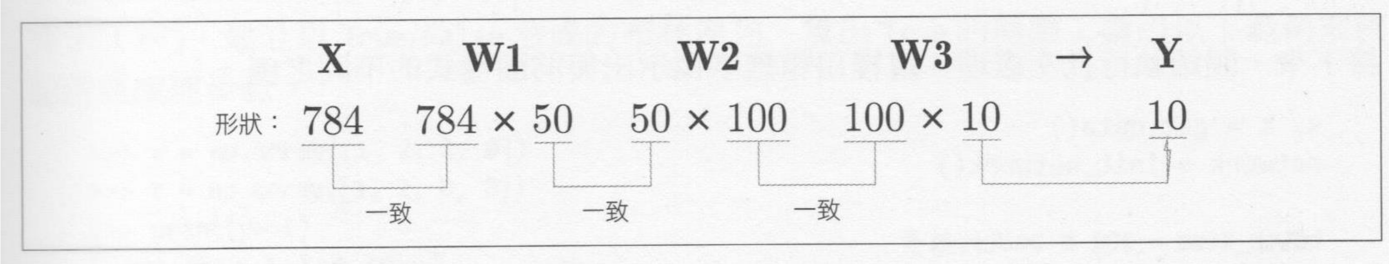
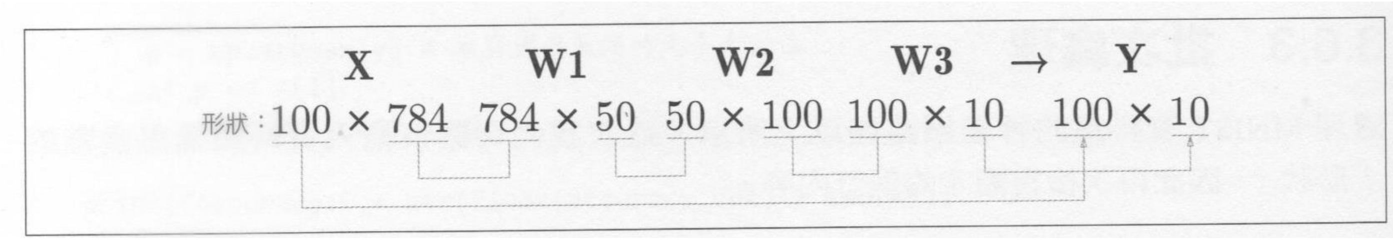
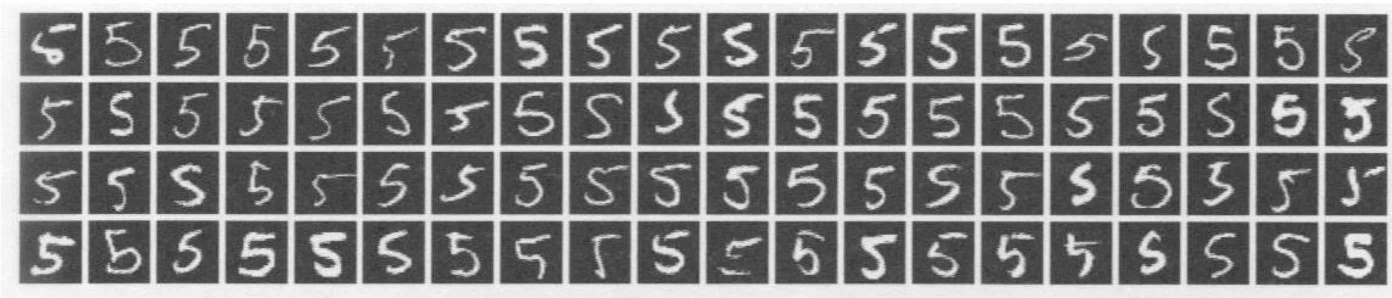
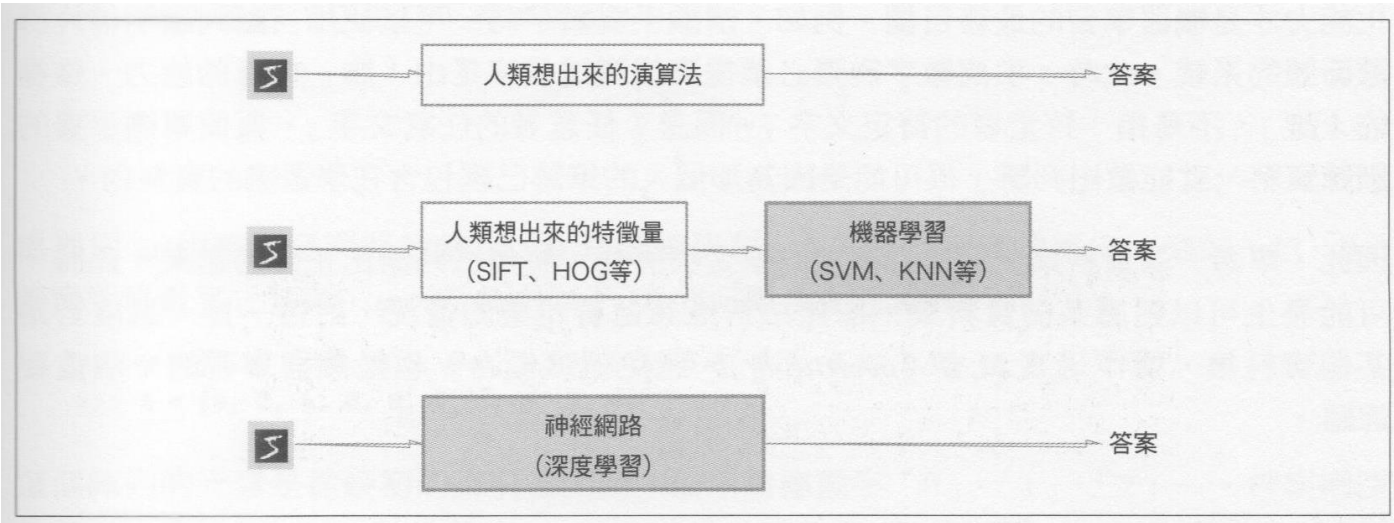
批次處理(batch)



