Neg log-likelihood

For sample i,

$$P(x^{(i)} \in C_1) = y^{(i)} = f(x^{(i)}; \theta)$$

Thus, our model tells us that the likelihood of observing $x^{(i)}$ is $y^{(i)}$. If $x^{(i)} \in C_0$, then the likelihood of observing $x^{(i)}$ is $1-y^{(i)}$.

Putting those together into one formula that works for both cases...

$$P(x^{(i)} \in C_{t^{(i)}}) = \left(y^{(i)}\right)^{t^{(i)}} \left(1-y^{(i)}\right)^{1-t^{(i)}} \quad \text{for } t^{(i)} \in \{0,1\}.$$

The likelihood of observing ALL the samples

$$x^{(i)} \in C_{t^{(i)}} \quad \text{for } i=1,\ldots,N$$

is the product of all their individual probabilities.

$$P\left(x^{(i)} \in C_{t^{(i)}} \text{ for } i=1,\ldots,N\right)$$

$$= \prod_{i=1}^{N} \left(y^{(i)}\right)^{t^{(i)}} \left(1-y^{(i)}\right)^{1-t^{(i)}}$$

Taking the negative log-likelihood...

$$-\ln P\left(x^{(i)} \in C_{t^{(i)}} \text{ for } i=1,\ldots,N\right)$$

$$= -\left[\sum_{i=1}^{N} \ln\left(y^{(i)}\right)^{t^{(i)}} + \ln\left(1-y^{(i)}\right)^{1-t^{(i)}}\right]$$

$$= -\sum_{i=1}^{N} t^{(i)} \ln y^{(i)} + (1-t^{(i)}) \ln(1-y^{(i)})$$

as required. ▨