2.1 Given :

- $\{x_n\}_{n=0}^{N-1}$ are N independent one hot encoded training samples where $x_n \in \mathbb{R}^M$
- m_n is the class of the n^{th} sample i.e. $x_{n,m_n} = 1$
- N_i is the number of training samples with class i

$\Rightarrow$ Therefore we can write N_i as

$$N_i = \sum_{n=0}^{N-1} x_{n,i}$$

- As, for training samples with class i

$x_n = (0, 0, \ldots \underset{i^{th}\text{ index}}{1}, 0, \ldots \underset{(M-1)^{th}\text{ index}}{0})$

and for training samples with class other than i

$x_n = (\ldots \underset{i^{th}\text{ index}}{0}, \ldots)$