Learning from multiple tasks

Multi-task learning

deeplearning.ai

Simplified autonomous driving example

$X^{(i)}$

Pedestrians

Cars

Stop signs
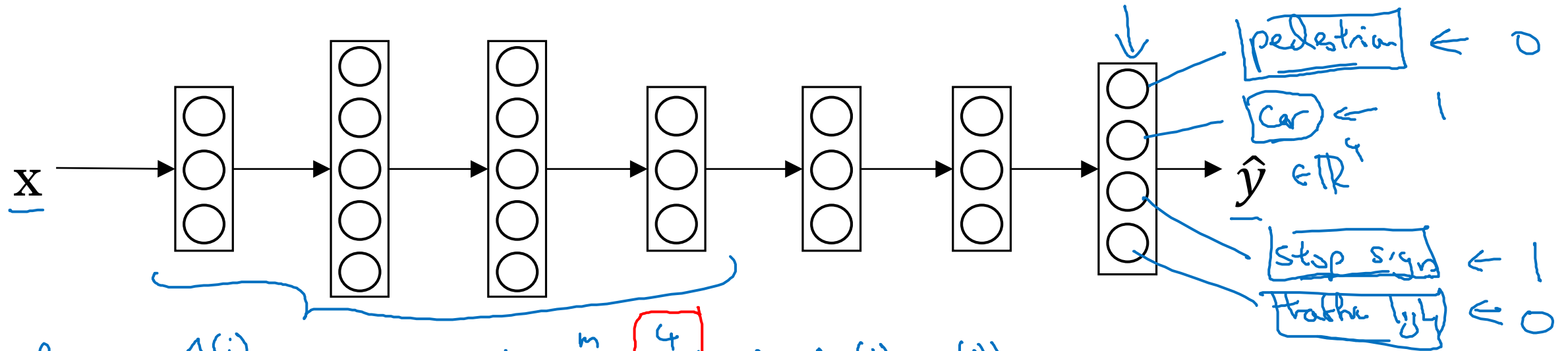
Traffic lights

$\vdots$

$$y^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \quad (4,1)$$

$$Y = \begin{bmatrix} | & | & | & & | \\ y^{(1)} & y^{(2)} & y^{(3)} & \cdots, & y^{(m)} \\ | & | & | & & | \end{bmatrix}$$

$(4, m)$

Andrew Ng

Neural network architecture

$$x \rightarrow \hat{y}$$

pedestrian ← 0
Car ← 1
$\hat{y} \in \mathbb{R}^4$
stop sign ← 1
traffic light ← 0

Loss: $\hat{y}^{(i)}$ $(4,1)$

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

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

→ 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 ←

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

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.

$A \quad \underline{1,000,000}$

$\downarrow \qquad \downarrow$

$B \quad 1,000$

$A_1 \qquad 1,000$

$A_2 \qquad 1,000$

$\vdots \qquad \vdots$

$A_{100} \qquad 1,000$

$99,000$
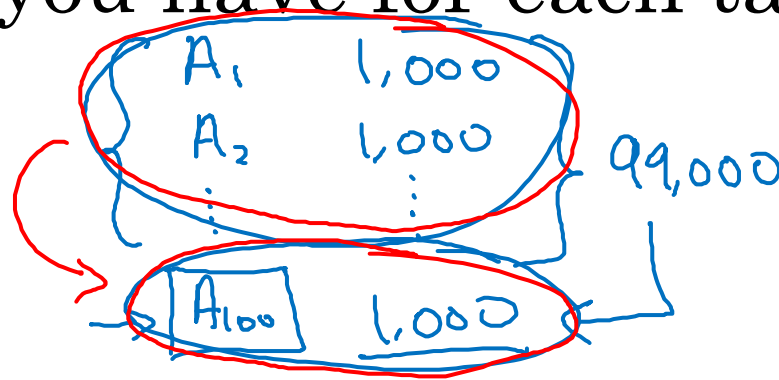
- Can train a big enough neural network to do well on all the tasks.