2. Training binary classifiers in situation where we do not have full access to the labels.

→ We have labels only for a subset of positive examples.

*

(a) Suppose that each $y^{(i)}$ and $x^{(i)}$ are conditionally independent given $t^{(i)}$

$$P(y^{(i)}=1 \mid t^{(i)}=1, x^{(i)}) = P(y^{(i)}=1 \mid t^{(i)}=1)$$

⇒ We assume a dataset $\{(x^{(i)}, t^{(i)}, y^{(i)})\}_{i=1}^{m}$

where $t^{(i)} \in \{0,1\}$ is the "true" label and where

$$y^{(i)} = \begin{cases} 1 & x^{(i)} \text{ is labeled} \\ 0 & \text{otherwise} \end{cases}$$

⇒ All labeled examples are positive, which is to say $P(t^{(i)}=1 \mid y^{(i)}=1) = 1$

⇒ We want to construct h such that $h(x^{(i)}) \approx P(t^{(i)}=1 \mid x^{(i)})$ as closely as possible, using only x and y.

$\Rightarrow P(y^{(i)}=1 \mid x^{(i)})$

$\Rightarrow P(y^{(i)}=1 \mid t^{(i)}=1, x^{(i)})\, P(t^{(i)}=1 \mid x^{(i)})$

$+ P(y^{(i)}=1 \mid t^{(i)}=0, x^{(i)})\, P(t^{(i)}=0 \mid x^{(i)})$

{Zero as for negative example there is no lable}

$\Rightarrow P(y^{(i)}=1 \mid t^{(i)}=1)\, P(t^{(i)}=1 \mid x^{(i)})$

{$y^{(i)}$ & $x^{(i)}$ are conditionally independent given $t^{(i)}=1$}

$\Rightarrow \boxed{\alpha\, P(t^{(i)}=1 \mid x^{(i)})}$

$$\frac{\sum_{i=1}^{m} 1\{y^{(i)}=1\}}{\sum_{i=1}^{m} 1\{t^{(i)}=1\}}$$

{Labeled example were selected uniformly at random from the set of positive examples}