

Unanswered Questions

課程中提到了梯度下降法 (Gradient Descent) 和牛頓法 (Newton's Method) 兩種尋找最佳解 θ 的方法。牛頓法需要計算並反轉一個 Hessian 矩陣，計算成本看似更高。在什麼樣的情況下，使用牛頓法會比梯度下降法更有優勢？它們各自的收斂速度和穩定性有何不同？