

Orange SVM 模型介面完整解說

你的畫面包含三大部分：

1. SVM Type (SVM 類型)
 2. Kernel (核函數)
 3. Optimization Parameters (最佳化參數)
-

① SVM Type：SVM 分類模式的選擇

SVM (C-SVM)

這是最常用的 SVM 模式，用在分類 (classification)。

Cost (C)：懲罰錯誤的強度

- C 越大：容忍錯誤越少 → 想把訓練集分得非常完美 → 容易過擬合
- C 越小：允許更多錯誤 → 模型比較平滑 → 容易泛化能力較好

👉 教學重點：

C 控制模型是「保守」還是「積極」。

ϵ -SVR (SVM 回歸)

你用的是分類問題 (生還 / 死亡)，因此不會用到這個。

ν -SVM (替代版 SVM)

使用 ν (nu) 參數控制錯誤率與支持向量數量，一般課堂上比較少使用。

Kernel (核函數)：決定 SVM 如何分界

這是 SVM 的核心！

● RBF (Radial Basis Function) — 預設、最常用、最實用

RBF 是 SVM 中最常見的核，它可以畫出彎曲的、非線性的分界線。

非常適合 Titanic 這種：

- 特徵之間有複雜關係
- 生還率不是線性分割

γ (gamma) 參數的意義

Gamma 決定「模型看事情的範圍」。

- γ 大：只看非常近的小區域 → 容易過擬合
- γ 小：看得很廣 → 分界很平滑
- auto : Orange 自動計算 (通常以 $1/n_features$)

👉 教學重點：

RBF = 常用、穩定、適合一般課堂案例。

○ Linear (線性)

畫出「直線」或「平面的」分界線。

Titanic 通常不是線性問題，效果比 RBF 差。

適合：

- 高維度 (特徵很多)

- 接近線性可分的資料集
-

○ Polynomial (多項式核)

使用多項式方式去彎曲空間。

參數多 (degree、coef0...)，調整困難。

不適合課堂初學者。

○ Sigmoid (類似神經網路的激活函數)

很少使用。

效果不一定好，常不穩定。

③ Optimization Parameters (最佳化參數)

✓ Numerical tolerance (數值容忍度)

決定演算法要多精準才算「找到最佳解」。

- 越小：要求越精準 → 訓練時間比較久
 - 0.001 已經很夠用 → Orange 的預設值
-

✓ Iteration limit (迭代次數上限)

最多允許演算法跑幾次。

- 預設 100
- Titanic 這種 dataset 不會太大，100 足夠
- 若 SVM 無法收斂，可改成 500 或 1000

4 底部的 Apply Automatically

啟用後模型會自動重新訓練
(對教學來說很方便)。

■ 總結：你這張圖中的設定代表什麼？

你現在的設定是：

- 分類用的 C-SVM
- 核函數選擇 RBF (最常用、適合 Titanic)
- $C = 1$ (中等偏保守、不容易過擬合)
- $\gamma = \text{auto}$ (交給 Orange 設定)
- $\text{Iteration} = 100$ (足夠)

➔ 這是一個標準、適合初學者、適合課堂示範的 SVM 設定

➔ 會得到穩定、可理解、可重現的結果
