

機器學習

Machine Learning

2023 (C) Elton Huang

非機器學習

2023 (C) Elton Huang



機器學習

2023 (C) Elton Huang

訓練：
training/fitting

訓練集

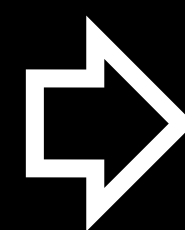


模型

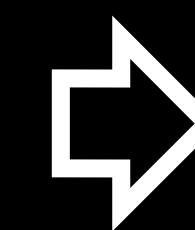


推論/預測：
inference/
prediction

輸入



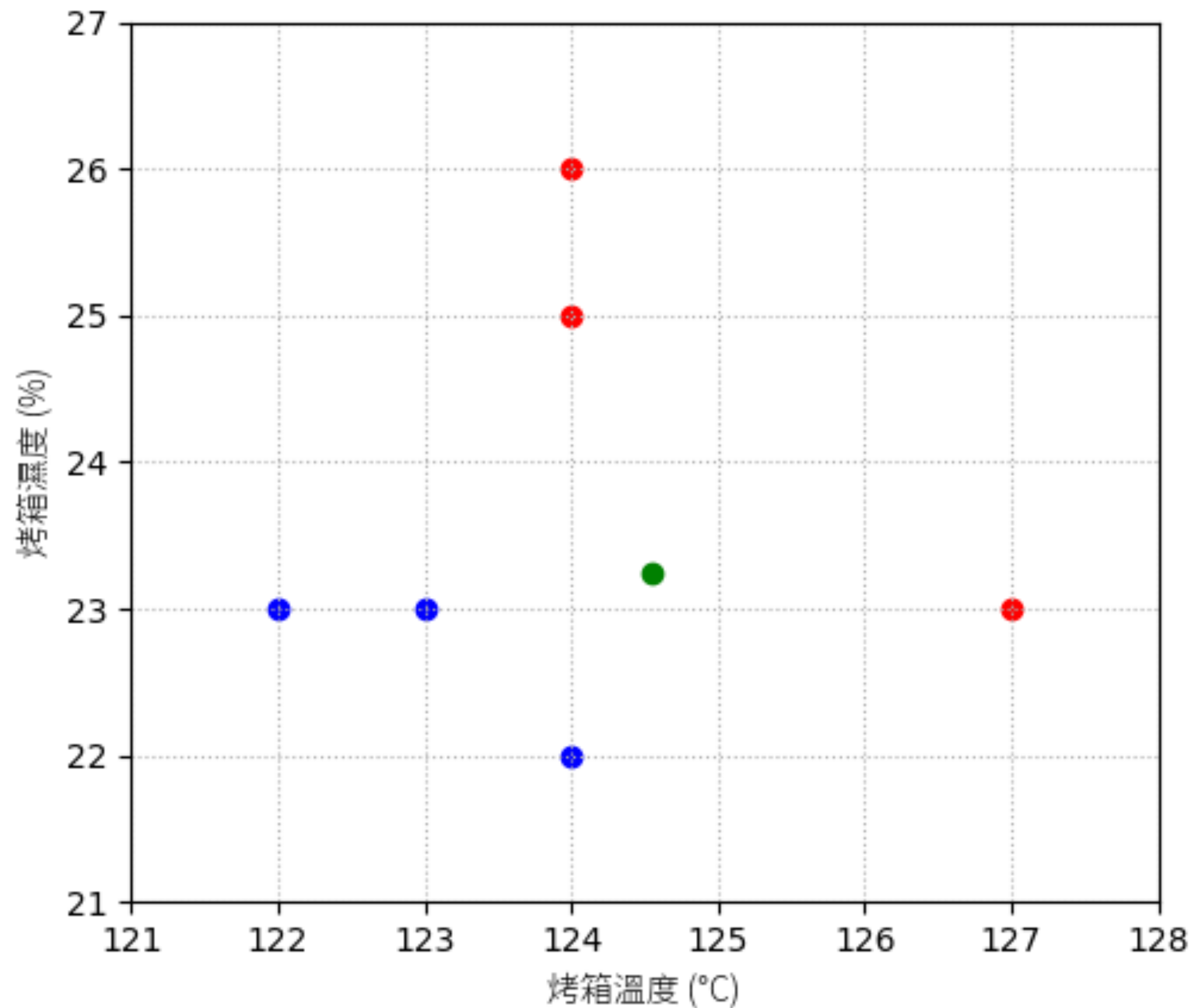
模型



輸出

Pizza 烘烤紀錄

	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	2	127	23	●
	3	124	25	●
	4	122	23	●
	5	124	26	●
	6	124	22	●
測試集	t	124.55	23.25	?

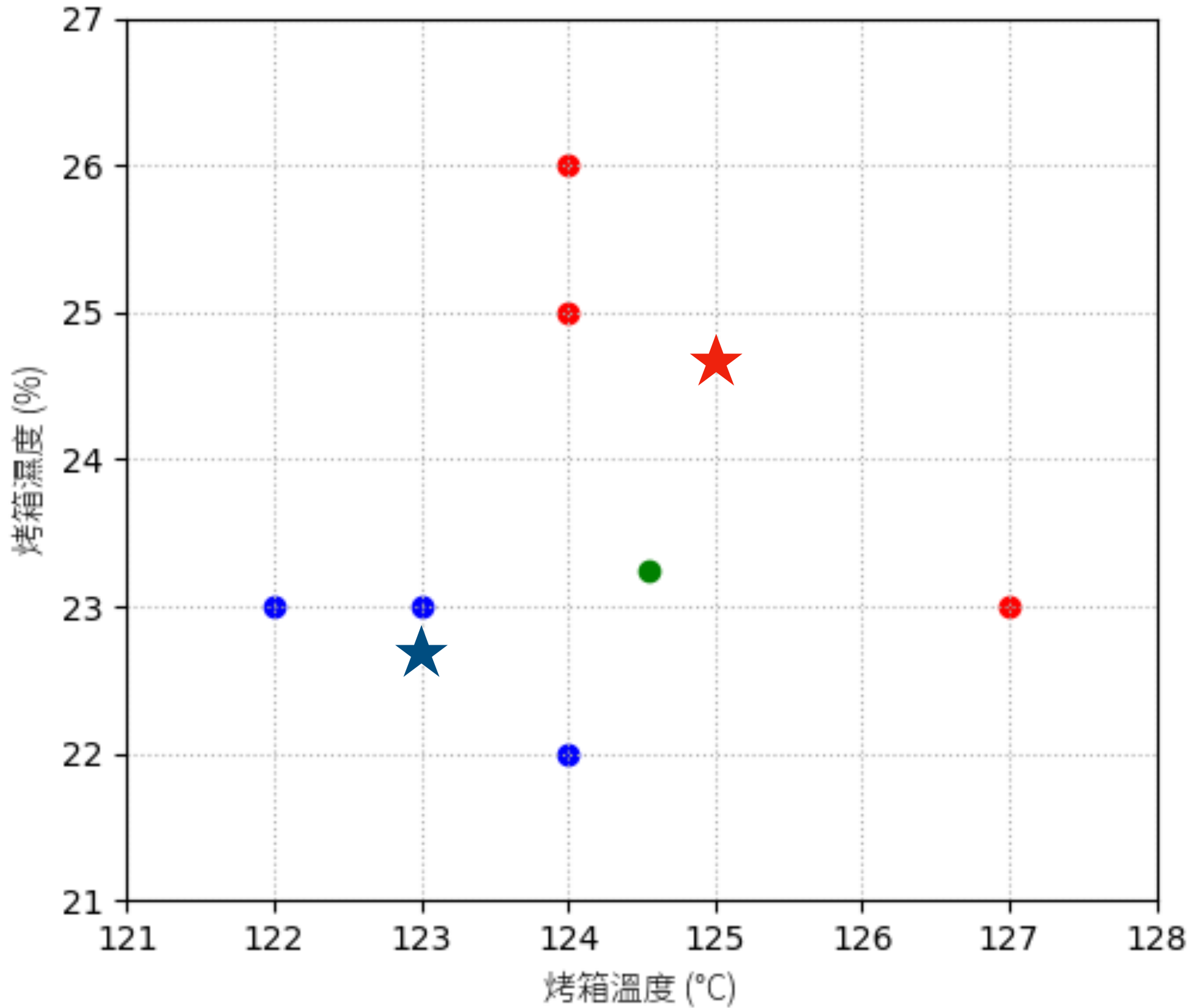


Pizza 烘烤紀錄				
	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	2	127	23	●
	3	124	25	●
	4	122	23	●
	5	124	26	●
	6	124	22	●
測試集	t	124.55	23.25	?

- 最短距離分類器
- kNN (k-最近鄰) 分類器
- 支持向量機
- 決策樹

- 最短距離分類器
- kNN (k-最近鄰) 分類器
- 支持向量機
- 決策樹

最短距離分類器



	編號	和★距離 ²	和★距離 ²	推論分類
測試集	t	2.7	2.2	●

	編號	溫度 °C	濕度 %	評價
訓練集	2	127	23	●
	3	124	25	●
	5	124	26	●
平均值		125.0	24.7	★

	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	4	122	23	●
	6	124	22	●
平均值		123.0	22.7	★

- 最短距離分類器
- kNN (k-最近鄰) 分類器
- 支持向量機
- 決策樹

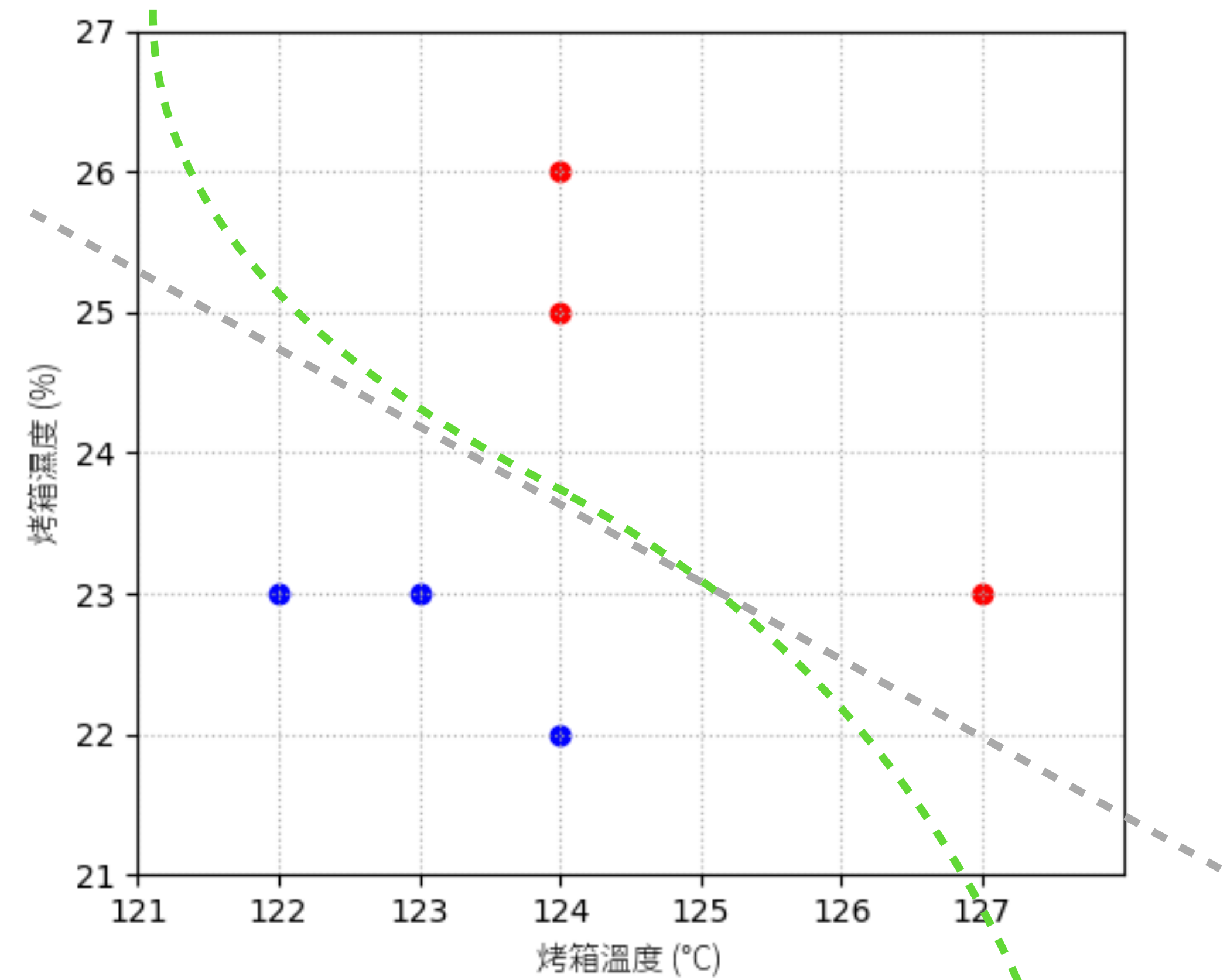
kNN (k-最近鄰) 分類器

	編號	溫度 °C	濕度 %	評價	和t距離²	順序	1NN 推論	3NN 推論	5NN 推論	7NN 推論
訓練集	1	123	23	●	2.5	2nd		✓	✓	✓
	2	127	23	●	6.1	4th			✓	✓
	3	124	25	●	3.4	3rd		✓	✓	✓
	4	122	23	●	6.6	5th			✓	✓
	5	124	26	●	7.9	6th				✓
	6	124	22	●	1.9	1st	✓	✓	✓	✓
	7	122	21	●	11.6	8th				
	8	126	26	●	9.7	7th				✓
測試集	t	124.55	23.25	?			●	●	●	●