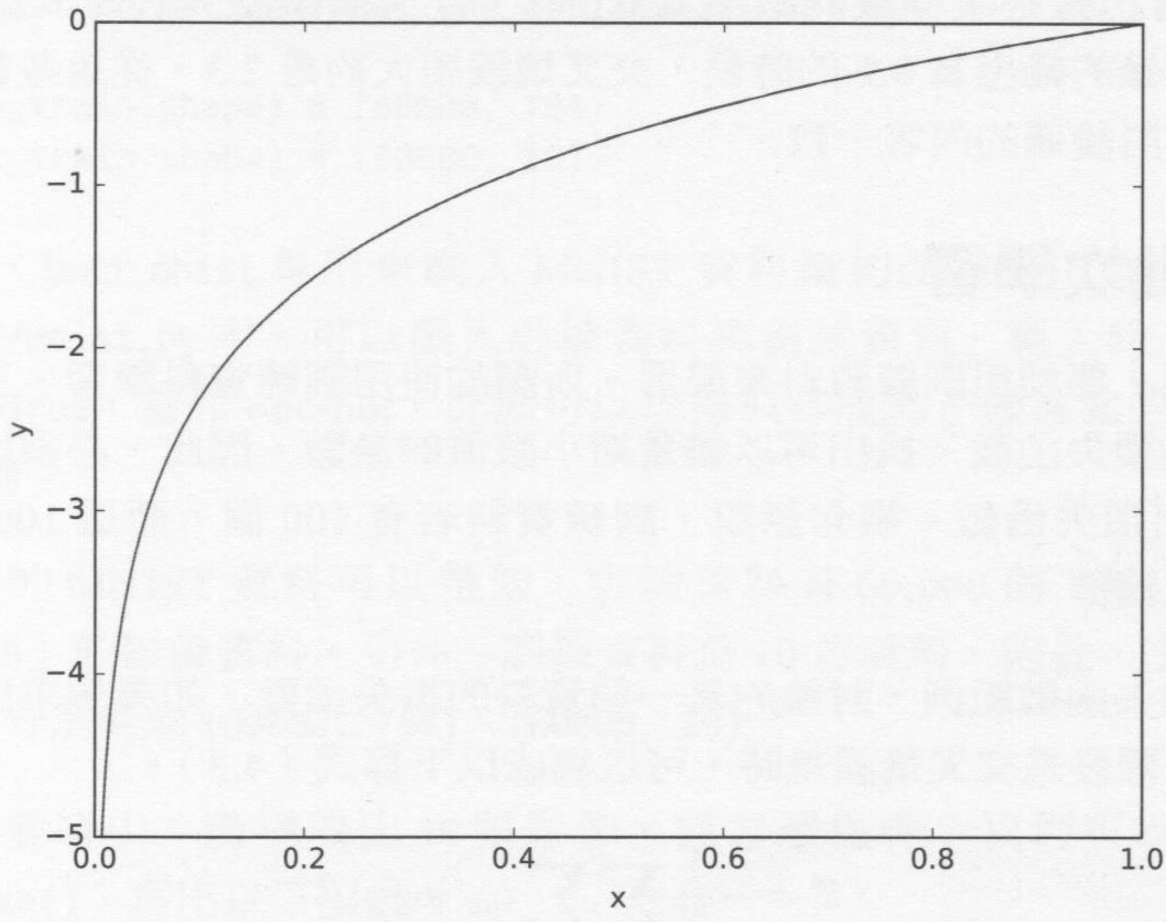



$$E = \frac{1}{2}\sum_{k}(y_k - t_k)^2$$

```
>>> y = [0.1, 0.05, 0.6, 0.0, 0.05, 0.1, 0.0, 0.1, 0.0, 0.0]
>>> t = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
```

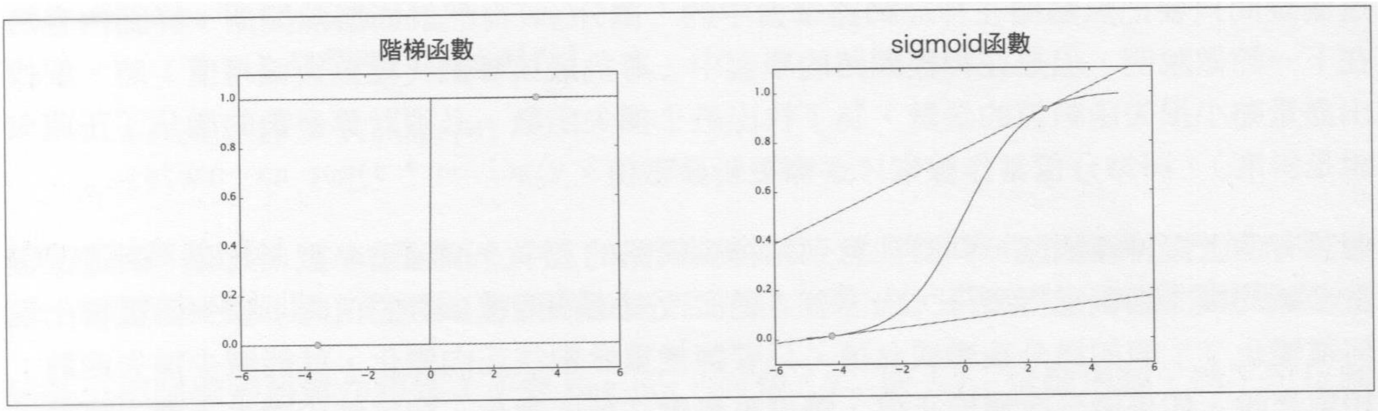
$$E = -\sum_{k} t_k \log y_k$$

log 是底為 $e$ 的自然對數（$\log_e$），$y_k$ 是神經網路的輸出，$t_k$ 是正確答案標籤。另外，$t_k$ 是只有正確答案標籤的索引值為 1，其餘為 0（one-hot）。因此，這個算式（4.2）實質上只有計算對應正確答案標籤為 1 的輸出之自然對數。假設「2」是正確答案標籤的索引值，對應的神經網路輸出為 0.6，交叉熵誤差是 $-\log\ 0.6 = 0.51$。另外，「2」的輸出為 0.1 時，會變成 $-\log\ 0.1 = 2.30$。換句話說，交叉熵誤差會隨著正確答案標籤的輸出結果而改變。

