# TP2

Thibault Duhamel 18026048, Heng Shi 18171434

February 15, 2019

## Question 1

The loss function is defined as follows:

$$E(\overrightarrow{w}) = \sum_{n}^{N} (t_n - \overrightarrow{w}^T.\overrightarrow{\phi_n})^2 + \lambda \overrightarrow{w}^T.\overrightarrow{w}$$

As we want to minimize the above expression, it is convenient to calculate its gradient:

$$\frac{dE(\overrightarrow{w})}{d\overrightarrow{w}} = \sum_{n}^{N} \big[ -2(t_n - \overrightarrow{w}^T.\overrightarrow{\phi_n})\overrightarrow{\phi_n}^T \big] + 2\lambda \overrightarrow{w}$$

Let us now set this gradient to zero in order to deduce $\overrightarrow{w}$:

$$\frac{dE(\overrightarrow{w})}{d\overrightarrow{w}} = 0$$

$$\sum_{n}^{N} \big[ -2(t_n - \overrightarrow{w}^T.\overrightarrow{\phi_n})\overrightarrow{\phi_n}^T \big] + 2\lambda \overrightarrow{w} = 0$$

$$\sum_{n}^{N} \big[ (\overrightarrow{w}^T.\overrightarrow{\phi_n})\overrightarrow{\phi_n}^T \big] - \sum_{n}^{N} \big[ t_n \overrightarrow{\phi_n}^T \big] + \lambda \overrightarrow{w} = 0$$

We can write the exact same equality using matrices, with $\Phi$ being the same matrix as defined in the book from Bishop:

$$\Phi^T \Phi \overrightarrow{w} - \Phi^T \overrightarrow{t} + \lambda \overrightarrow{w} = 0$$

$$\Phi^T \Phi \overrightarrow{w} + \lambda \overrightarrow{w} = \Phi^T \overrightarrow{t}$$

$$(\Phi^T \Phi + \lambda I) \overrightarrow{w} = \Phi^T \overrightarrow{t}$$

And finally, assuming the matrix $(\Phi^T \Phi + \lambda I)$ is invertible:

$$\boxed{\overrightarrow{w} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T \overrightarrow{t}}$$

## Question 2

In this question, we do not write the transpose symbol to simplify expressions.

The cross-entropy loss function is defined as follows:

$$E(\overrightarrow{w}) = -\sum_{n}^{N}\left[t_n ln(\sigma(\overrightarrow{w}.\overrightarrow{\phi_n})) + (1-t_n)ln(1-\sigma(\overrightarrow{w}.\overrightarrow{\phi_n}))\right]$$

Let us first compute the gradient of the sigmoid function:

$$\frac{d\sigma}{d\overrightarrow{w}} = \frac{\overrightarrow{\phi_n}e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{(1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}})^2}$$

Then, we can use this intermediate result to calculate the gradient of the 2 logarithms in the sum:

$$\frac{dln(\sigma)}{d\overrightarrow{w}} = \frac{\frac{d\sigma}{d\overrightarrow{w}}}{\sigma}$$

$$= \frac{\overrightarrow{\phi_n}e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{(1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}})^2} \times (1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}})$$

$$= \frac{\overrightarrow{\phi_n}e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}$$

and:

$$\frac{dln(1-\sigma)}{d\overrightarrow{w}} = \frac{-\frac{d\sigma}{d\overrightarrow{w}}}{1-\sigma}$$

$$= \frac{-\overrightarrow{\phi_n}e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{(1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}})^2} \times \frac{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}$$

$$= \frac{-\overrightarrow{\phi_n}}{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}$$

