

Tuning Hyperparameters With Bayesian Optimization

自動調整深度學習模型超參數之研究

自動調整深度學習模型超參數之研究

指導教授：謝孫源

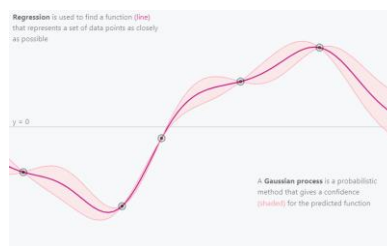
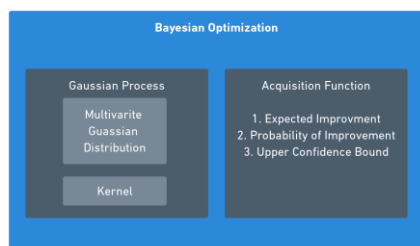
專題成員：黃濬程

開發工具：Python、Colaboratory

測試環境：Windows 10 家用版

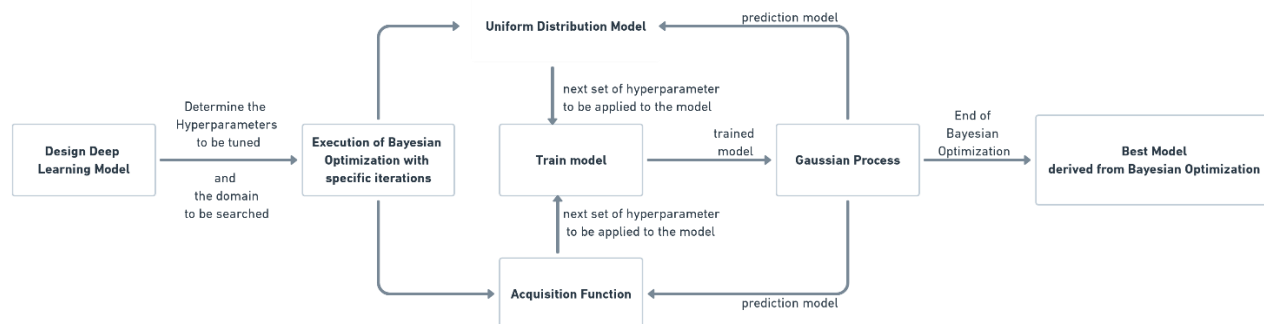
一、簡介：

在 ML, DL 快速發展的時代底下，縮短訓練一個深度學習模型的生命週期是很重要的課題。調整超參數便是其中一個方式。但調整超參數是不簡單的，因為這是一個高維度的問題，所以人類是較難去想像的。而 **Bayesian Optimization** 提供了 **programmer** 一個機率的方式去嘗試。透過這個演算法，訓練深度學習模型的過程不僅能更自動化也能更有效率。



而在這個專題中，除了深入了解背後的數學意義外；我也利用 **python** 和一些數學 **package** 來 **implement Bayesian Optimization**。並且在實現 **Bayesian Optimization** 後，我將其套用在多個 **model** 上；並且在這幾個例子中研究它的優點和缺點，最後針對其最大的缺點進行改善 - 也就是卡在局部最大值或局部最小值的問題。

以下為系統架構圖：



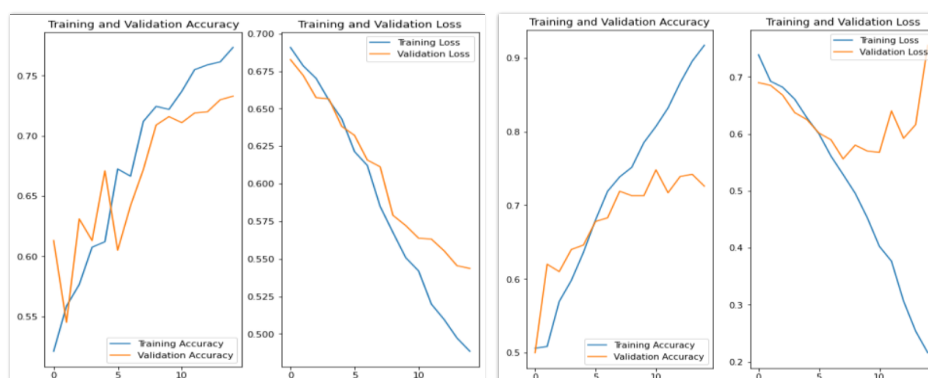
二、測試結果：

詳細狀況見 PPT 或是大專生計劃書，這裡僅列出其中兩個代表的測試結果。

➤ Application On KNN classifying dogs and cats

在這個例子中，Bayesian Optimization 試圖透過調整該 KNN Model 的 “number of dense layers” 和 “dropout rate of each dense layer” 來最佳化 Model 在 validation loss 上的表現。

下面兩張圖片為結果圖，左側的圖是 Bayesian Optimization 所挑出的 Model，右側的圖則是開源網站上其他開發者用同一組 dataset 訓練得到的。可以發現 Bayesian Optimization 的版本大幅改善 Model 在 Validation loss 上的表現；同時也代表 Bayesian Optimization 改變了 overfitting 的狀況。



➤ Problem of Local Maximum

當 Bayesian Optimization 搭配的是 Exploitation 特性較強的 Acquisition Function 時，由於 Exploration 能力較差，常會有卡在局部最大值或局部最小值的問題；而以下便是我的解決方案。

Solution: 在 Acquisition Function 開始決定下組嘗試的超參數前，Bayesian Optimization 會透過 uniform distribution model 去對目標函數取樣以彌補 Exploration 較差的劣勢。(類似於 Grid Search 的概念)

下圖為上述改善方案的結果圖，可以看到 Bayesian Optimization 在左上角的圖中已經有 10 個 uniform 的 datapoint；最後也在此 non-convex function 中找到 Global Maximum (原先的 Bayesian Optimization 版本會卡在局部最大值)

