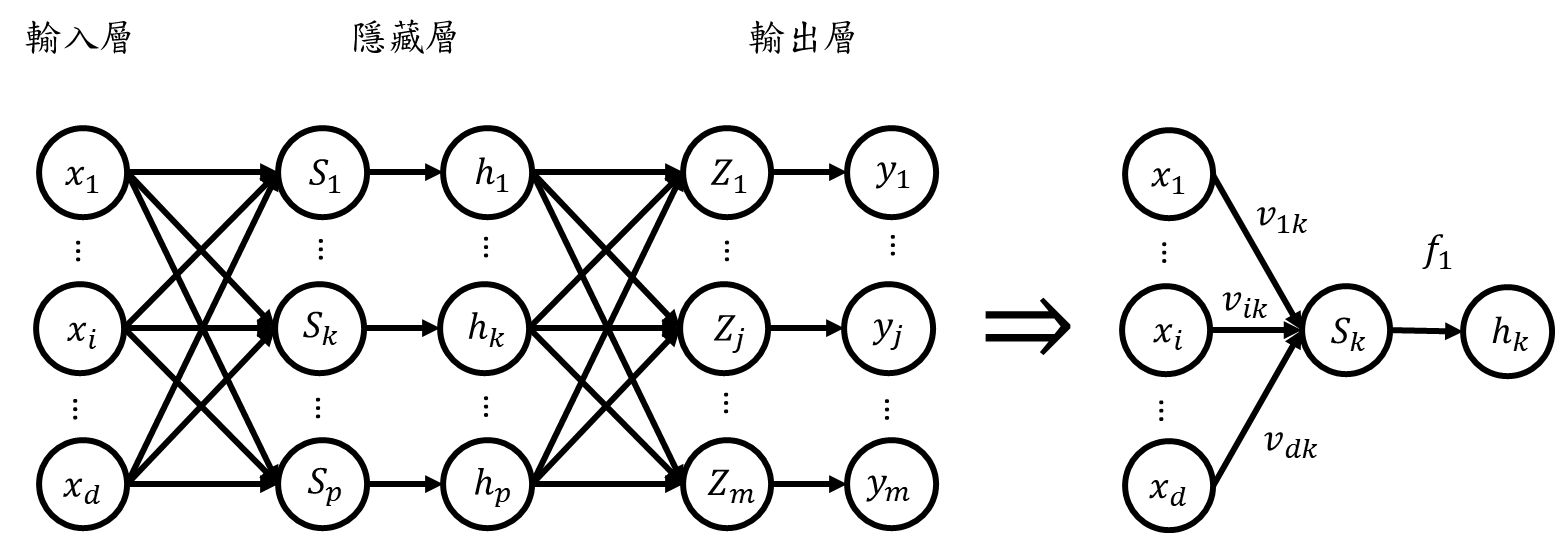
HW2

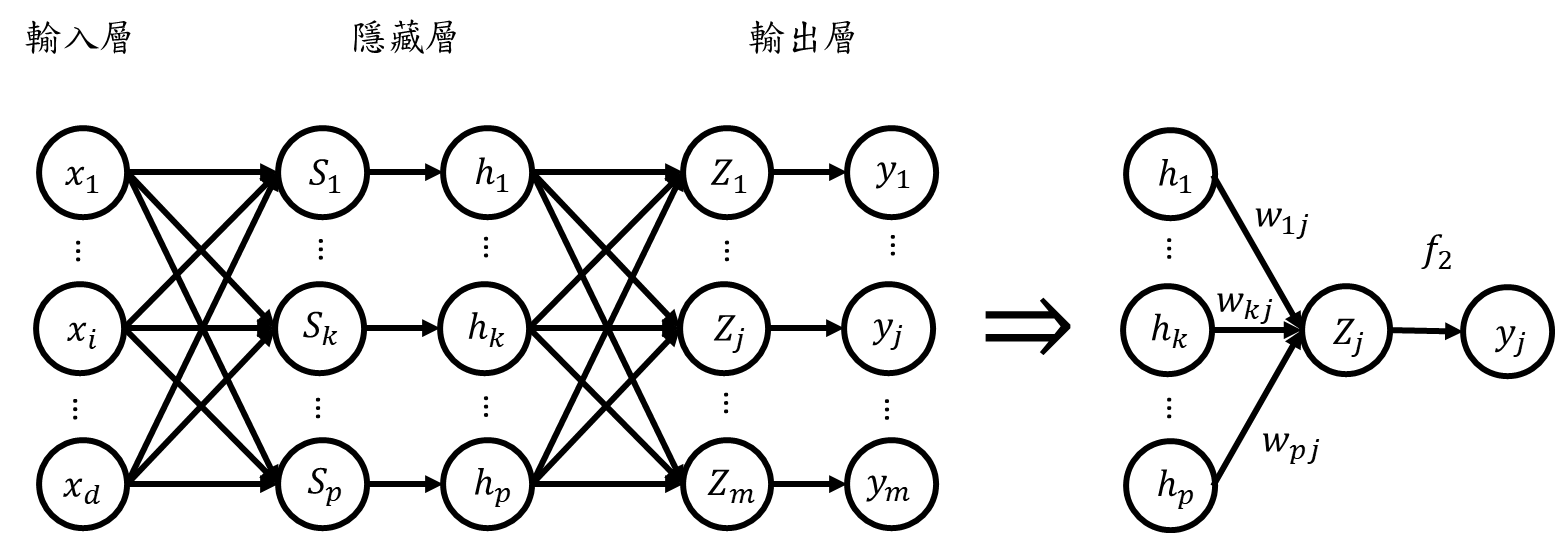
1. derivation of optimal W for a binary SVM
2. derivation of optimal W's for a 3-layer MLP
3. define a loss-function for SVM

**derivation of optimal W for a binary SVM**

1. **derivation of optimal W's for a 3-layer MLP**

前向傳遞(Forward propagation)：





1. **derivation of optimal W's for a 3-layer MLP**

運算法則

輸入X乘以權重W得到y，再通過啟動函數得到輸出（O）。 在這裡，啟動函數是sigmoid函數

E是loss函數值，這裡是輸出值（output）與真實值（target）的歐式距離

E的大小是評價感知器模型好壞的指標之一，w權重是描述這個感知器模型的參數，通過計算E來優化感知器模型，即優化w的值

表示第I層，第j個輸入連結第k個輸出的權值w。 以下先對一個權重（值）w求得感知器模型的梯度

現在把單個輸出的感知器模型推廣成多輸出感知器模型

**define a loss-function for SVM**

用來評估模型的預測值與真實值不一致的程度，也是神經網絡中優化的目標函數，神經網絡訓練或者優化的過程就是最小化損失函數的過程，損失函數越小，說明模型的預測值就越接近真實值，模型的健壯性也就越好。

Hinge Loss function：