- 最短距離分類器
- kNN (k-最近鄰) 分類器
- 支持向量機
- 決策樹

支持向量機

- 找一條線，
線的兩邊分屬不同的兩類
- 統計上的假設：
物以類聚、常態分佈

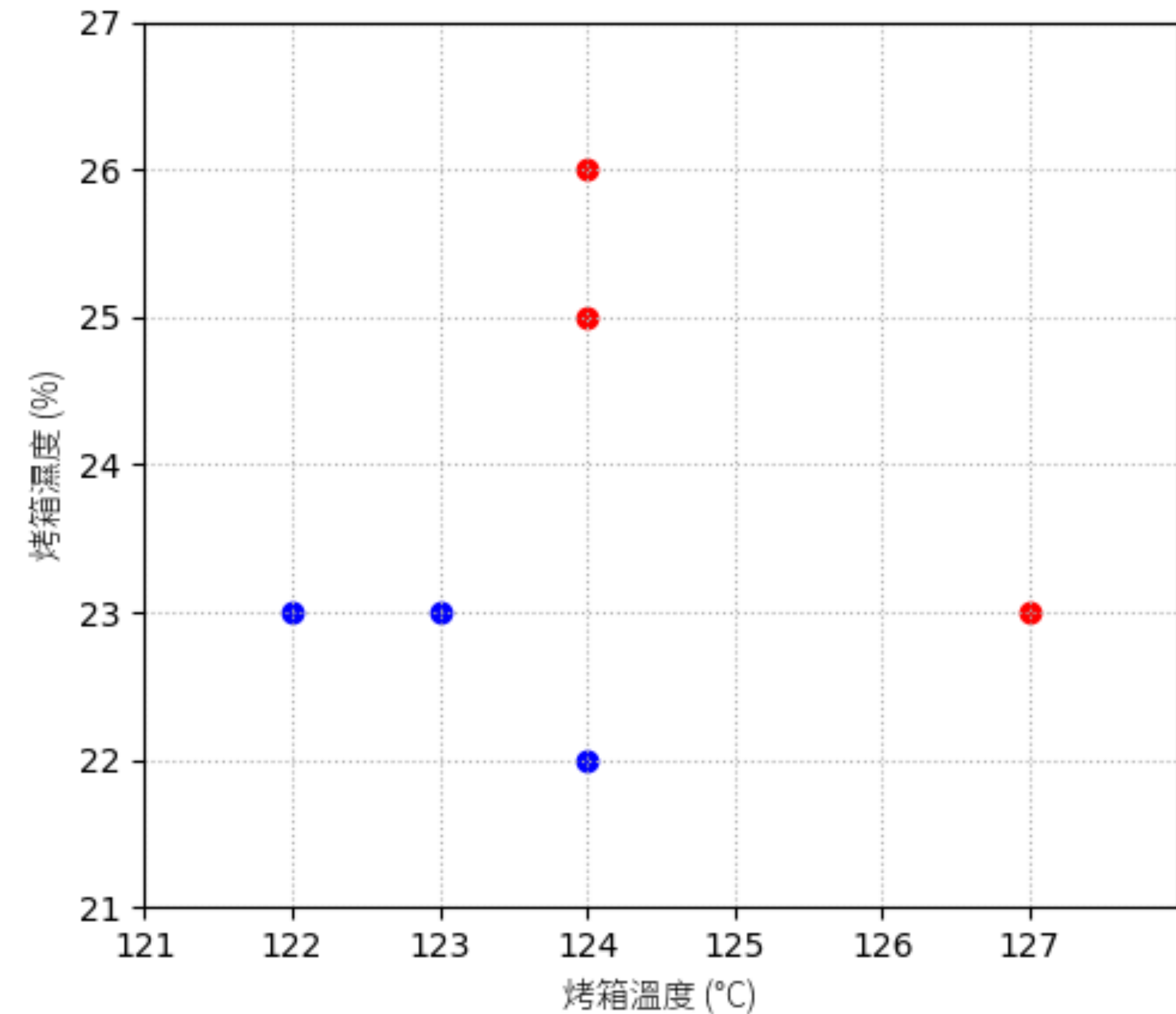


(線性) 支持向量機

(Linear) Support Vector Machine

一條能夠支持以下條件的 (直) 線

1. 能夠分隔兩集合
 2. 在眾多這樣的直線中找到
 - 2.1. 和兩集合最靠近的點等距
 - 2.2. 通過兩集合最近點的平行線所形成的緩衝區間帶 (margin) **最寬**
- 的唯一一直線

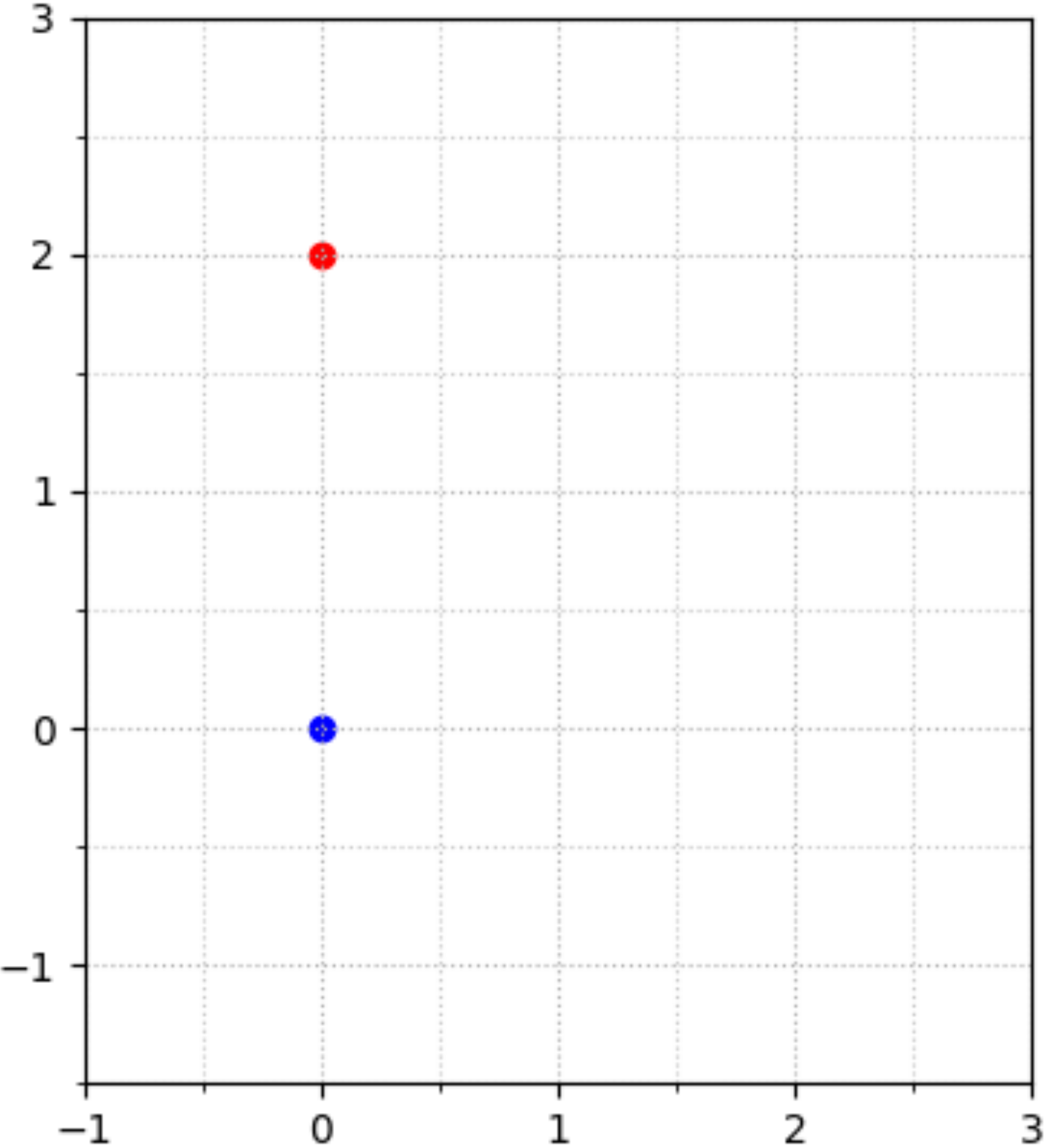


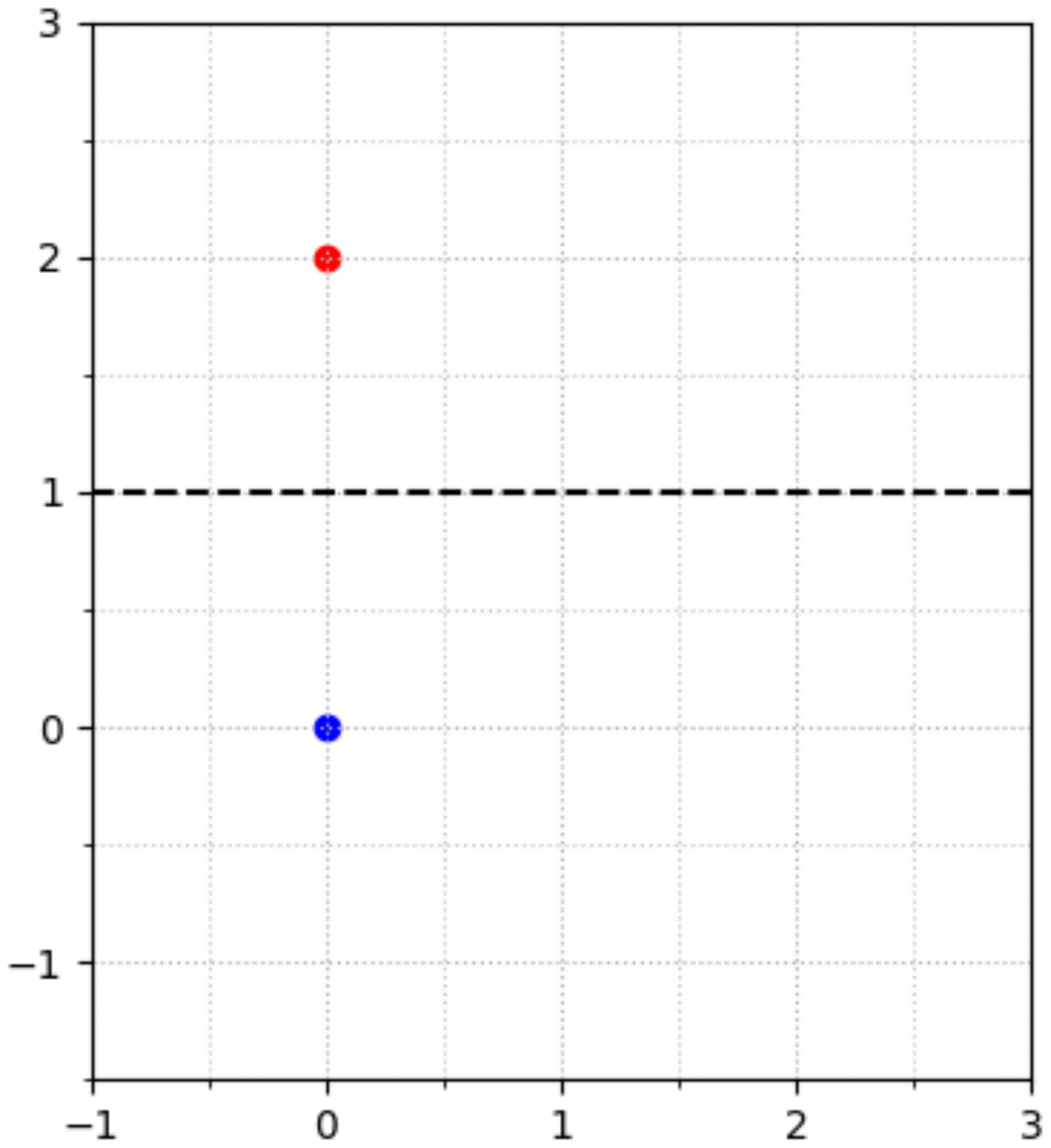
(線性) 支持向量機

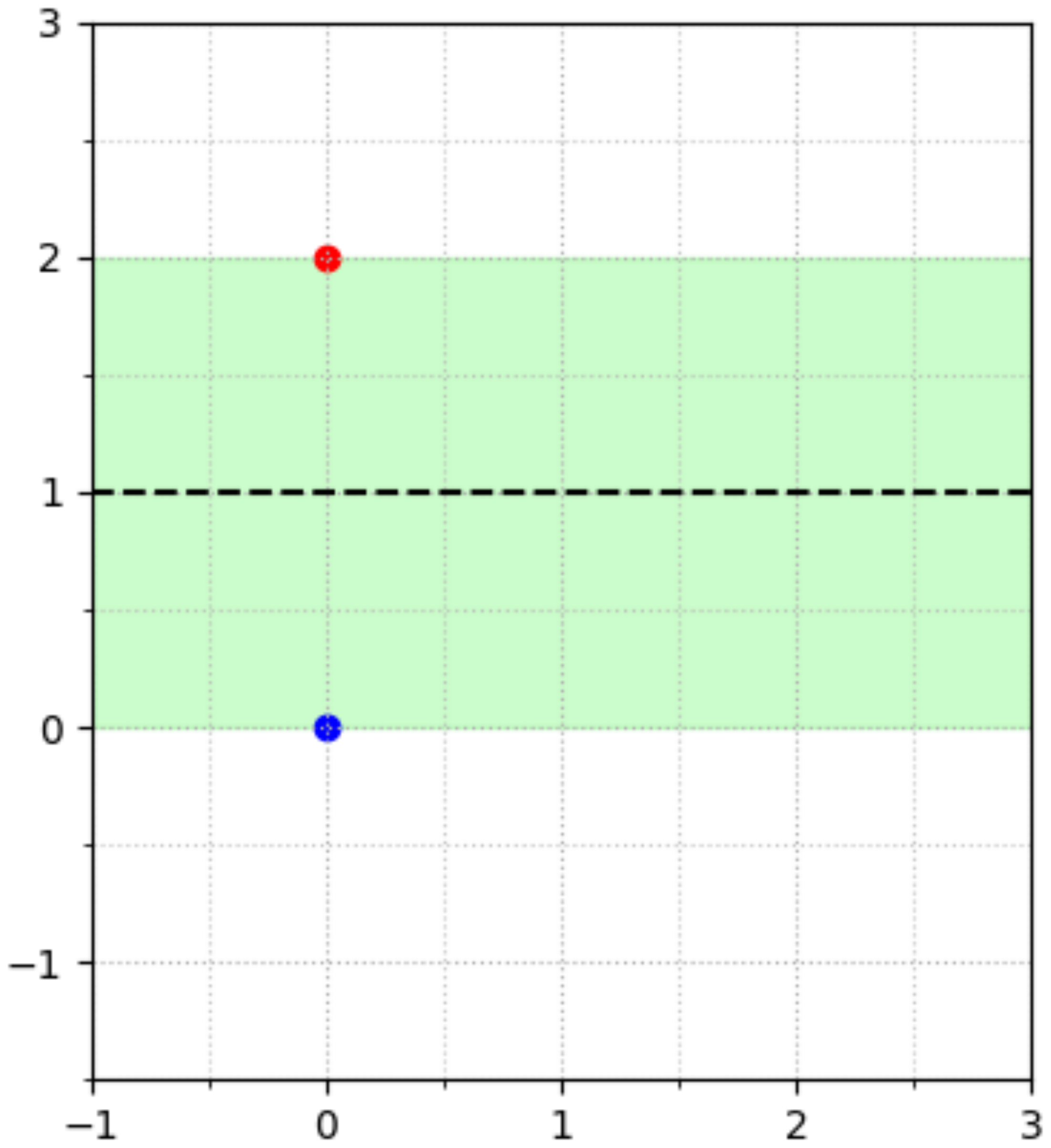
(Linear) Support Vector Machine

$$ax + by + c = 0$$

$$\underline{(a \quad b \quad c)} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$$