From those 2 expressions, we are now able to express the gradient of the whole sum:

$$\frac{dE(\overrightarrow{w})}{d\overrightarrow{w}} = -\sum_{n}^{N}\left[t_n \frac{\overrightarrow{\phi_n}e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}} + (1-t_n)\frac{-\overrightarrow{\phi_n}}{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}\right]$$

$$= -\sum_{n}^{N}\left[t_n \frac{e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}} + (1-t_n)\frac{-1}{1+e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}}\right]\overrightarrow{\phi_n}$$

$$= -\sum_{n}^{N} \left[ t_n y_n e^{-\overrightarrow{w}.\overrightarrow{\phi_n}} + (t_n - 1) y_n \right] \overrightarrow{\phi_n}$$

$$= -\sum_{n}^{N} \left[ t_n y_n (1 + e^{-\overrightarrow{w}.\overrightarrow{\phi_n}}) - y_n \right] \overrightarrow{\phi_n}$$

$$= -\sum_{n}^{N} \left[ t_n y_n \frac{1}{y_n} - y_n \right] \overrightarrow{\phi_n}$$

$$= -\sum_{n}^{N} \left[ t_n - y_n \right] \overrightarrow{\phi_n}$$

$$\boxed{\frac{dE(\overrightarrow{w})}{d\overrightarrow{w}} = \sum_{n}^{N} \left[ y_n - t_n \right] \overrightarrow{\phi_n}}$$

## Question 3

Let us first write down the definition of the entropy in this context:

$$H(X) = -p_1 log_2(p_1) - p_2 log_2(p_2) - p_3 log_2(p_3)$$

As we want to maximize this function (with $p_1 + p_2 + p_3 = 1$) we can express the problem with a Lagrange multiplier:

$$L = -p_1 log_2(p_1) - p_2 log_2(p_2) - p_3 log_2(p_3) + \lambda(p_1 + p_2 + p_3 - 1)$$

However, we do have another condition on those probabilities (that is $p_1 = 2p_2$), so we can modify the expression consequently:

$$L = -2p_2 log_2(2p_2) - p_2 log_2(p_2) - p_3 log_2(p_3) + \lambda(3p_2 + p_3 - 1)$$

$$L = -2p_2 - 2p_2 log_2(p_2) - p_2 log_2(p_2) - p_3 log_2(p_3) + \lambda(3p_2 + p_3 - 1)$$

$$L = -2p_2 - 3p_2 log_2(p_2) - p_3 log_2(p_3) + \lambda(3p_2 + p_3 - 1)$$

Let us now compute the gradients of such a function:

$$\begin{cases} \frac{dL}{dp_2} = -2 - 3log_2(p_2) - \frac{3}{ln(2)} + 3\lambda \\ \frac{dL}{dp_3} = -log_2(p_3) - \frac{1}{ln(2)} + \lambda \end{cases}$$

We can then set those gradients to zero:

$$\begin{cases} -2 - 3log_2(p_2) - \frac{3}{ln(2)} + 3\lambda = 0 \\ -log_2(p_3) - \frac{1}{ln(2)} + \lambda = 0 \end{cases}$$

Let us first divide the first equation by 3:

$$\begin{cases} -\frac{2}{3} - log_2(p_2) - \frac{1}{ln(2)} + \lambda = 0 \\ -log_2(p_3) - \frac{1}{ln(2)} + \lambda = 0 \end{cases}$$

As both equations contain the same terms, let us substract $L_2$ from $L_1$:

$$log_2(\frac{p_3}{p_2}) = \frac{2}{3}$$

It is now possible to extract a relation between $p_3$ and $p_2$:

$$p_3 = 2^{2/3}p_2$$

As all probabilities now depend on $p_2$, we can insert those terms inside the condition $p_1 + p_2 + p_3 = 1$:

$$2p_2 + p_2 + 2^{2/3}p_2 = 1$$

Which gives:

$$\boxed{p_2 = \frac{1}{2^{2/3} + 3}}$$

Now that we know $p_2$, it is eventually trivial to compute $p_1$ and $p_3$:

$$\boxed{p_1 = \frac{2}{2^{2/3} + 3} \text{ and } p_3 = \frac{2^{2/3}}{2^{2/3} + 3}}$$