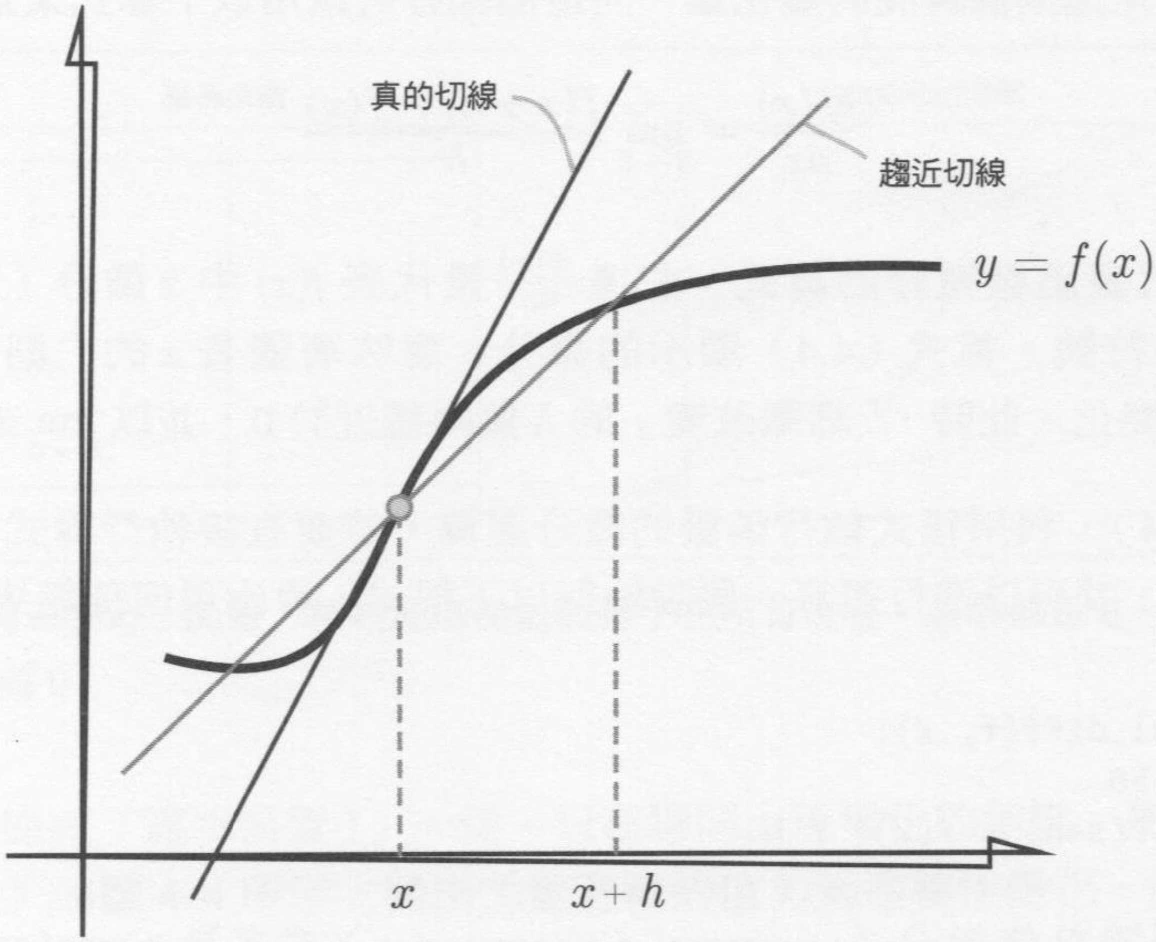



自然對數 $y = \log\ x$ 的圖表

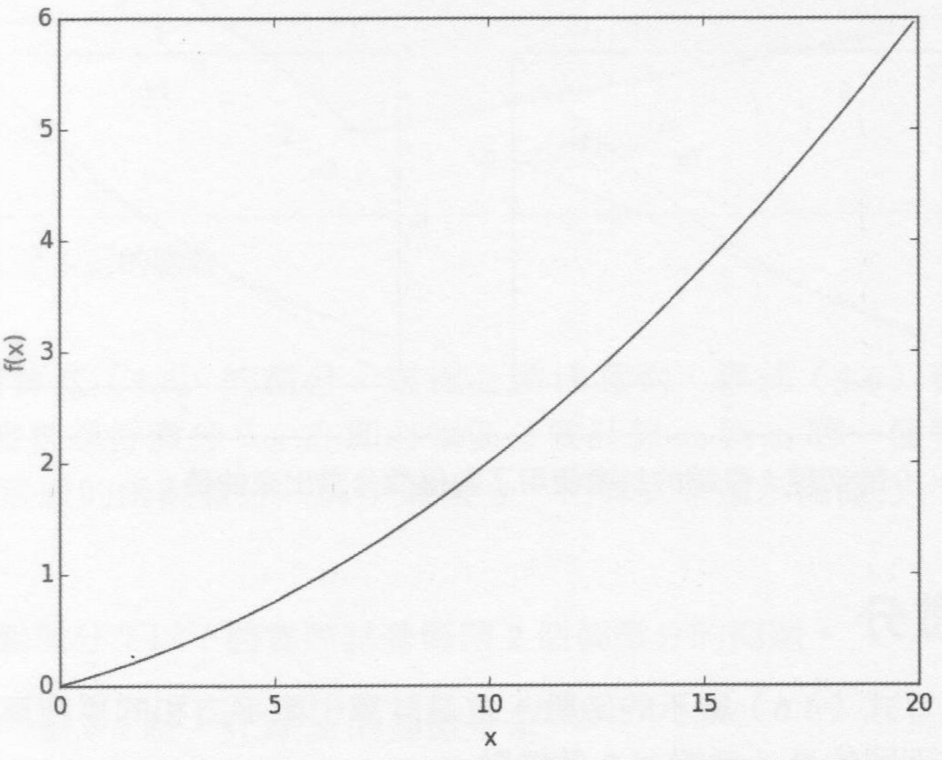
$$E = -\frac{1}{N}\sum_{n}\sum_{k} t_{nk} \log y_{nk}$$