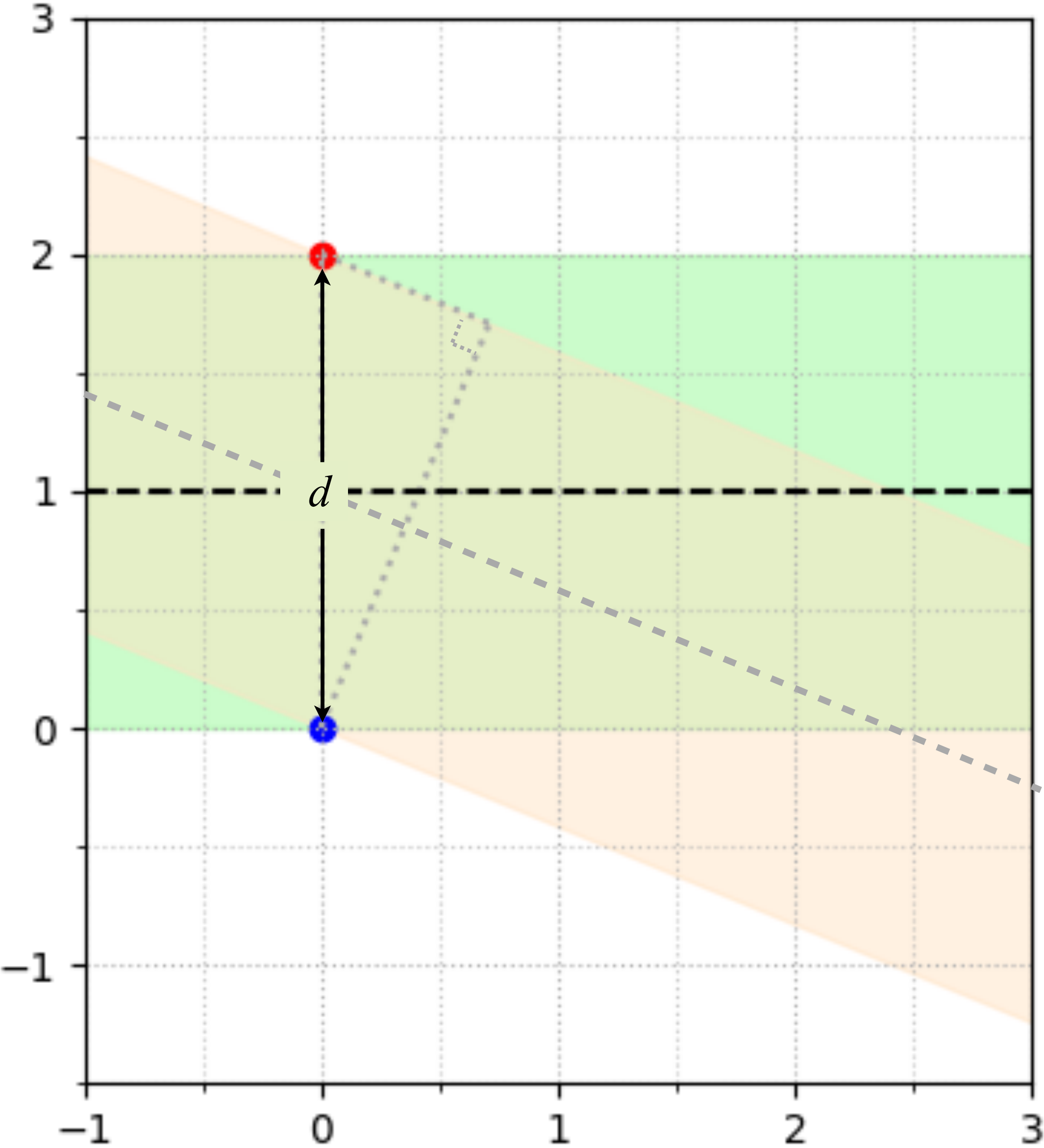






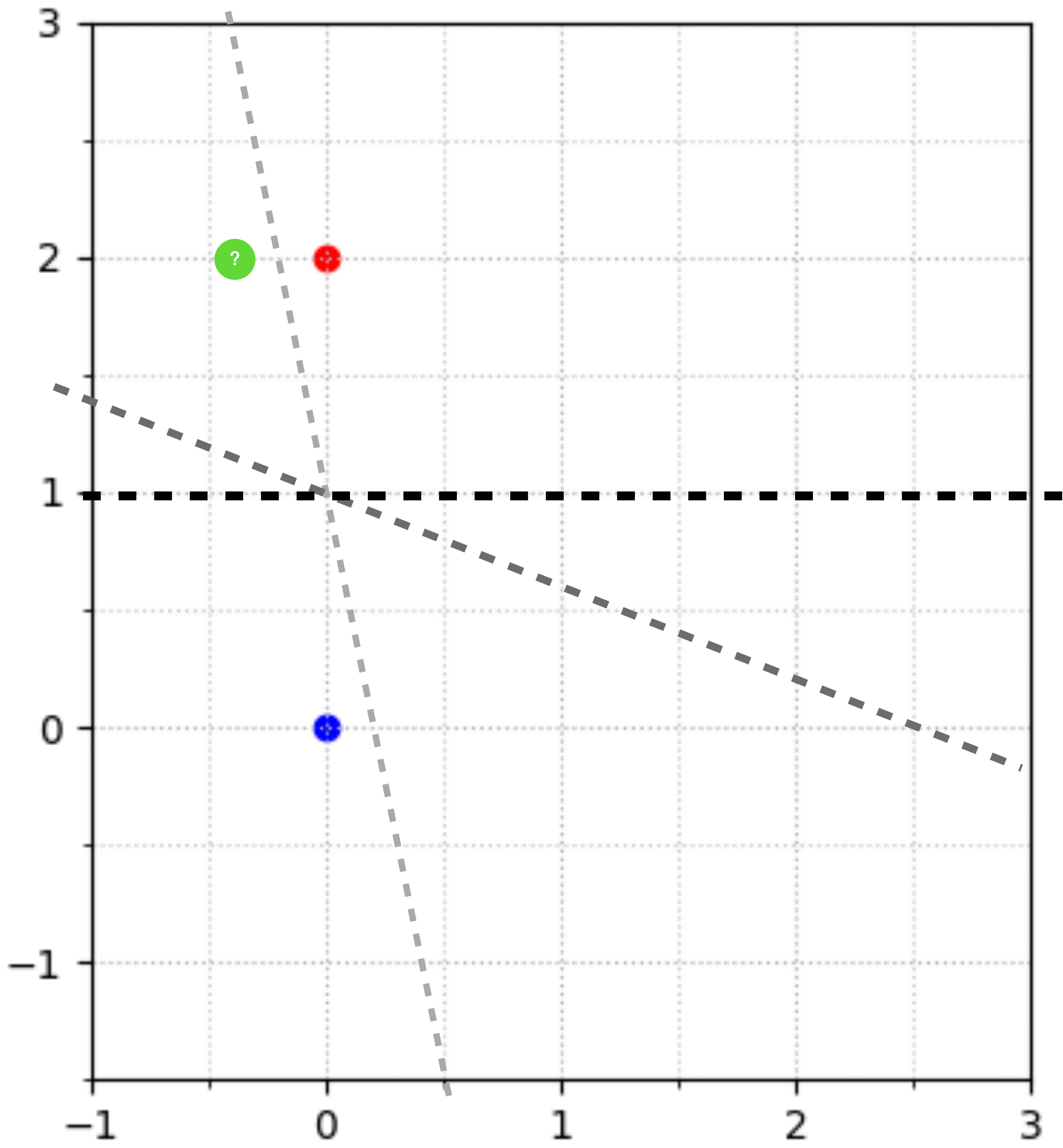
$ax + by + c = 0$

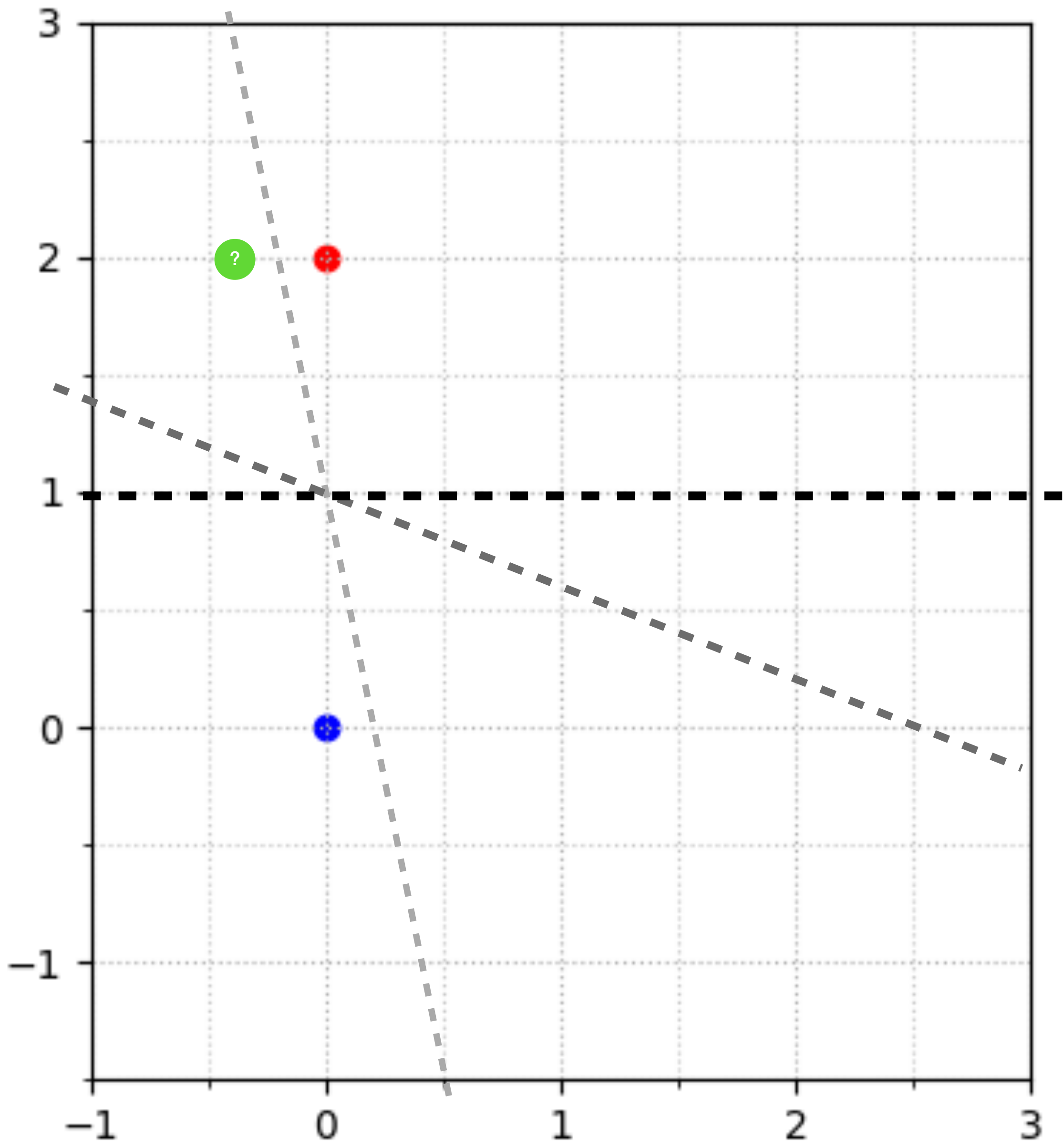
$$\text{帶寬} = \frac{d}{\sqrt{a^2 + b^2}} \cdot b$$



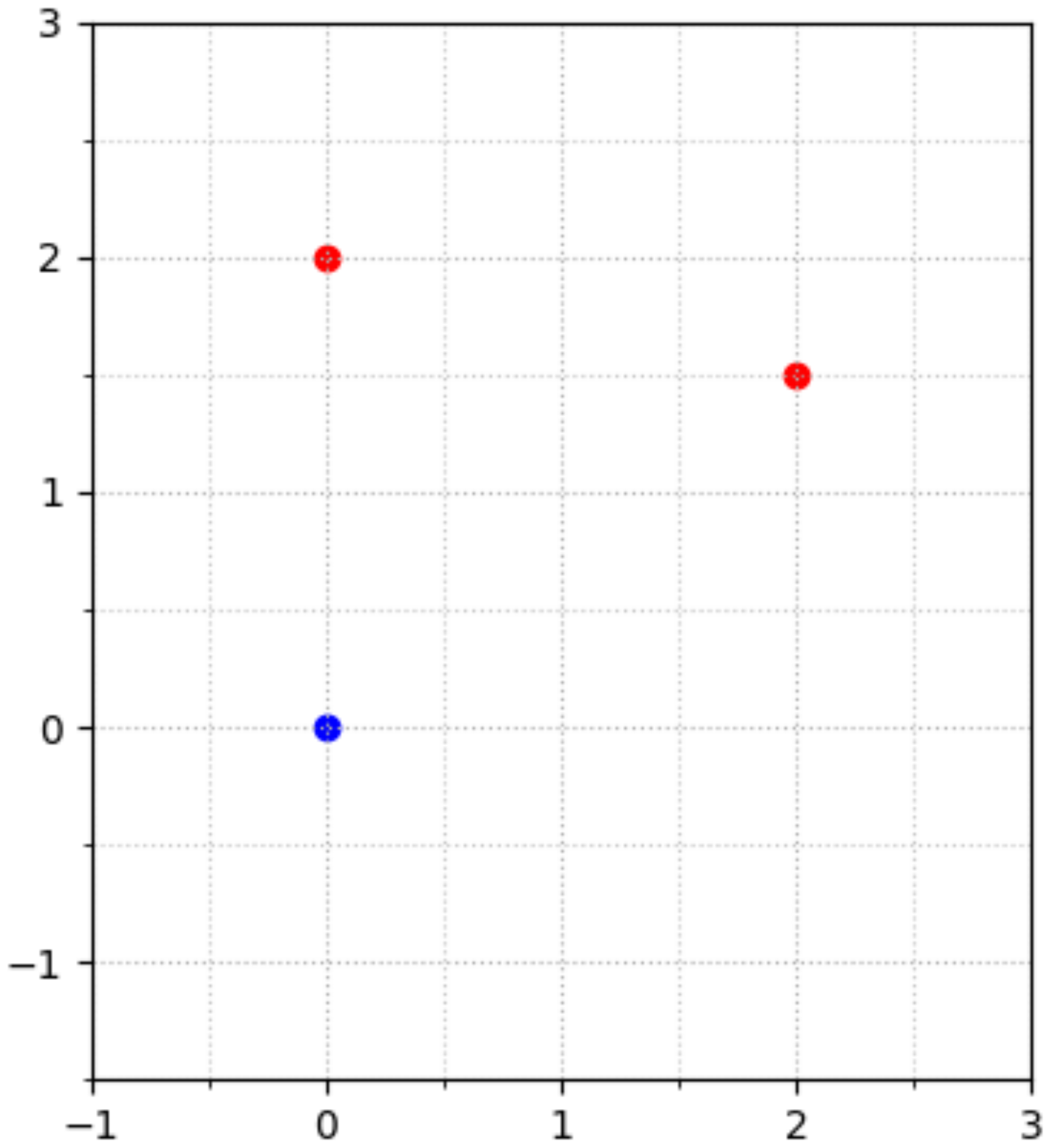
$$\text{帶寬} = \frac{4}{\sqrt{0^2 + 1^2}} \cdot 1 = 4$$

