Learning from multiple tasks

Multi-task learning

deeplearning.ai

Simplified autonomous driving example

Pedestrians

Cars

Stop Signs

Traffic lights
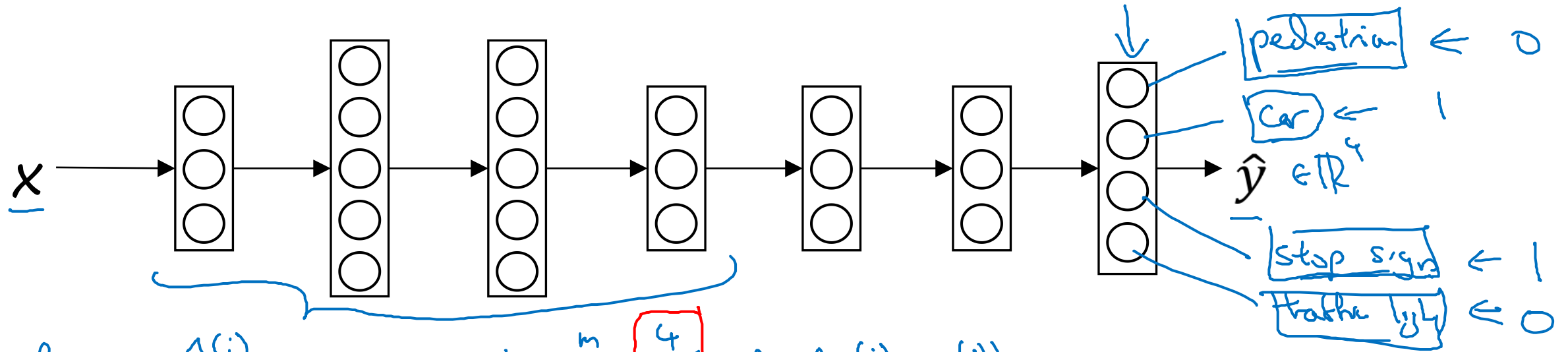
$\vdots$

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

$X^{(i)}$

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

$(4, m)$

Andrew Ng

Neural network architecture

$x \rightarrow$ [ ] [ ] [ ] [ ] [ ] [ ] [ ] $\rightarrow \hat{y} \in \mathbb{R}^4$

pedestrian $\leftarrow$ 0
Car $\leftarrow$ 1
stop sign $\leftarrow$ 1
traffic light $\leftarrow$ 0

$$\text{Loss}: \quad \overset{\wedge (i)}{y} \quad \rightarrow \quad \frac{1}{m} \sum_{i=1}^{m} \boxed{\sum_{j=1}^{4}} \mathcal{L}\left(\hat{y}_j^{(i)}, y_j^{(i)}\right)$$

(4,1)

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

$\hookrightarrow$ Usual logistic loss

$$-y_j^{(i)} \log \hat{y}_j^{(i)} - (1 - y_j^{(i)}) \log (1 - \hat{y}_j^{(i)})$$
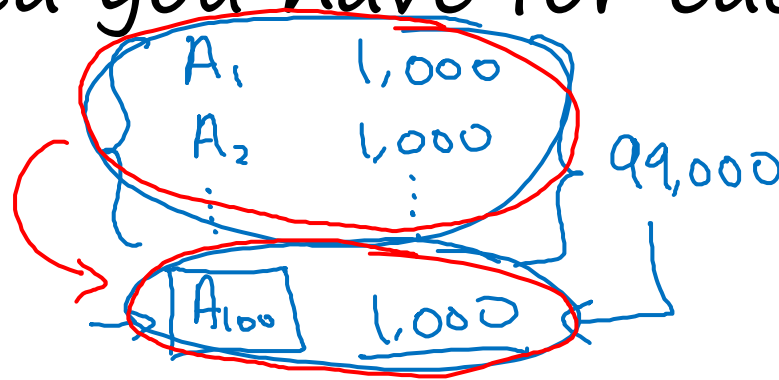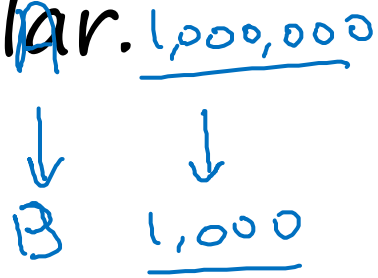
Unlike softmax regression:
One image can have multiple labels

Multi-task learning $\leftarrow$

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

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.

$$1,000,000$$
$$A \downarrow \qquad \downarrow$$
$$B \quad 1,000$$

$$
\begin{array}{ll}
A_1 & 1,000 \\
A_2 & 1,000 \\
\vdots & \vdots
\end{array}
\quad 99,000
$$

$$A_{100} \quad 1,000$$

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