電子四甲 C111112155 謝宗佑 機器學習HW1 9/25

**1. SVM 的訓練過程（簡單版）**

* 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：

min⁡w,b  12∥w∥2+C⋅誤差\min\_{w,b} \; \frac{1}{2}\|w\|^2 + C \cdot \text{誤差}w,bmin​21​∥w∥2+C⋅誤差

* 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

w∗=∑iαi∗yixiw^\* = \sum\_i \alpha\_i^\* y\_i x\_iw∗=i∑​αi∗​yi​xi​

（這裡的 αi\alpha\_iαi​ 由 QP 算法求出來，非手動算）。

* 1. **梯度下降法**：每次根據誤差更新：

Δw=−η∇wL,w←w+Δw\Delta w = -\eta \nabla\_w L,\quad w \leftarrow w + \Delta wΔw=−η∇w​L,w←w+Δw

**2. MLP（多層感知機）的訓練過程（簡單版）**

* 一個 MLP 就是：輸入 →（加權相加）→（激活函數）→ 下一層 → … → 輸出。
* 目標：讓輸出接近真實標籤（分類通常用 cross-entropy loss）。
* 做法：
  1. **前向傳播 (forward)**：算出預測值。
  2. **計算損失 (loss)**：衡量和正確答案的差距。
  3. **反向傳播 (backprop)**：算出每個權重對 loss 的梯度。
  4. **更新權重**：

Δw=−η∇wL,w←w+Δw\Delta w = -\eta \nabla\_w L,\quad w \leftarrow w + \Delta wΔw=−η∇w​L,w←w+Δw

┌─────────────┐

│ SVM 訓練 │

└──────┬──────┘

│

(1) 定義 hinge loss

L = ½‖w‖² + C Σ max(0,1 - y(wx+b))

│

(2) 計算梯度 (次梯度)

grad\_w = w - C Σ yx (誤差樣本)

│

(3) 更新權重

Δw = -η grad\_w

w ← w + Δw

│

(4) 重複直到收斂

┌─────────────┐

│ MLP 訓練 │

└──────┬──────┘

│

(1) 前向傳播

x → W1 → 激活 → W2 → softmax → ŷ

│

(2) 計算損失 (cross-entropy)

L = -Σ y log(ŷ)

│

(3) 反向傳播

求每層梯度 ∇W1 L, ∇W2 L

│

(4) 更新權重

Δw = -η grad\_w

w ← w + Δw

│

(5) 重複直到收斂

**重點**

* **SVM** → 透過 hinge loss + 梯度更新
* **MLP** → 透過 forward + backprop 算梯度再更新
* **共同點**：最後更新規則一樣

w∗=w+Δw,Δw=−η∇wLw^\* = w + \Delta w, \quad \Delta w = -\eta \nabla\_w Lw∗=w+Δw,Δw=−η∇w​L