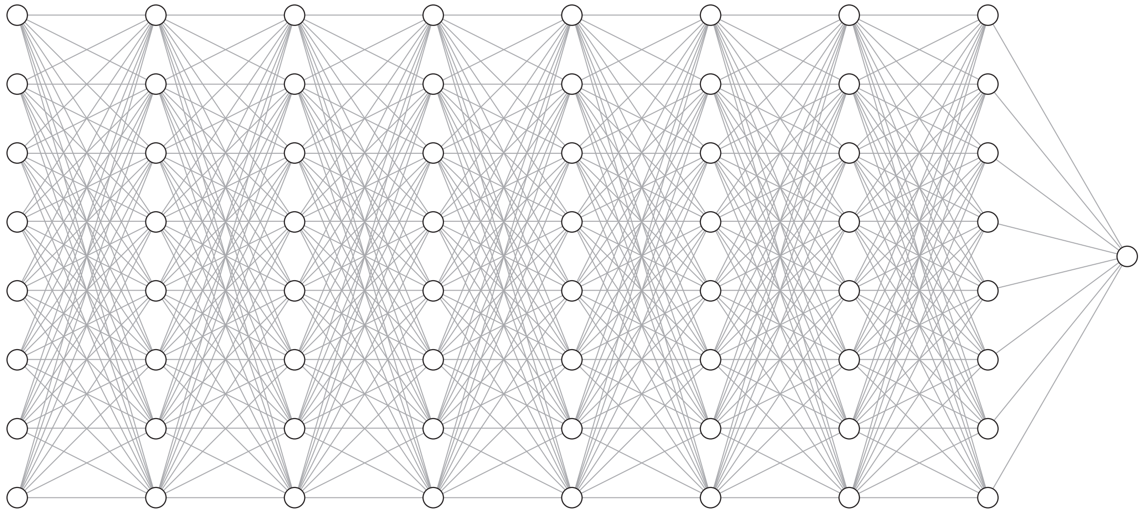
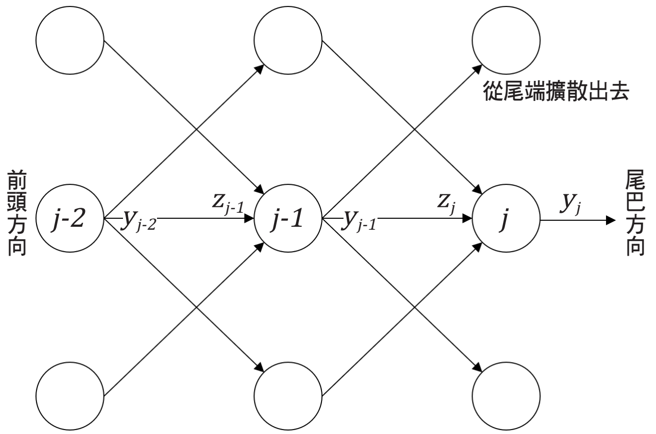
2-4-4 節

Bonus

反向傳播
Backpropagating





取上圖神經網路的一個局部

start!

$$\frac{\partial \mathcal{L}}{\partial z_j} = \frac{dy_j}{dz_j}\frac{\partial \mathcal{L}}{\partial y_j} =$$

iterate

$$\frac{\partial \mathcal{L}}{\partial y_{j-1}} = \sum_j \frac{dz_j}{dy_{j-1}}\frac{\partial y}{\partial z_j}$$

$$\frac{\partial \mathcal{L}}{\partial w_{j-1,j}} = \frac{\partial z_j}{\partial w_{j-1,j}}\frac{\partial \mathcal{L}}{\partial z_j}$$

從神經網路的任何一個 neuron j 開始，一開始 (上圖的 start!) 我們先計算損失函數 $\mathcal{L}$ 對該 neuron j (以下簡稱 j) 的輸出 y_j 的變化率 $\frac{\partial \mathcal{L}}{\partial y_j}$，運用 chain rule 我們就可以再算出 $\mathcal{L}$ 對 j 的輸入 z_j 的變化率：

$$\frac{\partial \mathcal{L}}{\partial z_j} = \frac{\partial y_j}{\partial z_j} \cdot \frac{\partial \mathcal{L}}{\partial y_j}$$

