

1. 上課有提到利用牛頓法來找到最佳解，但是以最佳化的角度來看，計算 Hessian matrix 是很貴的，應該不會比 stochastic gradient descent 來得快，而且因為 stochastic gradient descent 是可以一個 batch 的資料來去求解，但是 Hessian matrix 應該沒有這種性質可以做到 streaming learning?