

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

通常為了處理一個 task 我們可能去訓練了一個模型，而不同的 n 個 task 也就會去訓練出相對應的 n 種模型，而 lifelong learning 的中心概念在於不同的 task 都使用同一個模型去做訓練，並且可以同時處理它所學習過的 task，那在做 lifelong learning 主要會面臨到的問題是 knowledge retention、knowledge transfer、model expansion。

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)。

EWC：

核心概念在於不去調整對於 previous task 重要的參數，而去調整那些相對來說比較不重要的。

loss function :

$$\mathcal{L}_B = \mathcal{L}(\theta) + \sum_i \frac{\lambda}{2} F_i (\theta_i - \theta_{A,i}^*)^2$$

F 的實做方法：

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

F_i 對應到第 i 個參數的守衛，值愈大就代表說此參數對於 previous task 來說愈重要，因此就會盡量不去更動它，而每一個參數都可以用 backward 再取 gradient 的性質來計算出各自對應到的 F。

MAS：

核心概念與 EWC 雷同，只是計算 important weight 方式不同。

loss function :

$$\mathcal{L}_B = \mathcal{L}(\theta) + \sum_i \frac{\lambda}{2} \Omega_i (\theta_i - \theta_{A,i}^*)^2$$

Ω 的實做方法：

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

Ω 的作法就是對最後模型的 output vector 做 l2 norm 後取平方，再對各自的 weight 微分，並且取該 gradient 的絕對值。

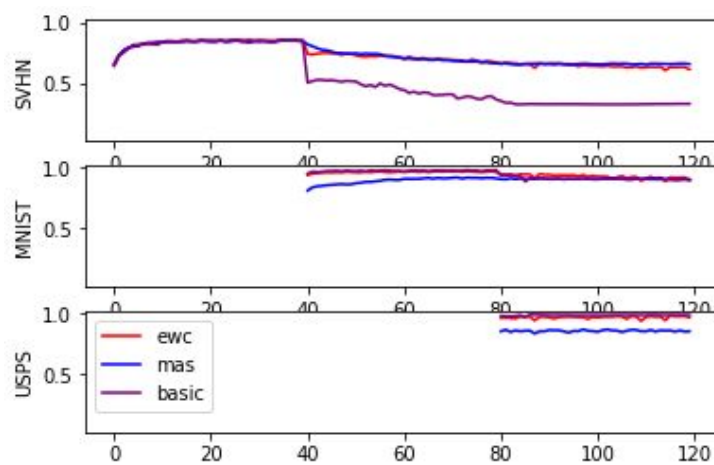
3. (1%) EWC 和 MAS 所需要的資料最大的差異是什麼？

EWC 需要計算參數對於 loss 的二次微分，因此需要有 label 的 data，而 MAS 則不需要。從上面的 F 、 Ω 公式也可看出前者需要 x 、 y ，後者則需要 x 。

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。

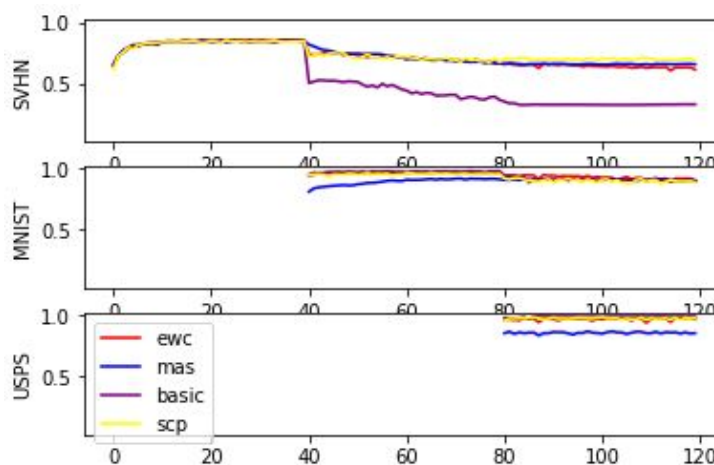
(EWC, MAS, baseline 比較圖 (2%) 與 EWC、MAS、SCP (或是你自己實做的演算法)、baseline 比較圖 (3%))

part1 :



EWC、MAS、baseline 的 λ 分別為 4.5、200、0。可發現 baseline 在訓練 SVHN 的整體表現較差，而 MAS 則是在訓練 USPS 的表現較差。

part2 :



EWC、MAS、baseline、SCP 的 λ 分別為 4.5、200、0、12。可發現 SCP 在三種狀況的表現和 EWC 的滿相近的，在訓練三種情況後的表現都不錯。