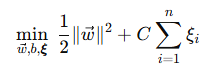
電子四甲 C111112155 謝宗佑 機器學習HW1 9/25

**1. SVM （支援向量機）的訓練過程**

* 我們要找到一條「分界線」（超平面）把 +1 / -1 的資料分開。
* 目標：讓這條線的「間隔 (margin)」最大。
* 數學上：



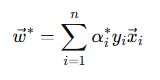
* w⃗：超平面的法向量（權重）
* b：偏差項（bias）
* ξi：每個樣本的**錯誤容許值**（slack variable）
* C：懲罰係數（控制間隔與錯誤之間的權衡）
* n : 樣本總數



這個問題沒有直接解，只能用「最佳化演算法」算出來。

解法有兩種思路：

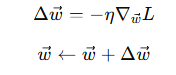
**1.對偶解法**：得到



### 說明：

* αi∗：對偶問題的最適解（由 QP 算法求出，不是手算）
* yi：第 i 筆資料的標籤（+1 或 -1）
* x⃗i：第 i筆資料的特徵向量
* n：訓練樣本的數量

**2.梯度下降法**：每次根據誤差更新：



### 其中：

* Δw⃗：權重的更新量（變化值）
* η：學習率（learning rate），控制更新步伐大小
* ∇w⃗L：損失函數 LLL 對權重 w⃗\vec{w}w 的梯度（偏導數）
* w⃗←w⃗+Δw⃗：用新的權重替代舊權重，完成一次更新

## 2. MLP（多層感知機）的訓練過程

* 一個 MLP 的結構是：

**輸入 → 加權相加 → 激活函數 → 下一層 → … → 輸出**

* **目標**：讓模型輸出接近真實標籤（分類問題通常用交叉熵損失函數 Cross-Entropy Loss）

### 訓練步驟：

1. **前向傳播 (Forward Propagation)**  
   將輸入資料依序通過每層的加權相加和激活函數，計算出模型的預測值。
2. **計算損失 (Loss Calculation)**  
   根據預測值與真實標籤，計算損失函數，衡量預測錯誤的大小。
3. **反向傳播 (Backpropagation)**  
   利用鏈式法則，從輸出層開始往回計算每個權重對損失的梯度。
4. **更新權重 (Weight Update)**  
   用梯度下降法更新權重：



* + η：學習率（learning rate）
  + ∇w⃗L：損失函數對權重的梯度

┌─────────────┐

│ SVM 訓練 │

└──────┬──────┘

│

(1) 定義 hinge loss：

\[

L = \frac{1}{2} \|w\|^2 + C \sum\_i \max(0, 1 - y\_i (w \cdot x\_i + b))

\]

│

(2) 計算梯度（次梯度 Subgradient）：

\[

\nabla\_w L = w - C \sum\_{i \in \text{錯誤樣本}} y\_i x\_i

\]

\*「錯誤樣本」指 \( 1 - y\_i (w \cdot x\_i + b) > 0 \) 的樣本

│

(3) 更新權重：

\[

\Delta w = - \eta \nabla\_w L

\]

\[

w \leftarrow w + \Delta w

\]

│

(4) 重複步驟 (2)(3) 直到收斂（梯度足夠小或達最大迭代次數）

### 幾個重點說明：

* **hinge loss** 是非光滑函數，所以梯度是用「次梯度(Subgradient)」概念。
* 梯度中，只有在「違反 margin 條件」的樣本（錯誤樣本）才會對梯度有貢獻。

┌─────────────┐

│ MLP 訓練 │

└──────┬──────┘

│

(1) 前向傳播 (Forward Propagation)：

輸入 \( x \) 經過權重 \( W\_1 \) → 激活函數 → 權重 \( W\_2 \) → softmax → 預測值 \( \hat{y} \)

│

(2) 計算損失 (Loss Calculation)：

使用交叉熵損失函數：

\[

L = -\sum\_i y\_i \log(\hat{y}\_i)

\]

│

(3) 反向傳播 (Backpropagation)：

計算各層權重的梯度：

\[

\nabla\_{W\_1} L, \quad \nabla\_{W\_2} L

\]

│

(4) 更新權重 (Weight Update)：

\[

\Delta W = - \eta \nabla\_W L

\]

\[

W \leftarrow W + \Delta W

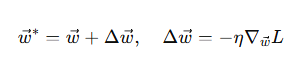
\]

│

(5) 重複步驟 (1)-(4) 直到收斂

## 重要重點整理

* **SVM**  
  透過 **hinge loss** 搭配 **梯度（次梯度）更新** 來訓練模型。
* **MLP**  
  透過 **前向傳播（forward）** 計算預測值，接著用 **反向傳播（backprop）** 計算梯度，再更新權重。
* **共同點**  
  兩者的權重更新公式相同：

****

也就是說，權重都透過「沿著損失函數梯度的反方向」做更新。