$$\text{帶寬} = \frac{4}{\sqrt{2^2 + 5^2}} \cdot 5 \approx 3.7$$



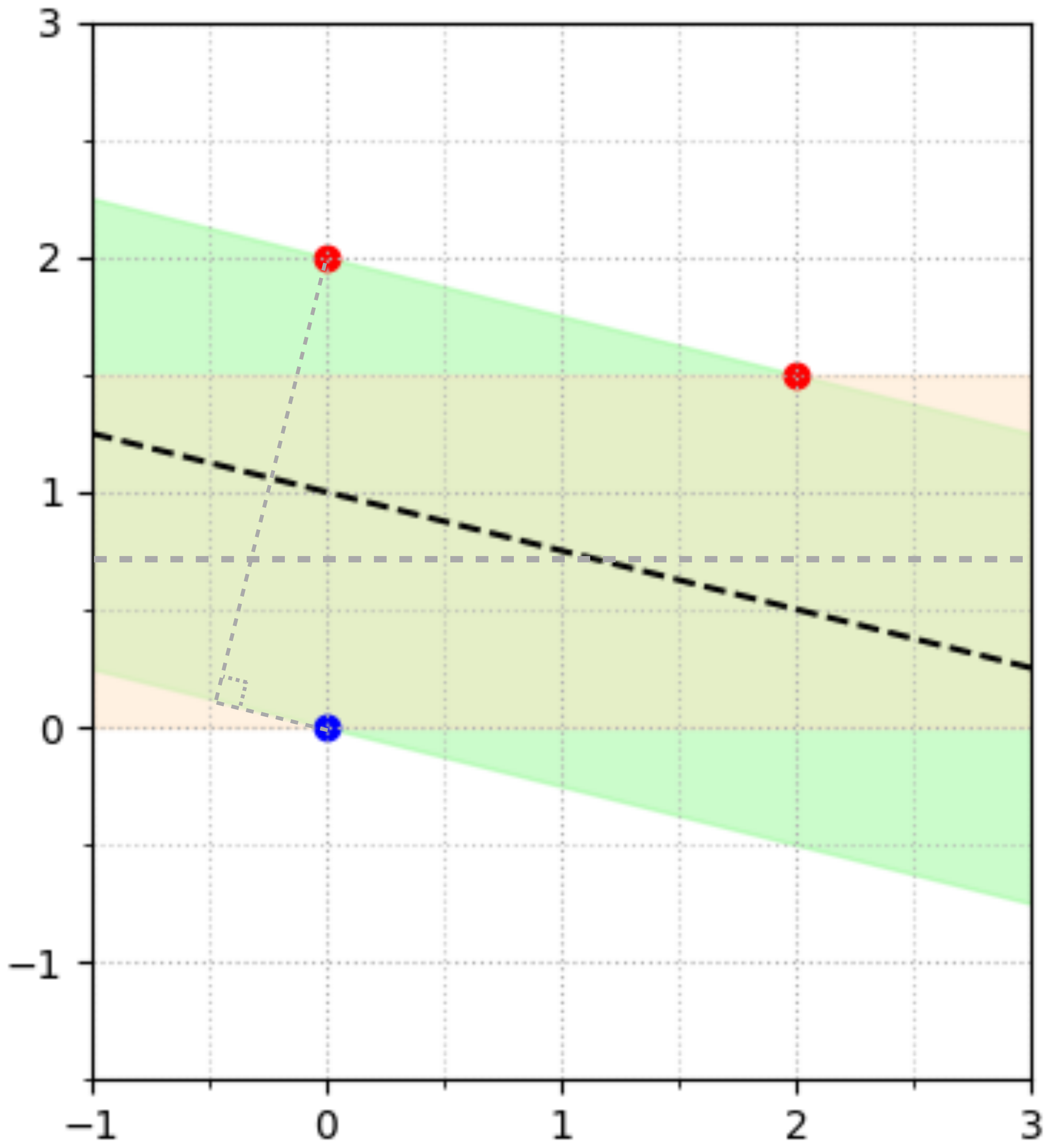


注意：和線性回歸的不同



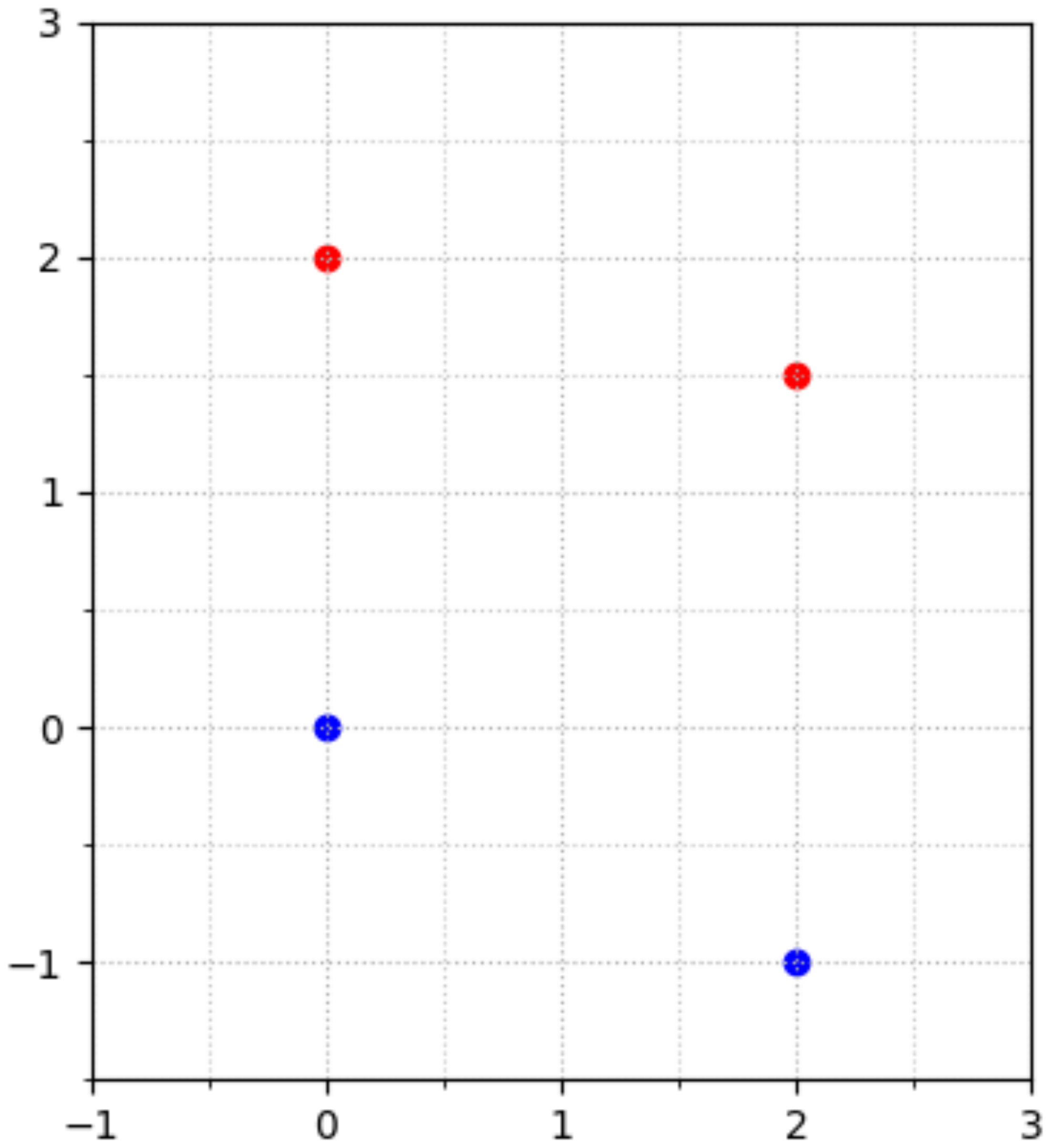
$ax + by + c = 0$

帶寬 = $\frac{d}{\sqrt{a^2 + b^2}} \cdot b$



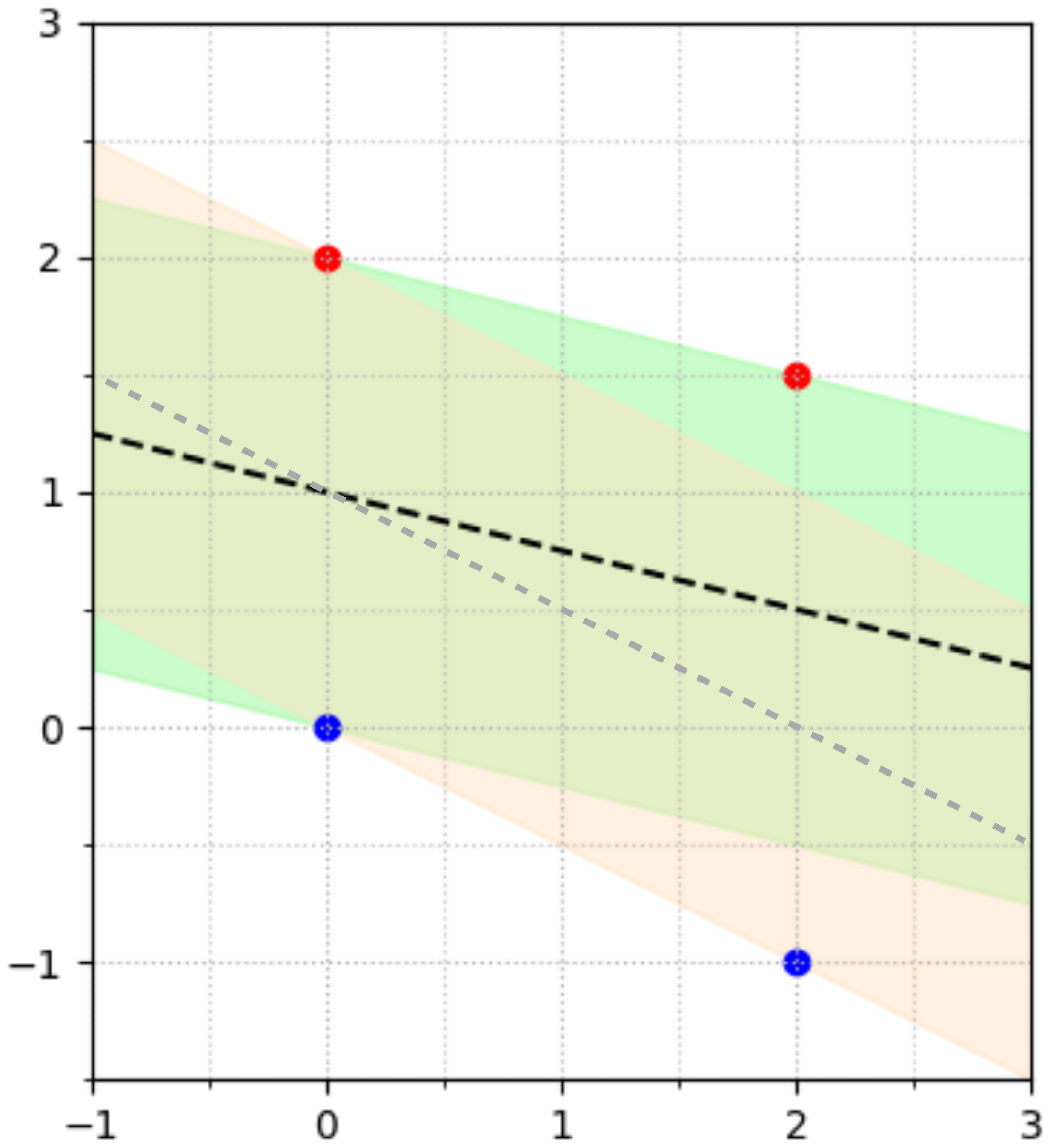
帶寬 = $\frac{3}{\sqrt{0^2 + 1^2}} \cdot 1 = 3$