其中 $\frac{\partial \mathcal{L}}{\partial y_j}$ 是已知，所以我們只要算 $\frac{\partial y_j}{\partial z_j}$ 就好了，而 y_j 是在 forward pass 就已經算好了。所以 $\frac{\partial y_j}{\partial z_j}$ 很快就可以算出來。而 $\frac{\partial \mathcal{L}}{\partial z_j}$ 就只是二者相乘就好了。

接著來算前一個 (從上圖，z_j 前一個就是 y_{j-1})：

$$\frac{\partial \mathcal{L}}{\partial y_{j-1}} = \sum_j \frac{\partial z_j}{\partial y_{j-1}} \cdot \frac{\partial \mathcal{L}}{\partial z_j}$$

為什麼要 $\sum_j$ 呢？因為 z_j 對 $\mathcal{L}$ 的影響會從 j 那一層的各 neuron 往尾端的各 neuron (各路線) 擴散出去，所以要每條連線都要加起來。這其實是多變數 chain rule 的概念。

然後就以此類推可以算出：

$$\frac{\partial \mathcal{L}}{\partial z_{j-1}} = \frac{\partial y_{j-1}}{\partial z_{j-1}} \cdot \frac{\partial \mathcal{L}}{\partial y_{j-1}}$$

這裡一樣把已算好的 $\frac{\partial \mathcal{L}}{\partial y_{j-1}}$ 拿來乘就好, 而 y_{j-1} 已經在 forward pass 算好了, 所以只要算 $\frac{\partial y_{j-1}}{\partial z_{j-1}}$ 就好了！

同樣的:

$$\frac{\partial \mathcal{L}}{\partial y_{j-2}} = \sum_{j-1} \frac{\partial z_{j-1}}{\partial y_{j-2}} \cdot \frac{\partial \mathcal{L}}{\partial z_{j-1}} \qquad \text{以此類推...}$$

所以這樣從 $j \to j-1 \to j-2 \to$ ..... 就可以把全部 neuron 的輸入對 $\mathcal{L}$ 的影響率 $\frac{\partial \mathcal{L}}{\partial z_j}$ 和輸出對 $\mathcal{L}$ 的影響率 $\frac{\partial \mathcal{L}}{\partial y_j}$ 都算出來, 這就是使用偏微分 chain rule 做反向傳播的過程。

但是, 不要忘了！我們要的是 $\frac{\partial \mathcal{L}}{\partial w_j}$！因為我們要用 $w_j = w_j - \eta \frac{\partial \mathcal{L}}{\partial w_j}$ 來改變一點點 w_j 值, 以便做 SGD 啊！算 $\frac{\partial \mathcal{L}}{\partial w_j}$ 可以說是前述反向傳播的副產品, 因為：

$$\frac{\partial \mathcal{L}}{\partial w_{j-1,j}} = \frac{\partial z_j}{\partial w_{j-1,j}} \cdot \frac{\partial \mathcal{L}}{\partial z_j}$$

其中 $\frac{\partial \mathcal{L}}{\partial z_j}$ 我們已經算出來了, 而 z_j 和 $w_{j-1,j}$ 的關係是：

$$z_j = w_{j-1,j} \cdot y_{j-1} + b_{j-1,j}$$

$$\Rightarrow \frac{\partial z_j}{\partial w_{j-1,j}} = y_{j-1}$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial w_{j-1,j}} = y_{j-1} \cdot \frac{\partial \mathcal{L}}{\partial z_j}$$

其中 $\frac{\partial \mathcal{L}}{\partial z_j}$ 已算好, y_{j-1} 也在前向傳播 (forward pass) 時算好了, 所以 $\frac{\partial \mathcal{L}}{\partial w_{j-1,j}}$ 就二者相乘就好了！這就是反向傳播的原理。

*參考資料：Geoferry Hinton 的 YouTube 教學, 網址：請上 youtube 搜尋 Geoffery Hinton Lecture 3.4