Learning from multiple tasks

Multi-task learning

deeplearning.ai

Simplified autonomous driving example

$X^{(i)}$

$y^{(i)}$ $(4, 1)$

$$\begin{array}{ll}
\text{Pedestrians} & 0 \\
\text{Cars} & 1 \\
\text{Stop signs} & 1 \\
\text{Traffic lights} & 0 \\
\vdots & \vdots
\end{array}$$

$$Y = \left[y^{(1)} \quad y^{(2)} \quad y^{(3)}, \cdots, y^{(m)} \right]$$

$(4, m)$

Neural network architecture

pedestrian ← 0

car ← 1 $\in \mathbb{R}^4$

stop sign ← 1

traffic light ← 0
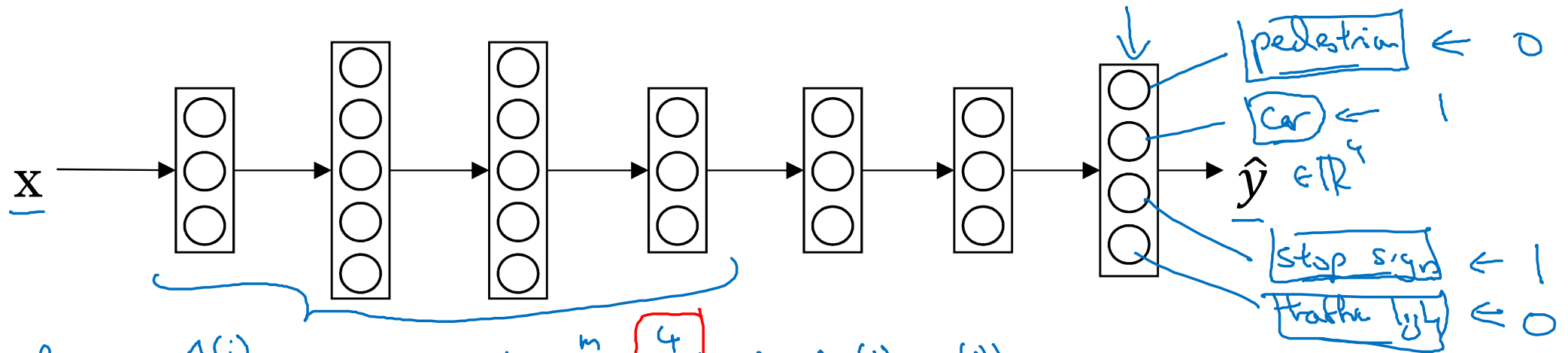
Loss: $\hat{y}^{(i)}$ $(4,1)$

$\rightarrow \dfrac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{4} \mathcal{L}\left(\hat{y}_j^{(i)}, y_j^{(i)}\right)$

Sum only over value of j with $0/1$ label.

$\rightarrow$ Usual logistic loss

$-y_j^{(i)} \log \hat{y}_j^{(i)} - (1-y_j^{(i)}) \log (1-\hat{y}_j^{(i)})$

Unlike softmax regression:
One image can have multiple labels

Multi-task learning ←

$$Y = \begin{bmatrix} 1 & 1 & 0 & ? & \cdots \\ 0 & 1 & 1 & 1 & \cdots \\ ? & ? & 1 & ? & \\ ? & ? & 0 & ? & \end{bmatrix}$$

Andrew Ng

When multi-task learning makes sense

- Training on a set of tasks that could benefit from having shared lower-level features.
- Usually: Amount of data you have for each task is quite similar.

$A \quad 1,000,000$

$\downarrow \qquad \downarrow$

$B \quad 1,000$

$A_1 \quad 1,000$
$A_2 \quad 1,000$
$\vdots \qquad \vdots \qquad 99,000$
$A_{100} \quad 1,000$
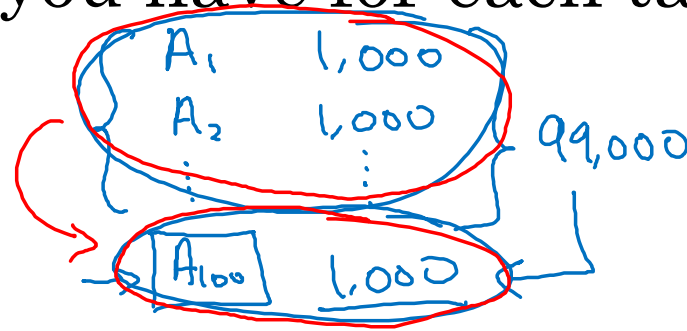
- Can train a big enough neural network to do well on all the tasks.