帶寬 = $\frac{4}{\sqrt{1^2 + 4^2}} \cdot 4 \approx 3.88$



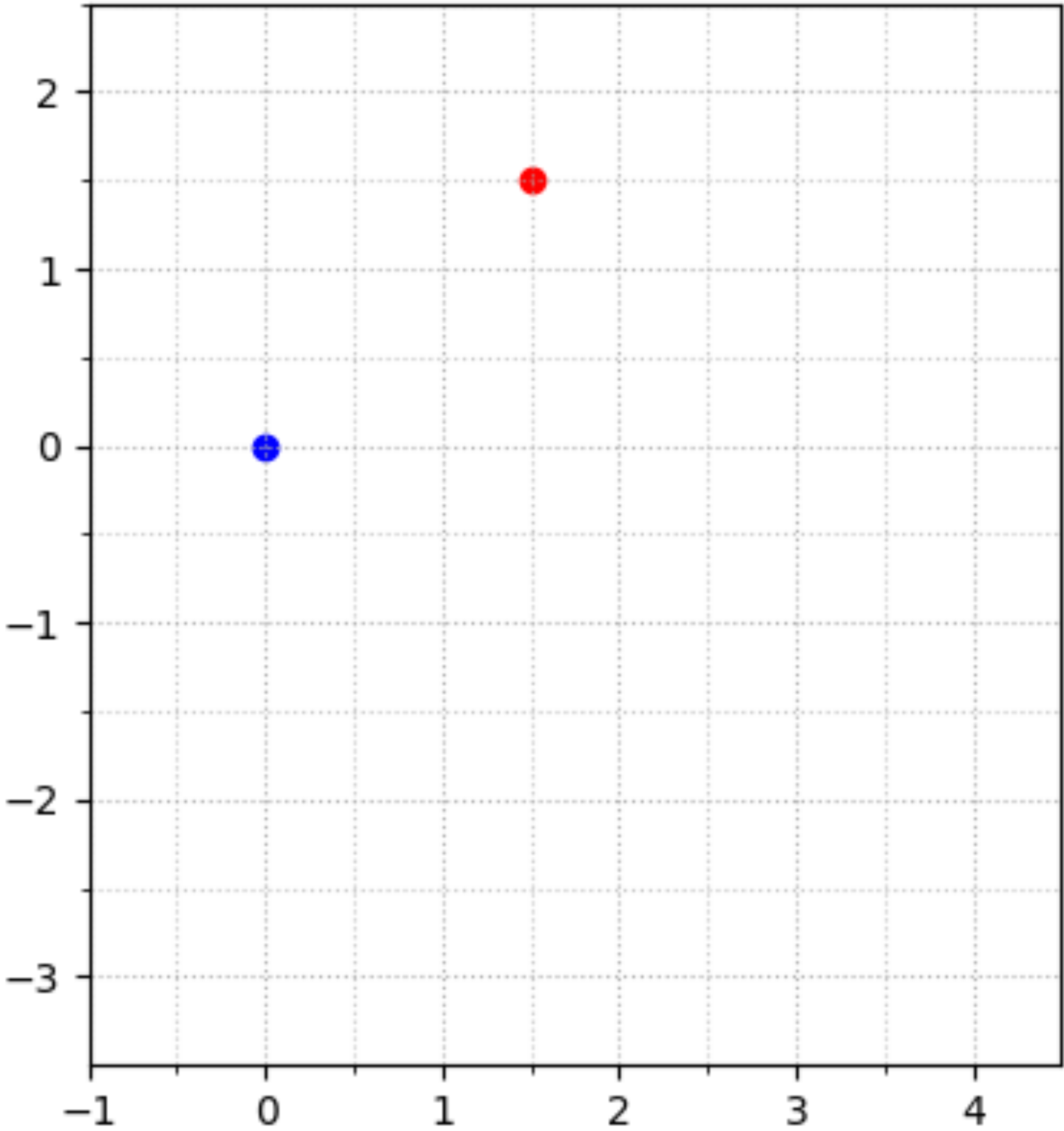
$ax + by + c = 0$

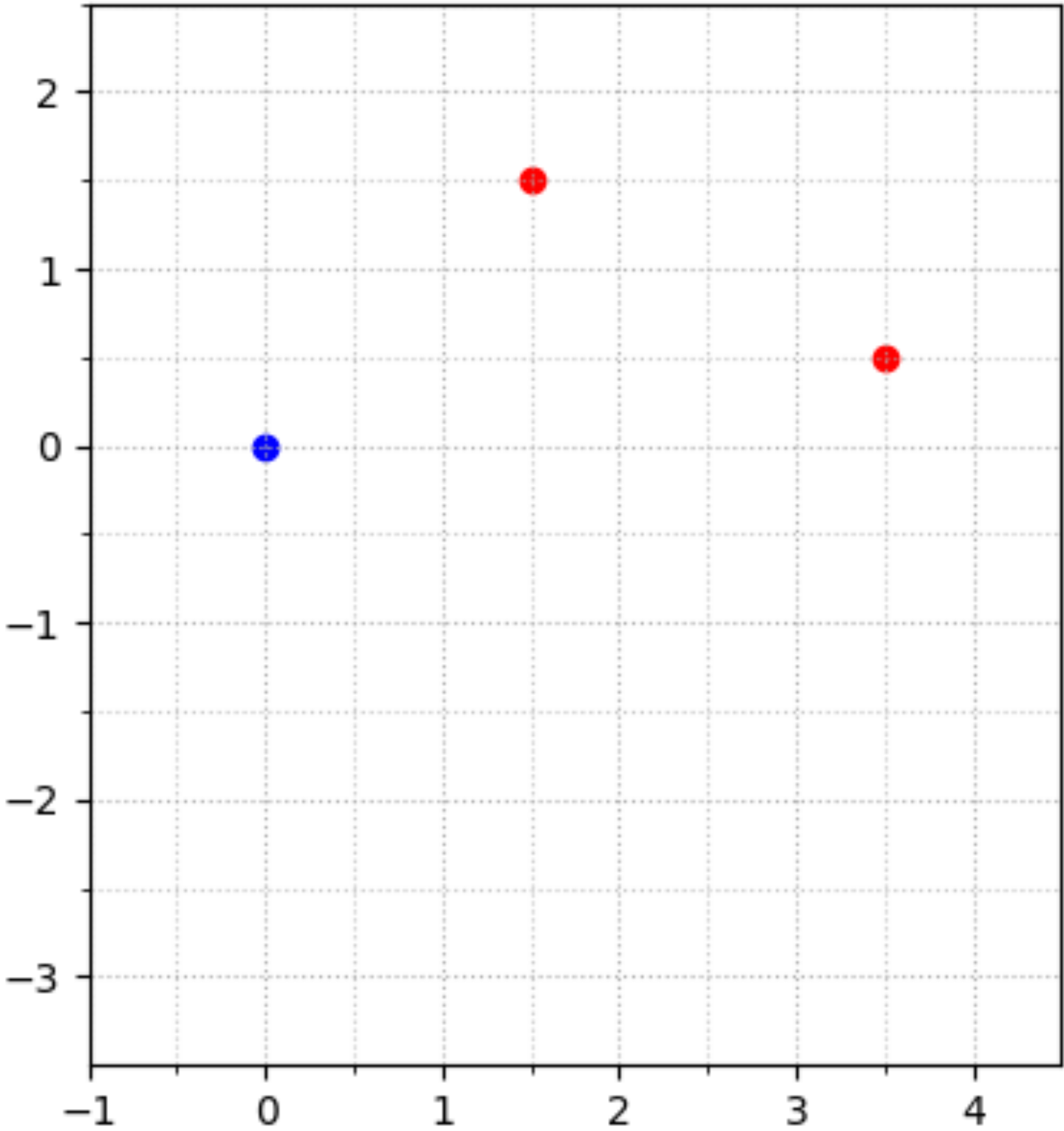
$$\text{帶寬} = \frac{d}{\sqrt{a^2 + b^2}} \cdot b$$

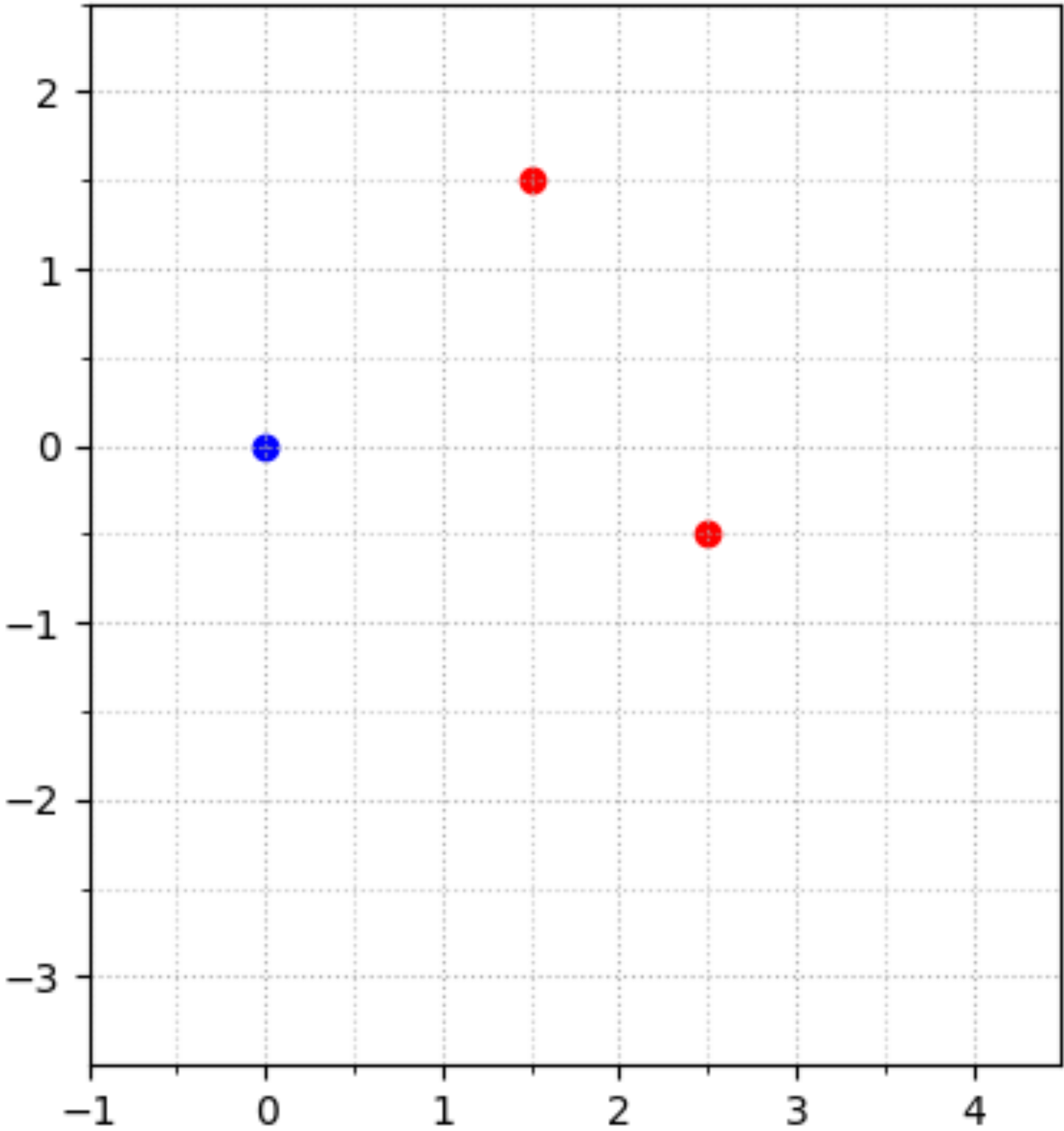


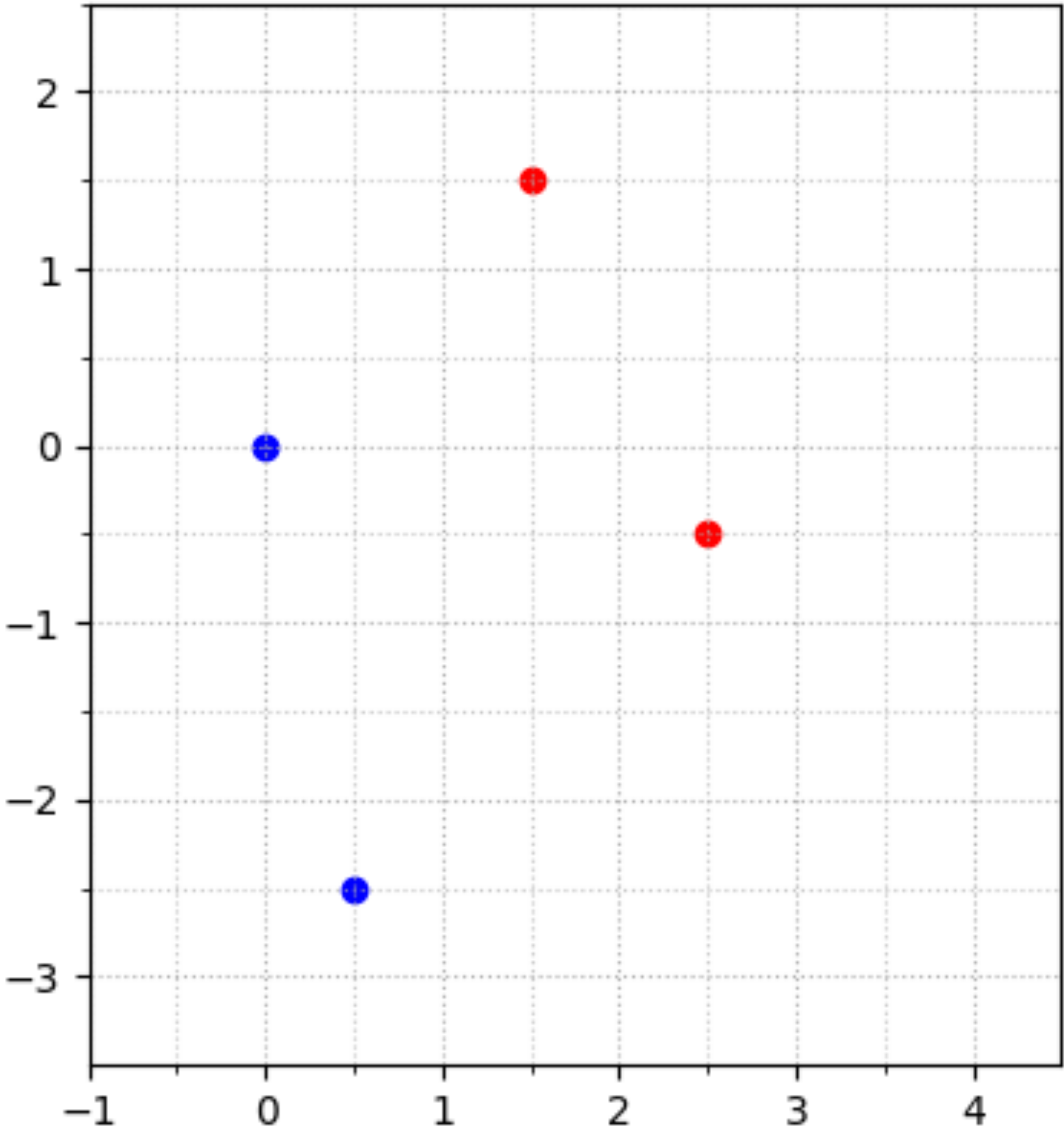
$$\text{帶寬} = \frac{4}{\sqrt{1^2 + 4^2}} \cdot 4 \approx 3.88$$

$$\text{帶寬} = \frac{4}{\sqrt{2^2 + 4^2}} \cdot 4 \approx 3.58$$









如何找出線性支持向量機？

請從以上的範例說明歸納出找出線性支持向量機的步驟

給定兩集合分別包含平面中兩點座標：

$$A = \{(x_1, y_1), (x_2, y_2)\}$$

$$B = \{(x_3, y_3), (x_4, y_4)\}$$

以下為找出劃分 A 、 B 兩集合之線性支持向量機的步驟：

1. ...