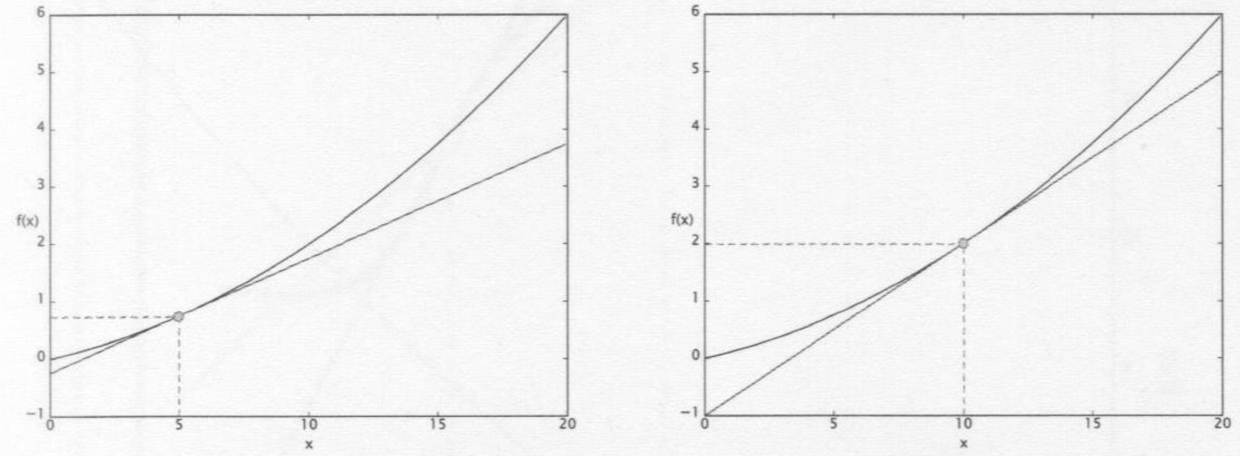

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



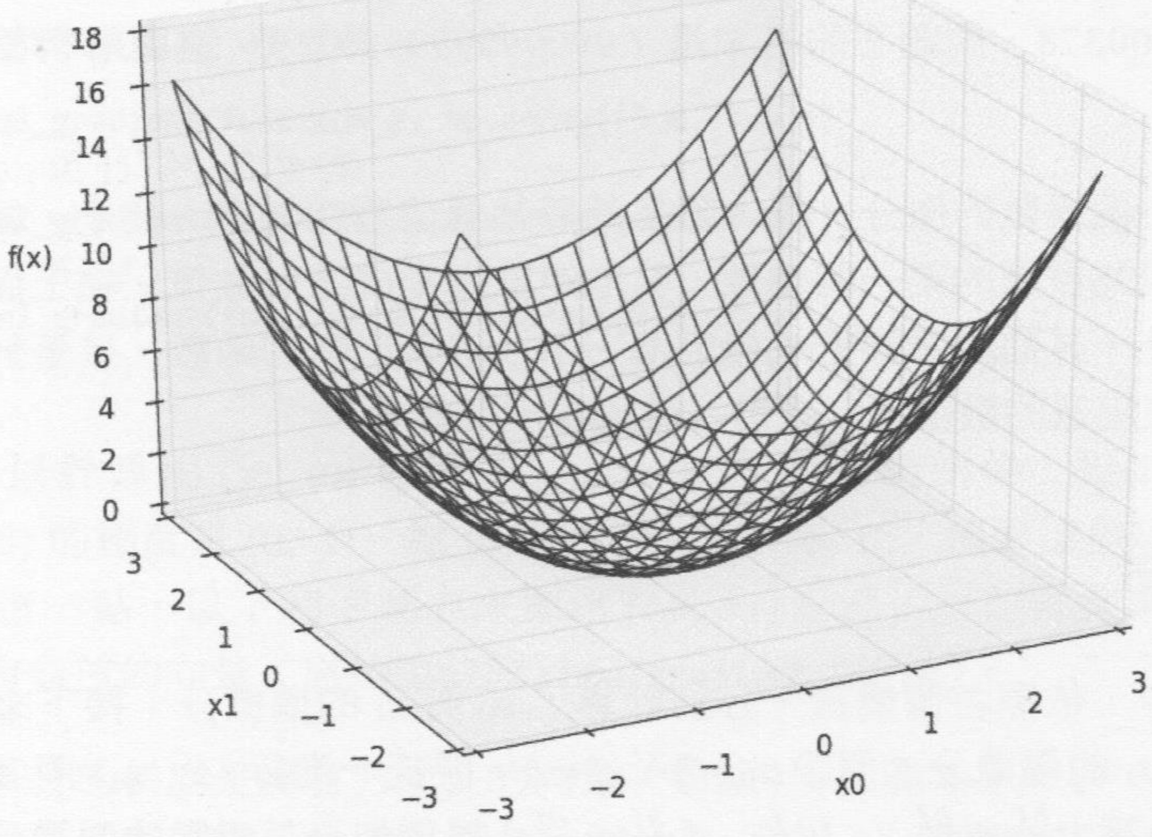

真的微分（真的切線）與數值微分（趨近切線）的值不同



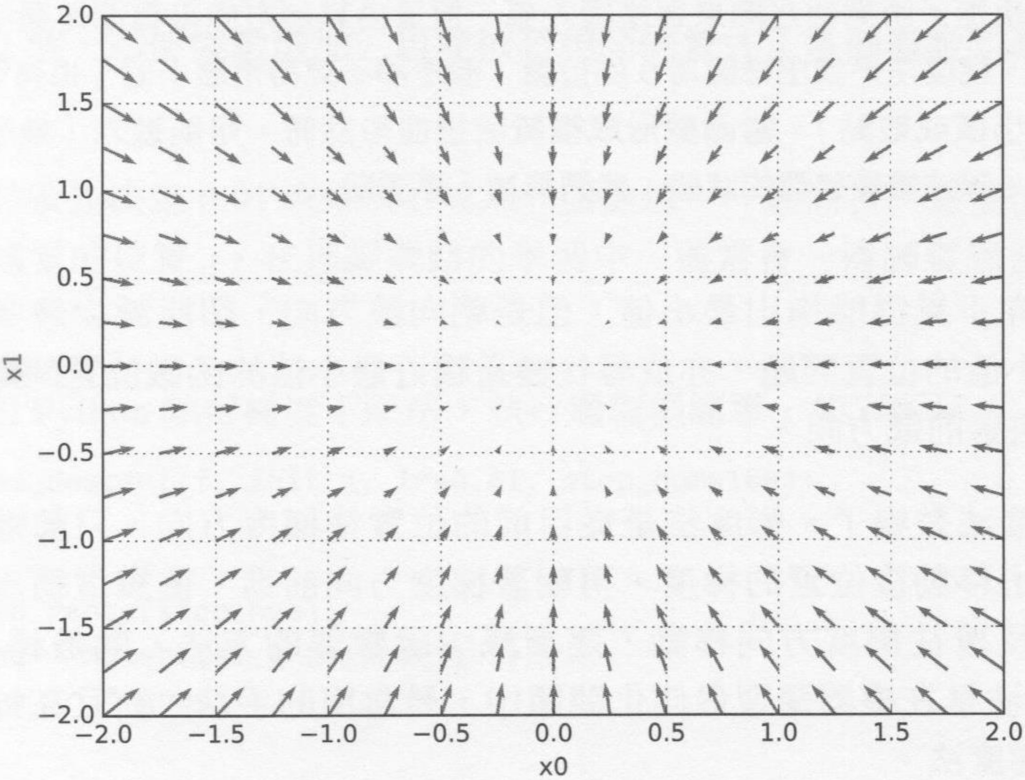

$f(x) = 0.01x^2 + 0.1x$ 的圖表




