- Transfer learning was a sequential process, learn from A transfer to B, A → B
- Multitask learning parallelizes this

Also → these can be thought of as methods to use when you dont have much data for predictions

deeplearning.ai

Learning from multiple tasks

Multi-task learning

Simplified autonomous driving example

An autonomous car needs to identify

- Pedestrians
- cars
- Stop signs
- Traffic lights

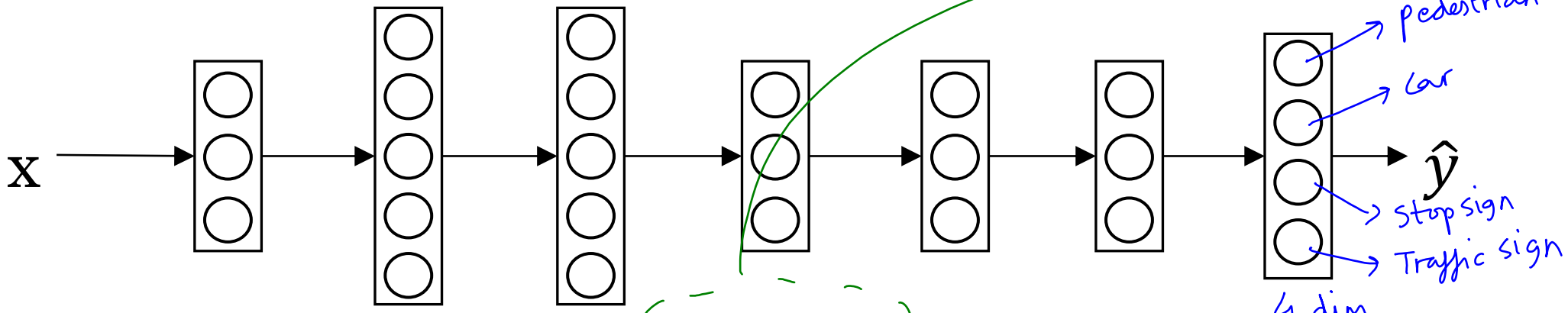
for proper operation

- If Image = $x^{(i)}$ then o/p has 4 labels → $y^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$ (4×1) = [pedestrian, car, stop sign, Traffic light]

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

$Y = 4 \times m$

Neural network architecture



If Y is only partially labelled

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

→ then just omit the values of y_j that dont have 0/1

Multi task learning will still work!

Loss: $\frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{4} L\left(y_j^{(i)}, \hat{y}_j^{(i)}\right)$,

$L\left(y_j^{(i)}, \hat{y}_j^{(i)}\right) = -y_j^{(i)} \log \hat{y}_j^{(i)} - (1 - y_j^{(i)}) \log (1 - \hat{y}_j^{(i)})$

(logistic loss)

- Diff b/w this & Softmax in the last layer

Softmax assigned a single label to a single example, this can have multiple labels per example → can have car & traffic lights.

(either a cat or dog or rabbit or horse)

This is doing multi task learning

- solves 4 problems
  - Is there a car? Is there a STOP sign? Is there pedestrian etc.
- could have used 4 diff NN – 1 to solve each problem, but since here the features may be shared eg stop sign occurs together w/ traffic lights OR pedestrians are often in the Img having traffic light

Hence it may be better to have it in 1 NN.

- Also more expensive to do 4 v/s 1

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