Gradient descend 的方法

之前，我們是先沿著 x 軸做最佳化，再沿著y軸，再沿著x軸，以此類推

但其實方向不一定要限於 x, y 軸，且不需只對兩個方向做輪流

假設，對目標方程式 f(x, y) 而言，若目前的點是 x0, y0

且對 x, y 偏微分的結果為

f(x, y) /x  = a1, f(x, y) /y = a2

at (x, y) = (x0, y0)

則方向應該選為 (a1, a2)，然後在

(x0 + c\*a1, y1 + c\*a2) 的直線上

找出 c 使得 f(x, y) 為最小

接下來再設 (x0 + c\*a1, y1 + c\*a2) 為新的 (x0, y0)，重覆以上的步驟

直到沿著 x, y 的偏微分皆近於0為止

(但在機器學習上會稍微有點不同，根本不需找到最佳的 c

直接設 c = *λ*

*λ* 被稱作是 learning rate

*λ* 值會隨著 iteration 的次數而減小)