$x=5$、$x=10$ 的切線：直線的斜率使用了數值微分求出來的值



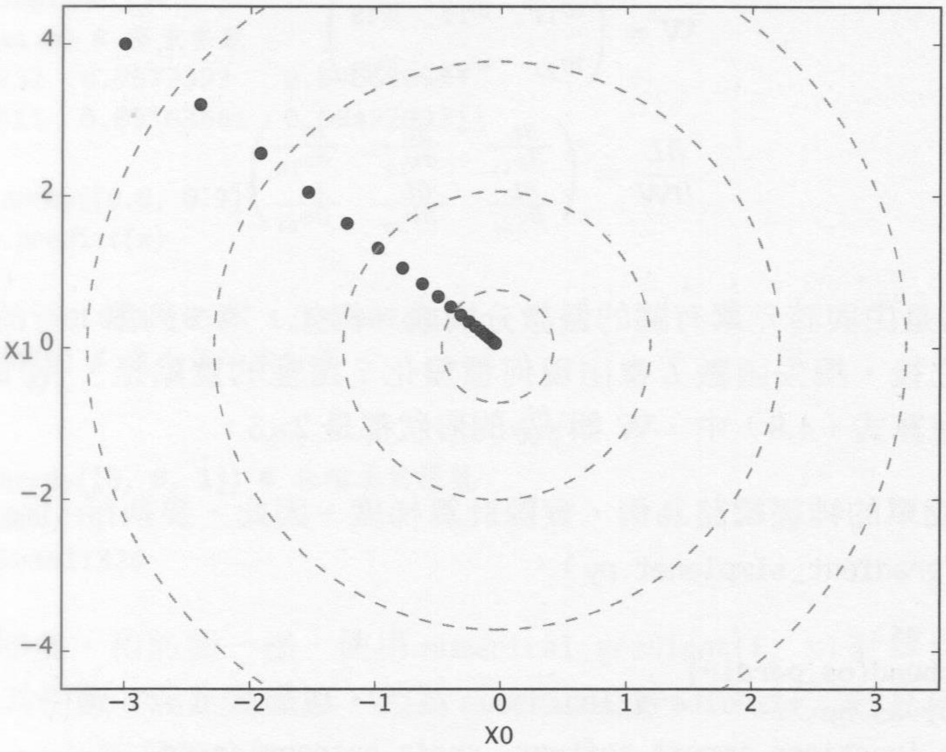

$f(x_0, x_1) = x_0^2 + x_1^2$的圖表




$f(x_0, x_1) = x_0^2 + x_1^2$的梯度

$$x_0 = x_0 - \eta \frac{\partial f}{\partial x_0}$$

$$x_1 = x_1 - \eta \frac{\partial f}{\partial x_1}$$




$f(x_0, x_1) = x_0^2 + x_1^2$ 的梯度法更新流程：虛線代表函數的等高線

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$$

$$\frac{\partial L}{\partial \mathbf{W}} = \begin{pmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \end{pmatrix}$$