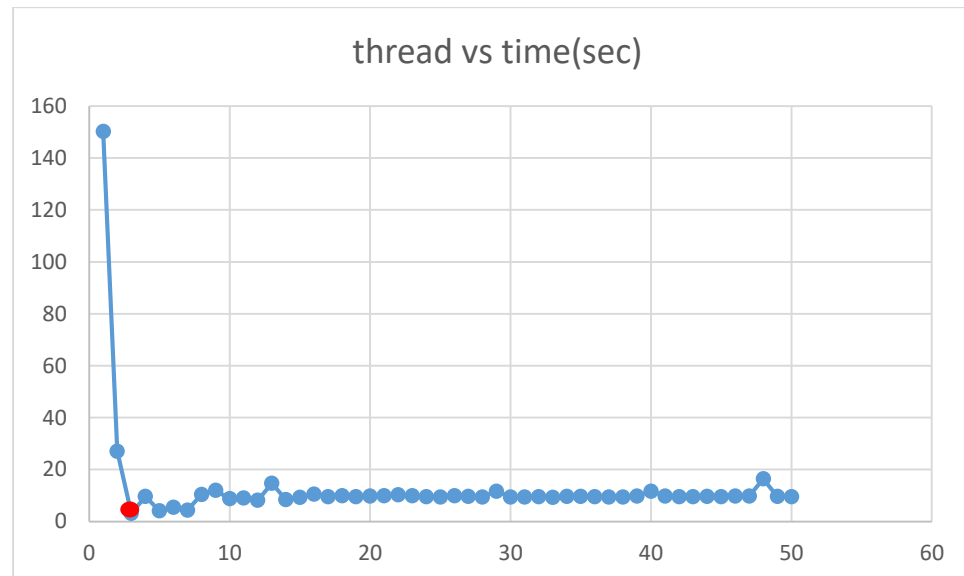


## HW4 Report b06902104 吳由由

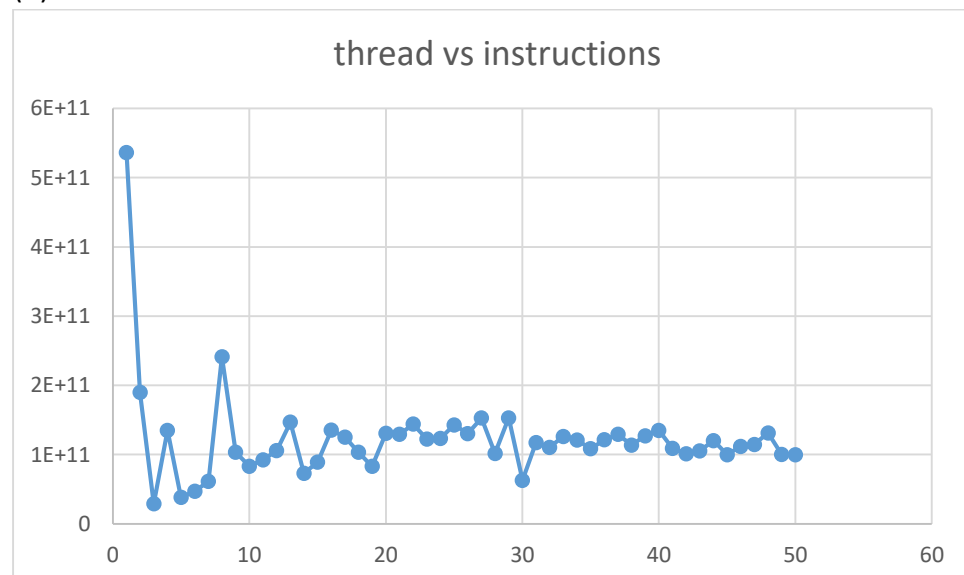
(b). thread 開在 train decision tree 的地方，一個 thread train 一棵 tree，thread 做完後再找下一棵還沒 train 過的樹來 train，因為在 train decision tree 時，tree 之間彼此是平行的，所以可以用 thread 分工。

(c). 1000 棵樹



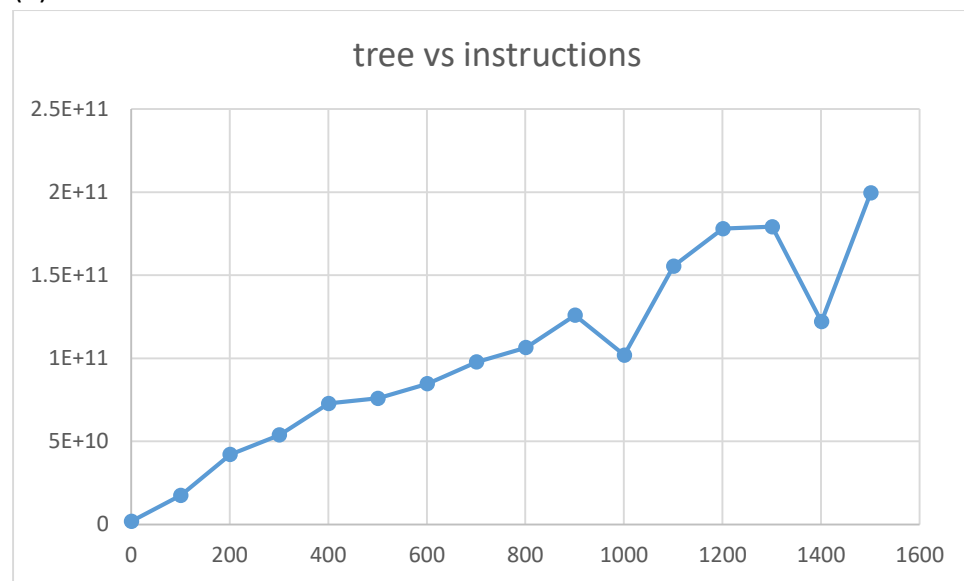
當 thread 的數量只有 1 時，整個 train 花非常多時間，隨著 thread 數量提升後執行速度大幅下降，但達到最佳執行速度後更多的 thread 也不會讓整個 train 的過程更省時。

(d).



當到達最佳 thread 數之前，instruction 大約隨著 thread 數增加而減少，在最佳執行速度的 thread 數時 instruction 也達最低點，在那之後 thread 和 instruction 數大約沒有相關維持在一定值附近，thread 數量增加不對 instruction 有正/負相關影響。

(e).



從這張圖可以看出 tree 的數量越多，instruction 的數量越多，tree 的數量和 instruction 的數量大約呈現正相關

(f). 我發現並不是資料量越大 train 的 tree 越多就保證正確率會越高，當資料量多到一定的程度時可能會產生 overfitting, 如果有很多 tree，他們的結果可能會互相干擾，有時候會產生正確率非常差的結果(1x%)。一棵樹的 training data 少一點，用多棵樹投票可以達到每次正確率都高於 85%，確保不會突然出現特別差的結果。