

1. (2%) 請以中文說明一下 lifelong learning 的中心概念是什麼?

lifelong learning 關注的是構建一套算法或者系統，使之能持續不斷的學習，並能從過去的學習任務中積累經驗和知識，來幫助解決新任務。但深度學習領域一直存在一個比較嚴重的問題——“災難性遺忘”，即一旦使用新的數據集去訓練已有的模型，該模型將會失去對原數據集識別的能力。lifelong learning 便主要關注如何再不持續遺忘的狀態下持續學習解新任務的方法。

2. (2%) 列出 EWC, MAS 的作法是什麼? 根據你的理解，說明一下大概的流程該怎麼做 (不要貼 code)。

$$\frac{L'(\theta)}{\theta_i} = \frac{L(\theta)}{\theta_i} + \lambda \sum_i \frac{b_i}{\theta_i} (\theta_i - \theta_i^b)^2$$

EWC 的基礎思想是把已經訓練好的模型中的比較重要的參數用正則化項保護起來，讓它們變得不那麼容易被更新，從而舊的知識就不會被完全洗掉。我們在基礎的損失函數上增加一個守門員，每個參數受到自己在上一個任務訓練完畢後，最後的參數值的約束。lambda 是一般的係數，其中的 F 評估參數重要程度。在我們訓練的參數中，有的參數是不重要的，但是有的參數是十分重要的， $\theta^b$  反應的是這個模型中所有的參數。如果  $b_i$  值大，就意味著這個參數是十分重要的，改變的話對之前的任務學習就會忘記了。 $\theta_i$  反應的是我們更新後的參數，之後  $\theta_i^b$  反應的是我們更新前的參數。我們原來的 loss function 僅僅是對參數  $\theta$  進行更新，但是沒有考慮這個參數是否是重要的。我們新的 loss function 就考慮了這一點，在之前的條件下，我們加入了新的一項。如果  $b_i$  很大的話，後一項如果括號中的變大，就會使新的損失函數瞬間變大，如果  $b_i$  很小的話，即使括號中的內容變大，損失函數依然不會很大。

$$F = [\nabla \log(p(y_n|x_n, \theta_A^*)) \nabla \log(p(y_n|x_n, \theta_A^*))^T]$$

上式所提到的 F 即前面所提的  $b_i$ 。 $p(y_n|x_n, \theta_A^*)$  指的就是模型在給定之前 task 的 data  $x_n$  以及給定訓練完 task A 存下來的模型參數  $\theta_A^*$  得到  $x_n$  對應的 label 的 posterior probability. 也就是對這個  $p(y_n|x_n, \theta_A^*)$  取 log 再取 gradient 並且平方。

而 MAS 大部份和 EWC 相同，但是重新定義了參數重要性的估量，加入了 unsupervised 的 setting。將  $b_i$  改為  $\Omega_i$ ，對最後模型的最後一層做 l2 norm 後平方再對各自的 weight 取 gradient 再絕對值。

$$\Omega_i = \left\| \frac{\partial \ell_2^2(M(x_k; \theta))}{\partial \theta_i} \right\|$$

先前討論的方法需要 supervised data 基於 loss 的重要性權重估計，因此僅限於可用的 training label data。相比之下，MAS 可以在 unsupervised 的情況下估算重要性權重保留的 data，因此能夠進行特定的參數調整。

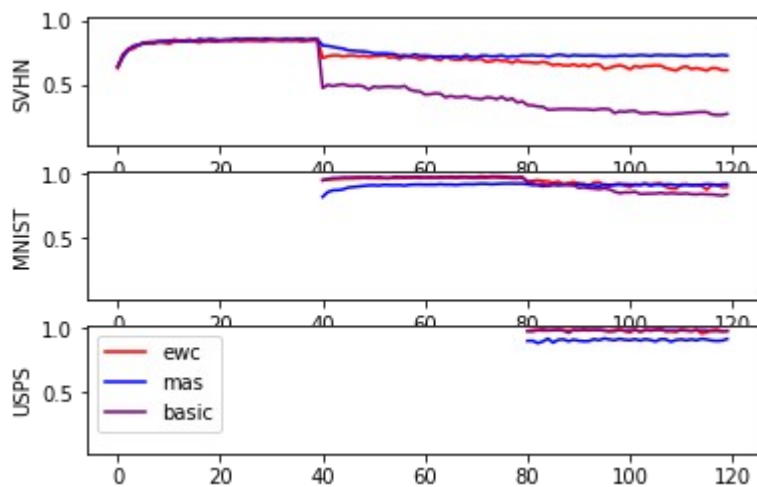
3. (1%) EWC 和 MAS 所需要的資料最大的差異是什麼？

MAS 可以是 unsupervised 的 data，但 EWC 一定是要有 label 的 supervised data。

4. (5%) 秀出 part1 及 part2 最後結果比較圖，並分析一下結果，以及你跑的實驗中有什麼發現。

( EWC, MAS, baseline 比較圖 (2%) 與 EWC、MAS、SCP (或是你自己實做的演算法)、baseline 比較圖 (3%) )

part 1.



由 part 1 的圖能觀察到，在 SVHM 這個 task 裡一開始大家的 accuracy 都相同，但當將同一個 model train 到 MNIST 上時，在 SVHM 這個 task，雖然大家都有下降，但 basic 的下降幅度最大，直接爆掉，而 ewc 雖然只有小幅下降，但再剛進入 MNIST 的 training 時，似乎還是有一定的降幅，可能是上述中的 F 雖然進行守衛，但仍然遺失部分重要的 weight，而 mas 似乎從同到瑋都有守住，更動重要的 weight 較少，當中的原因應該是我的 EWC 守衛的  $\lambda$  可能要再高些。USPS 以及剛開始 train mnist 則能觀察出，mas 先天，再還沒進入新的 task 時本身的 accuracy 就較低，所以或許 MAS，可能在更動 weight 侷限性較大。