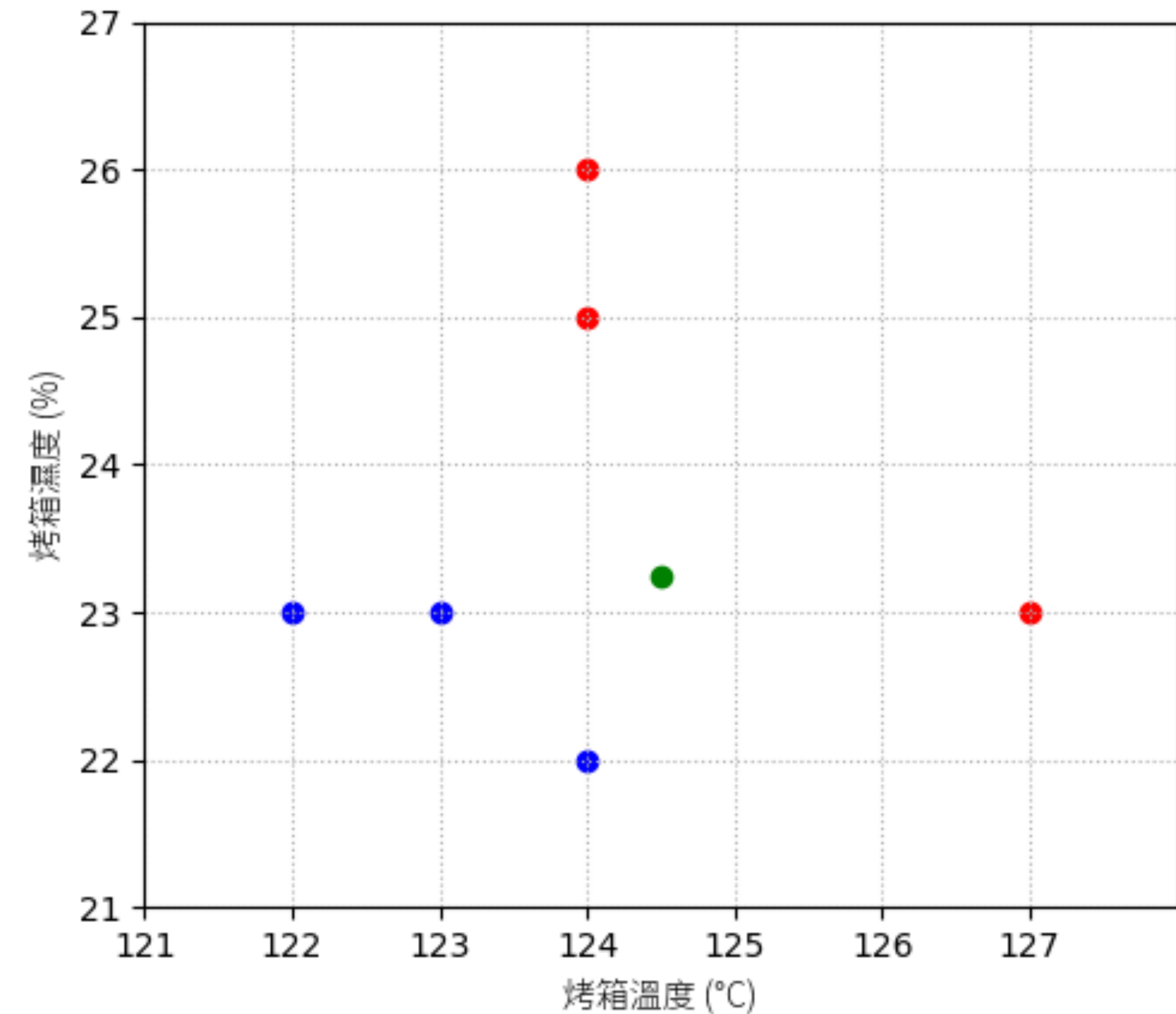
2. ...

(線性) 支持向量機

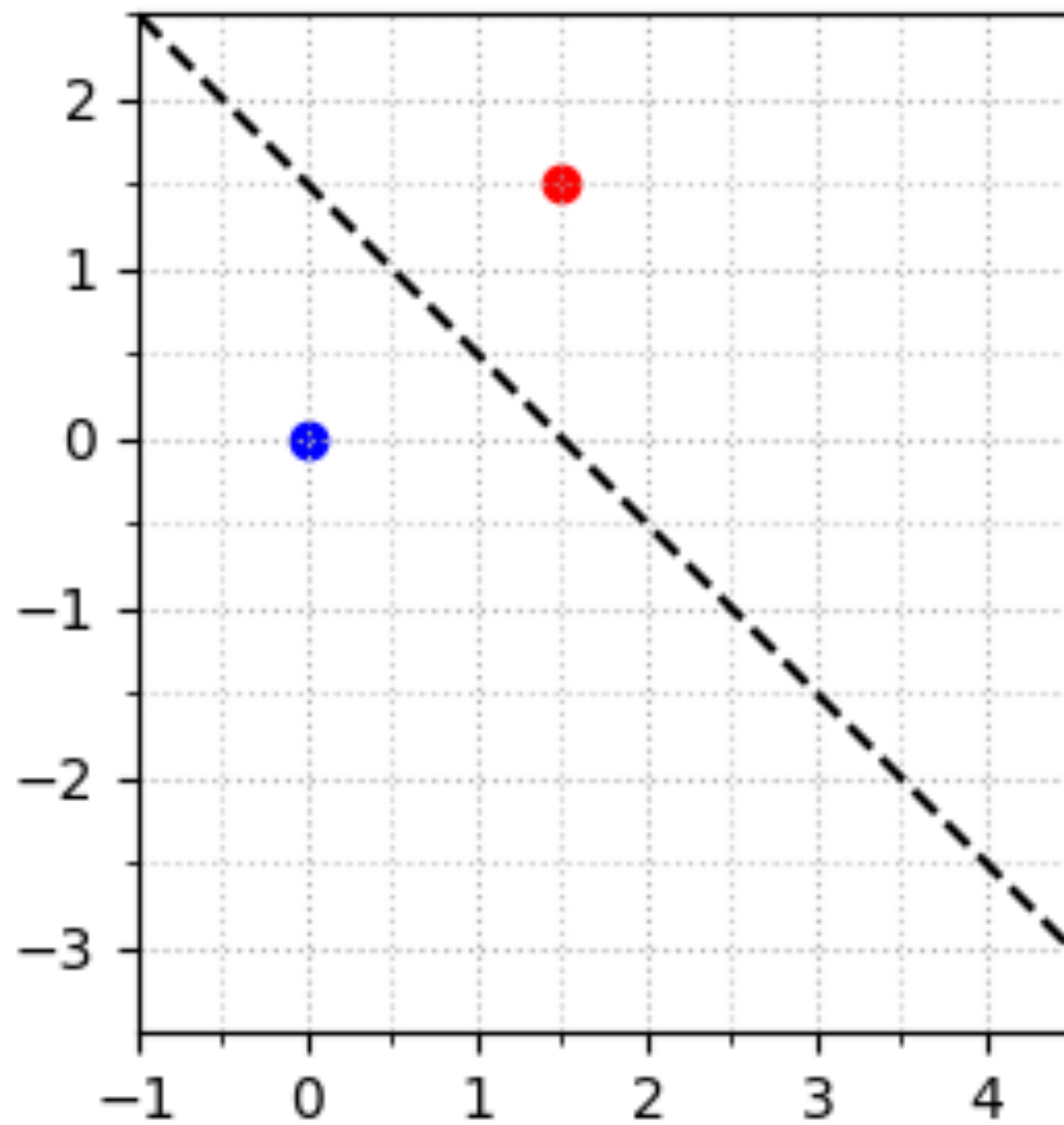
(Linear) Support Vector Machine

一條能夠支持以下條件的 (直) 線

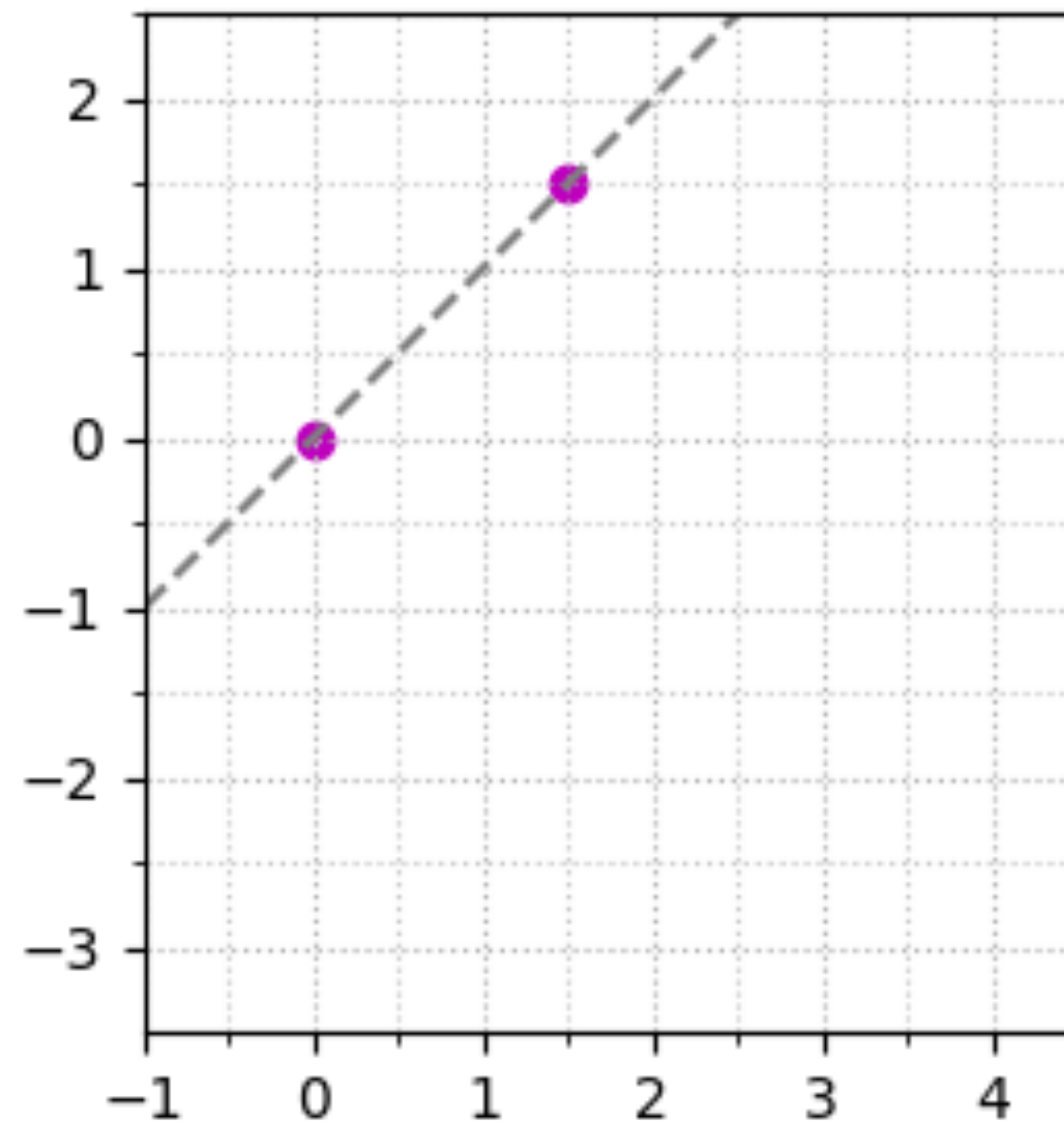
1. 能夠分隔兩集合
 2. 在眾多這樣的直線中找到
 - 2.1. 和兩集合最靠近的點等距
 - 2.2. 通過兩集合最近點的平行線所形成的緩衝區間帶 (margin) **最寬**
- 的唯一一直線



