1}{\exp(t_i y_m - 1(x_i))} \log(1 + \exp(-t_i w \phi(x_i))))$$

$$= \sum_{i=1}^{N} 1_{\{t_i \neq \phi(x_i)\}} c_i \log(1 + e^w) + \sum_{i=1}^{N} (1 - 1_{\{t_i \neq \phi(x_i)\}}) c_i \log(1 + e^{-w})$$

$$= \log\left(\frac{1 + e^w}{1 + e^{-w}}\right) \sum_{i=1}^{N} 1_{\{t_i \neq \phi(x_i)\}} c_i + \log(1 + e^{-w}) \sum_{i=1}^{N} c_i$$

Thus we can choose $\phi$ independently from its weight in order to minimize $a$. Furthermore it holds

$$\log\left(\frac{1 + e^w}{1 + e^{-w}}\right) = \log\left(\frac{e^w(e^{-w} + 1)}{1 + e^{-w}}\right) = w$$

This simplifies the calculation of the partial derivative of $E$:

$$\frac{\partial E(\phi, w)}{\partial w} = \sum_{i=1}^{N} 1_{\{t_i \neq \phi(x_i)\}} c_i - \frac{1}{1 + e^{-w}} \sum_{i=1}^{N} c_i = 0$$

$$\Leftrightarrow e^{-w} = \frac{\sum_{i=1}^{N} c_i}{\sum_{i=1}^{N} 1_{\{t_i \neq \phi(x_i)\}} c_i} - 1 = \frac{\sum_{i=1}^{N} (1 - 1_{\{t_i \neq \phi(x_i)\}}) c_i}{\sum_{i=1}^{N} 1_{\{t_i \neq \phi(x_i)\}} c_i} = \frac{b}{a}$$

$$\Leftrightarrow w = - \log \frac{b}{a} = \log \frac{a}{b}$$
Exercise 2: Committee

The exercise and the solution are taken from the book *Pattern Recognition and Machine Learning (Bishop)*. Note that the following enumeration is used:

\[ \alpha_m \geq 0, \sum_{m=1}^{M} \alpha_m = 1 \]  
(14.56)

\[ y_{\min}(x) \leq y_{\text{com}}(x) \leq y_{\max}(x) \]  
(14.57)

To prove that (14.57) is a sufficient condition for (14.56) we have to show that (14.56) follows from (14.57). To do this, consider a fixed set of \( y_m(x) \) and imagine varying the \( \alpha_m \) over all possible values allowed by (14.57) and consider the values taken by \( y_{\text{com}}(x) \) as a result. The maximum value of \( y_{\text{com}}(x) \) occurs when \( \alpha_k = 1 \) where \( y_k(x) \geq y_m(x) \) for \( m \neq k \), and hence all \( \alpha_m = 0 \) for \( m \neq k \). An analogous result holds for the minimum value. For other settings of \( \alpha \), \( y_{\min}(x) < y_{\text{com}}(x) < y_{\max}(x) \), since \( y_{\text{com}}(x) \) is a convex combination of points, \( y_m(x) \), such that

\[ \forall m: y_{\min}(x) \leq y_{\text{com}}(x) \leq y_{\max}(x) \]

Thus, (14.57) is a sufficient condition for (14.56). Showing that (14.57) is a necessary condition for (14.56) is equivalent to showing that (14.56) is a sufficient condition for (14.57). The implication here is that if (14.56) holds for any choice of values of the committee members \( \{y_m(x)\} \) then (14.57) will be satisfied. Suppose, without loss of generality, that \( \alpha_k \) is the smallest of the \( \alpha \) values, i.e. \( \alpha_k \leq \alpha_m \) for \( k \neq m \). Then consider \( y_k(x) = 1 \), together with \( y_m(x) = 0 \) for all \( m \neq k \). Then \( y_{\min}(x) = 0 \) while \( y_{\text{com}}(x) = \alpha_k \) and hence from (14.56) we obtain \( \alpha_k \geq 0 \). Since \( \alpha_k \) is the smallest of the \( \alpha \) values it follows that all of the coefficients must satisfy \( \alpha_m \geq 1 \). Similarly, consider the case in which \( y_m(x) = 1 \) for all \( m \). Then \( y_{\min}(x) = y_{\max}(x) = 1 \), while \( y_{\text{com}}(x) = \sum_m \alpha_m \). From (14.56) it then follows that \( \sum_m \alpha_m = 1 \), as required.

Exercise 3: Programming

```
    cd build
    cmake .
    make
    ./adaboost 20
```

for-loop, output in file, plotting the file with gnuplot