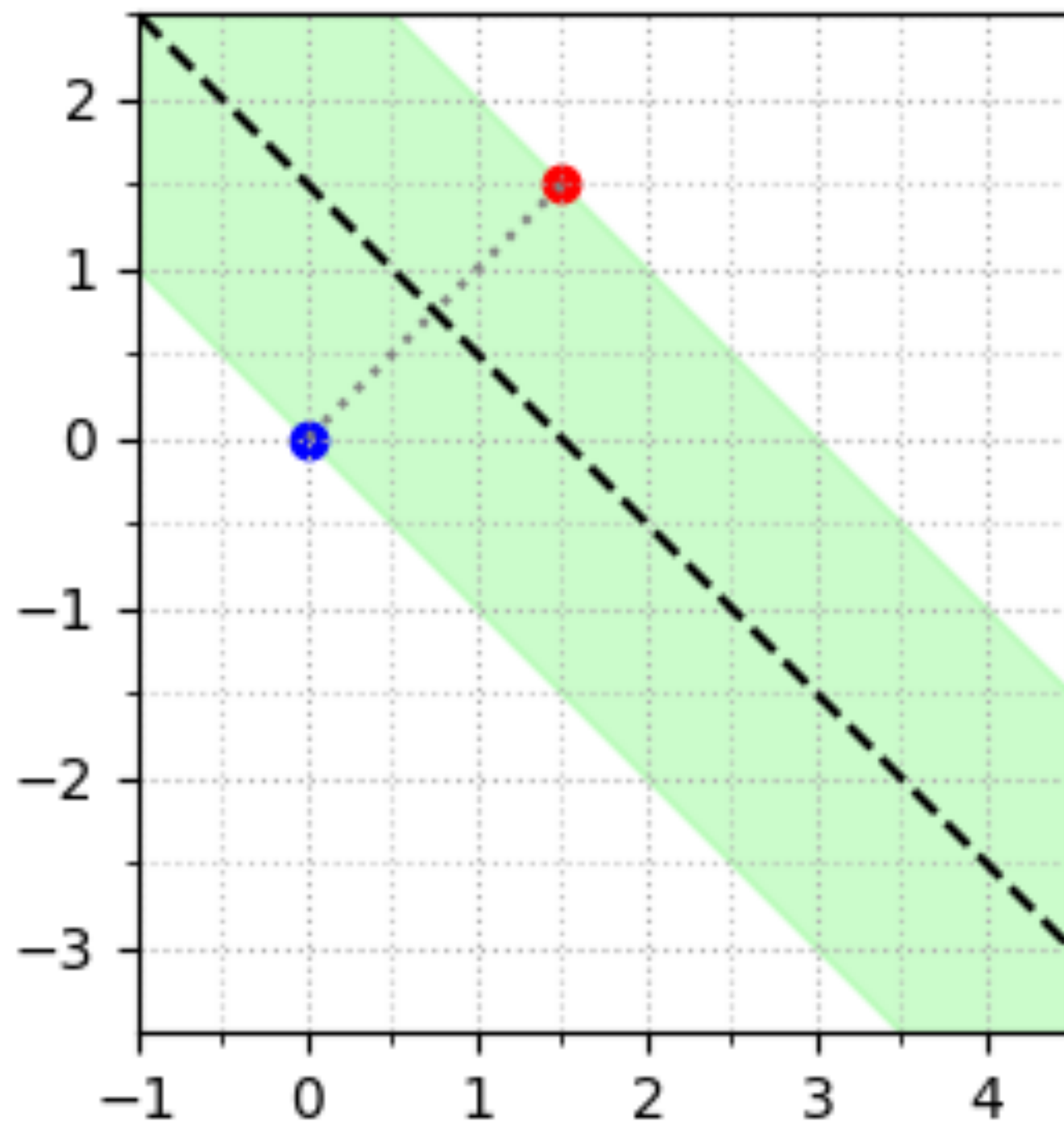
線性支持向量機
Linear Support Vector Machine



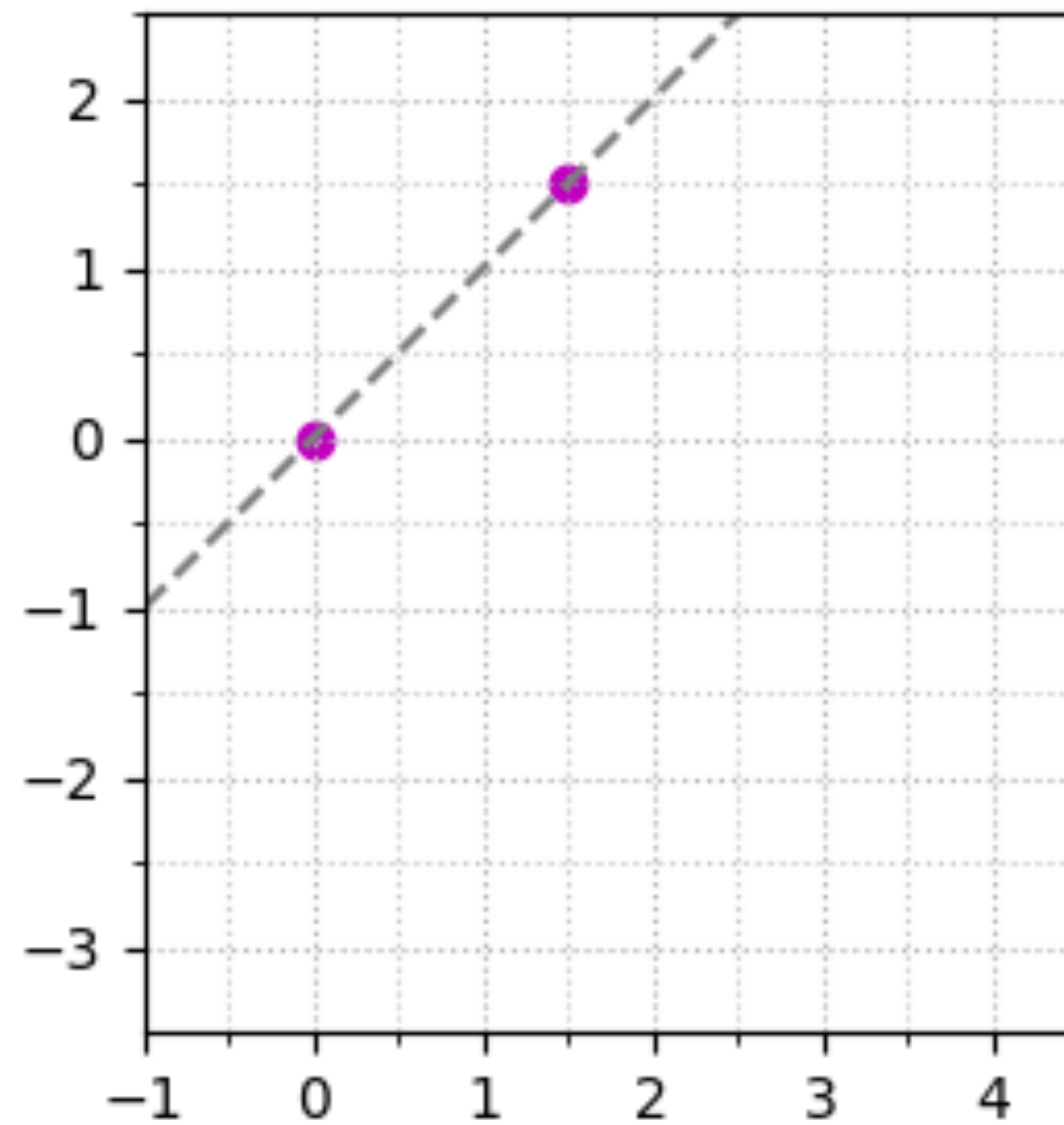
線性回歸
Linear Regression



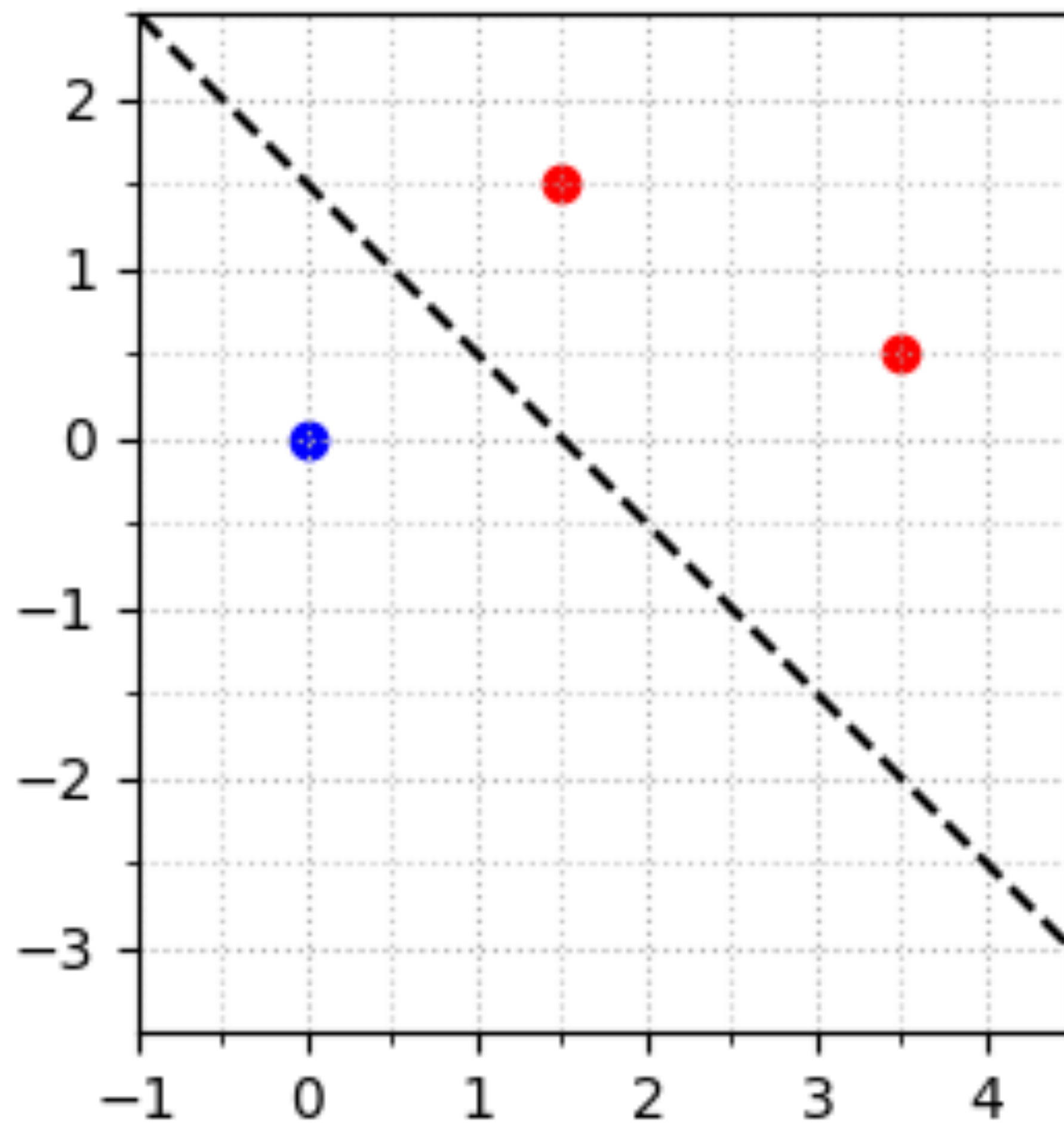
線性支持向量機
Linear Support Vector Machine



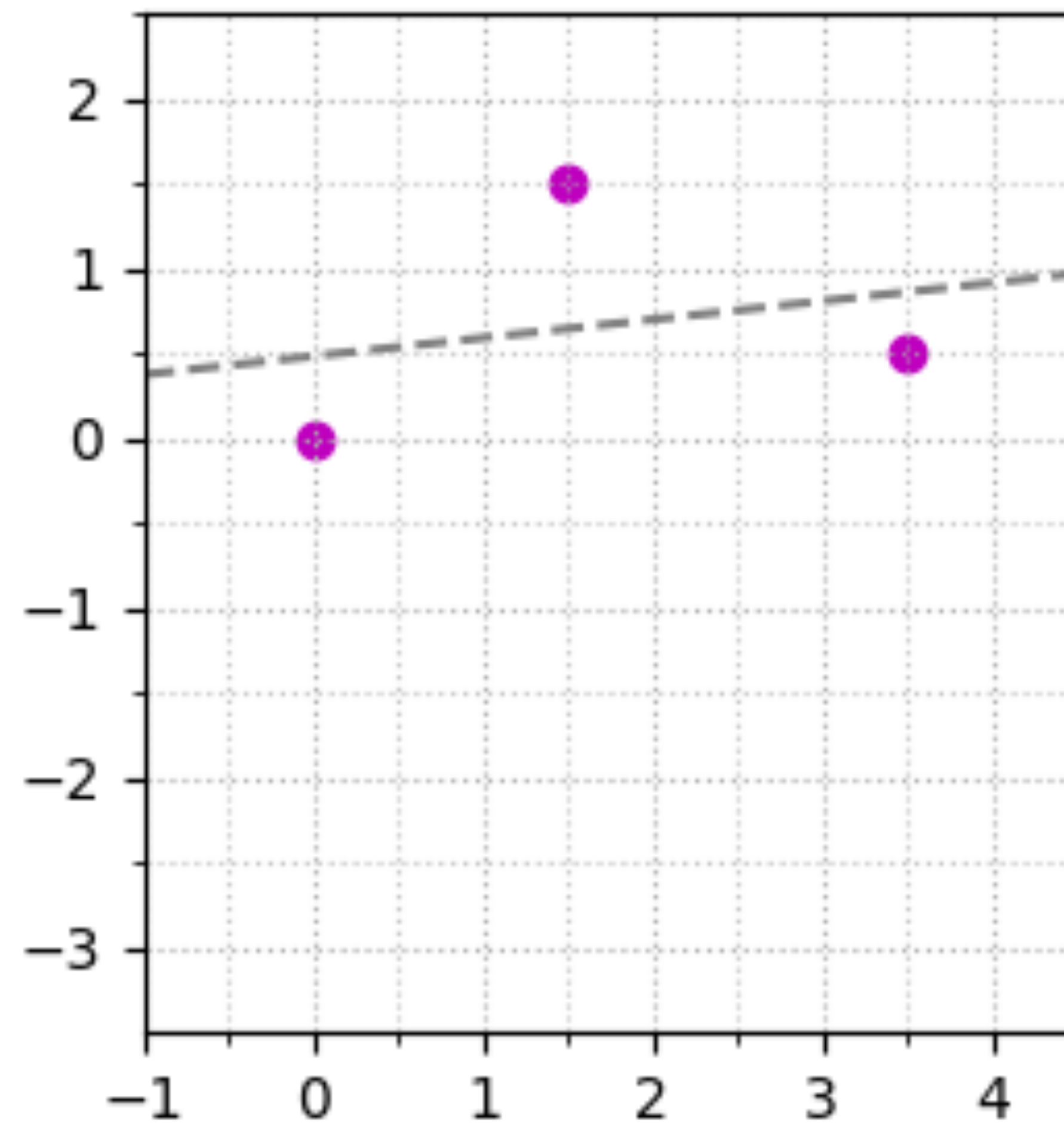
線性回歸
Linear Regression



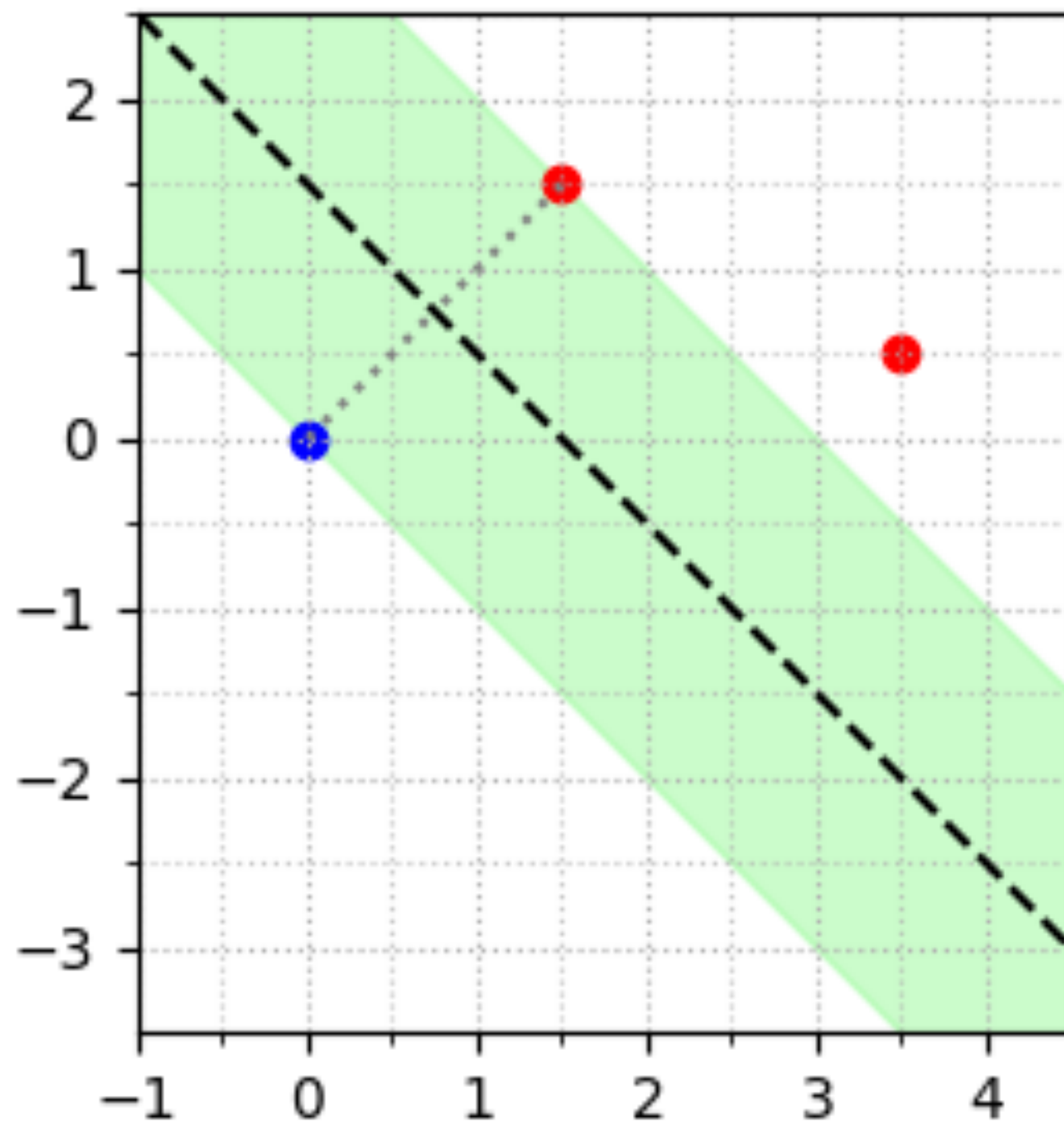
線性支持向量機
Linear Support Vector Machine



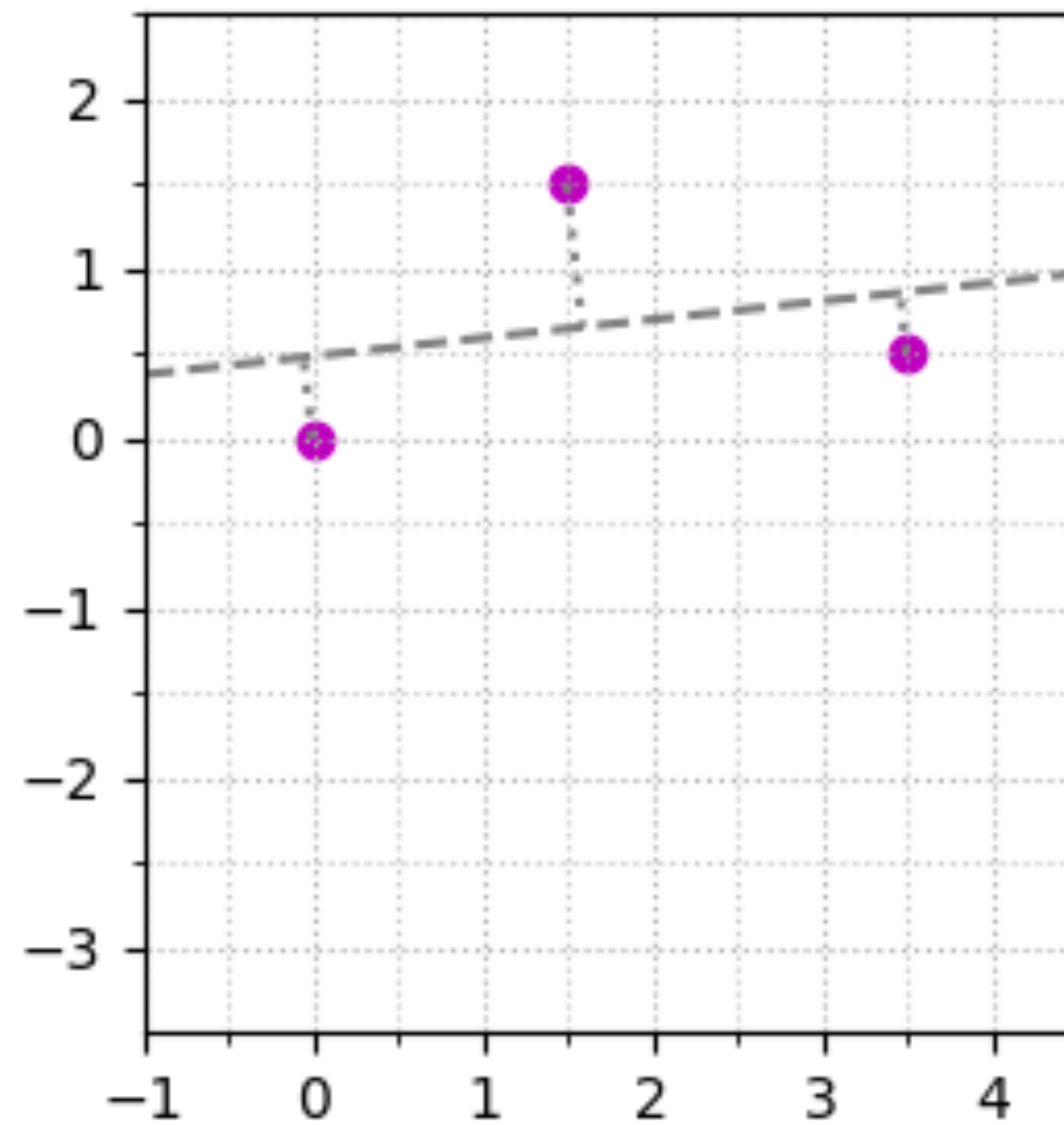
線性回歸
Linear Regression



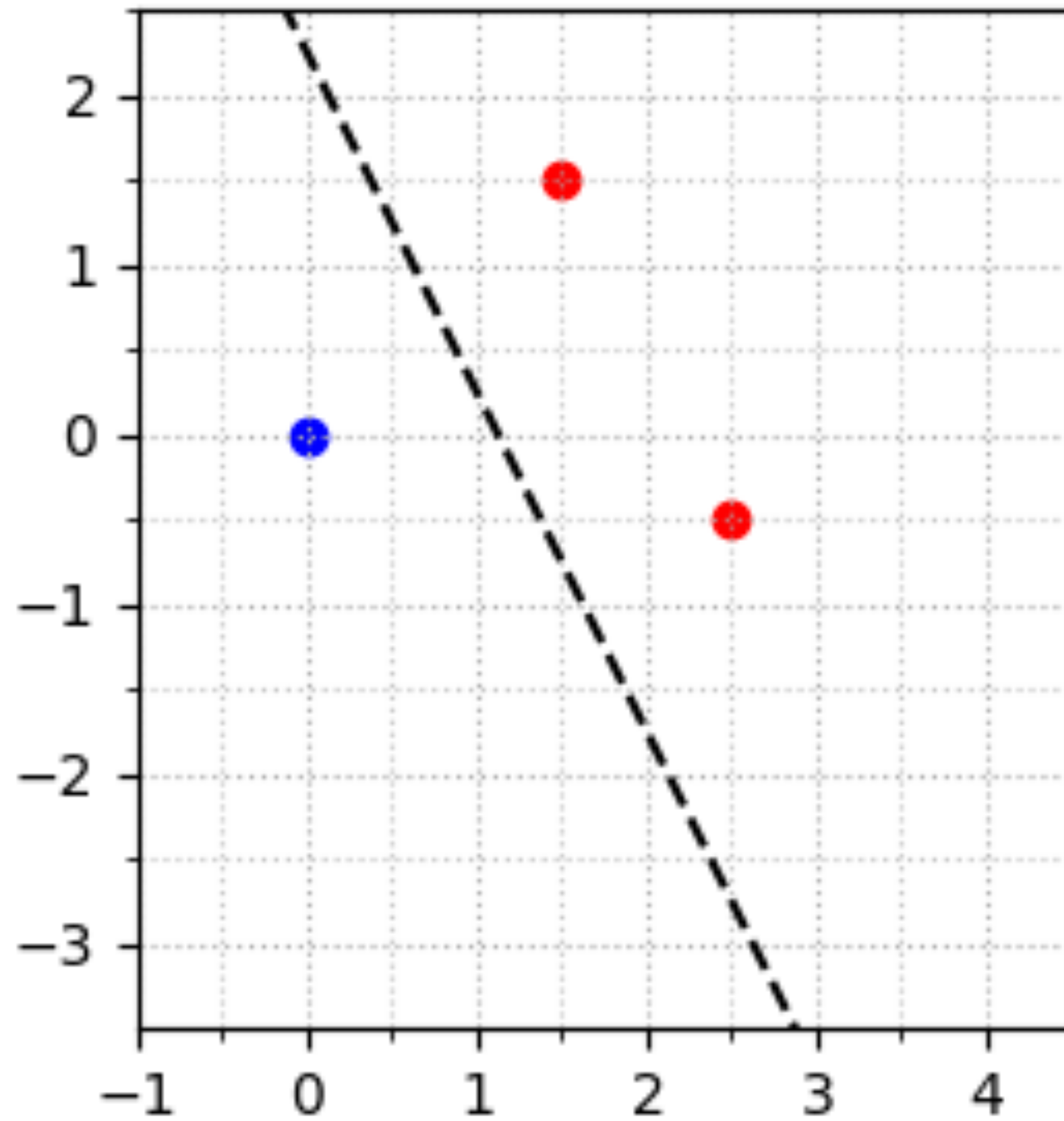
線性支持向量機
Linear Support Vector Machine



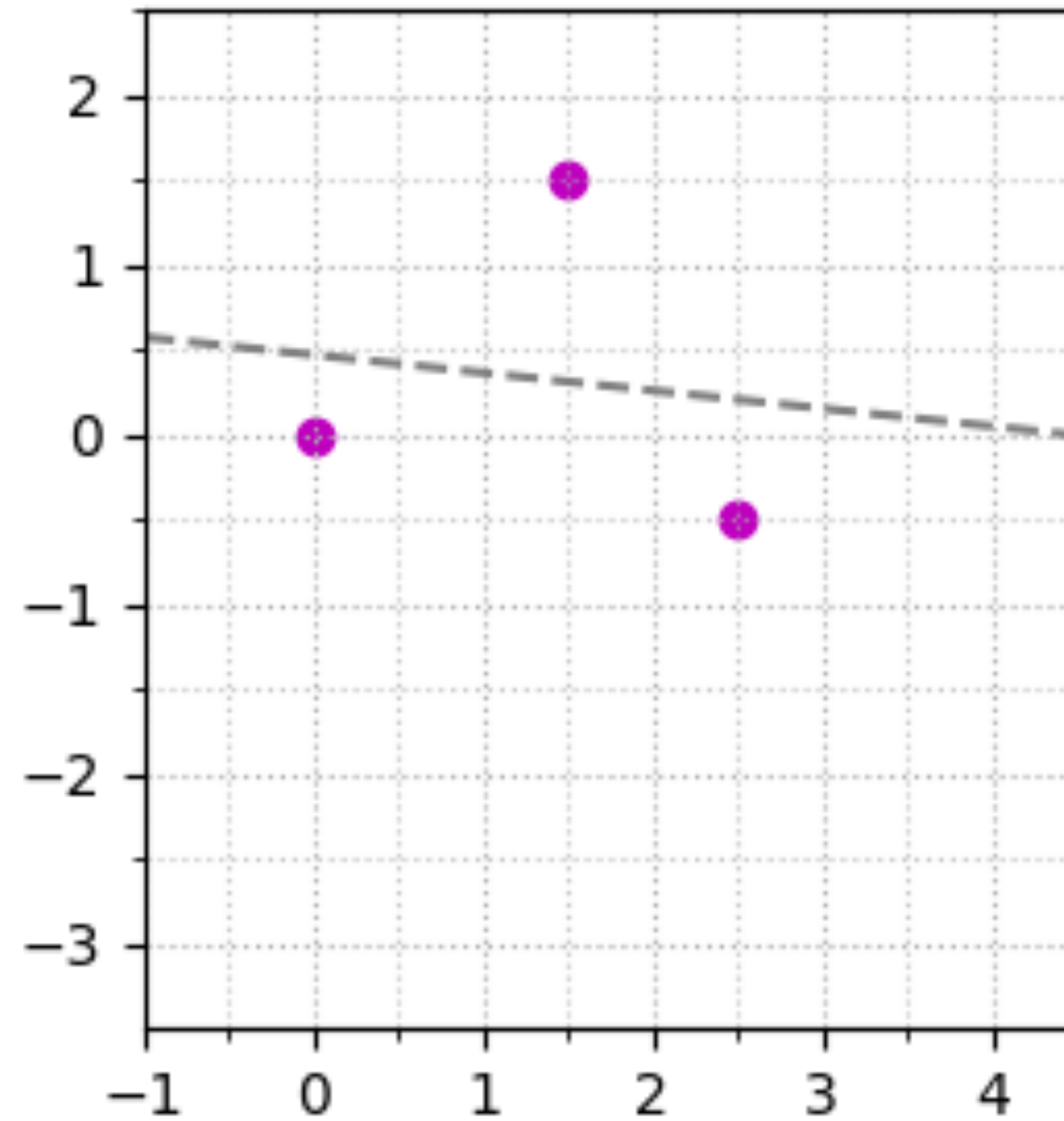
線性回歸
Linear Regression



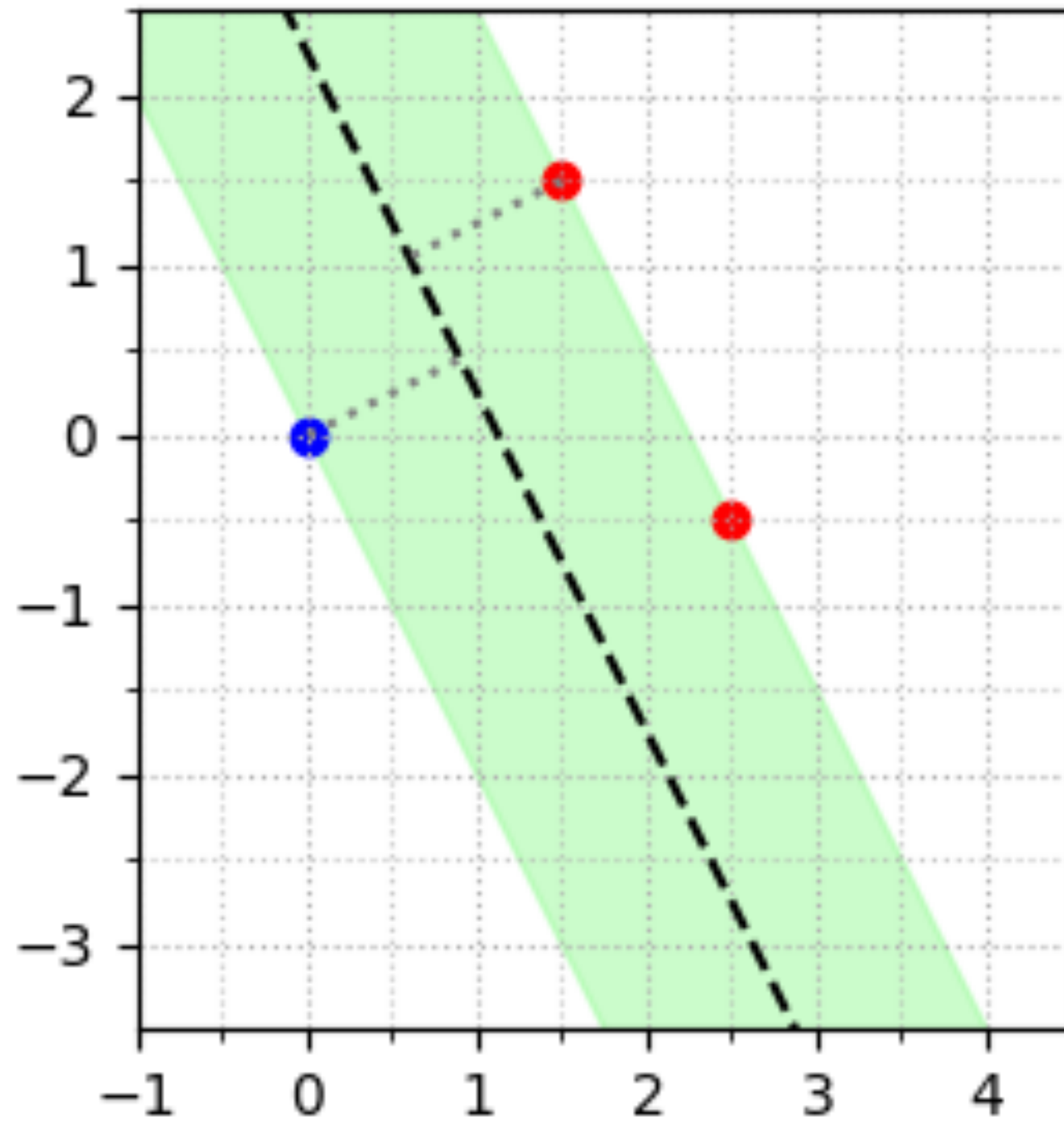
線性支持向量機
Linear Support Vector Machine



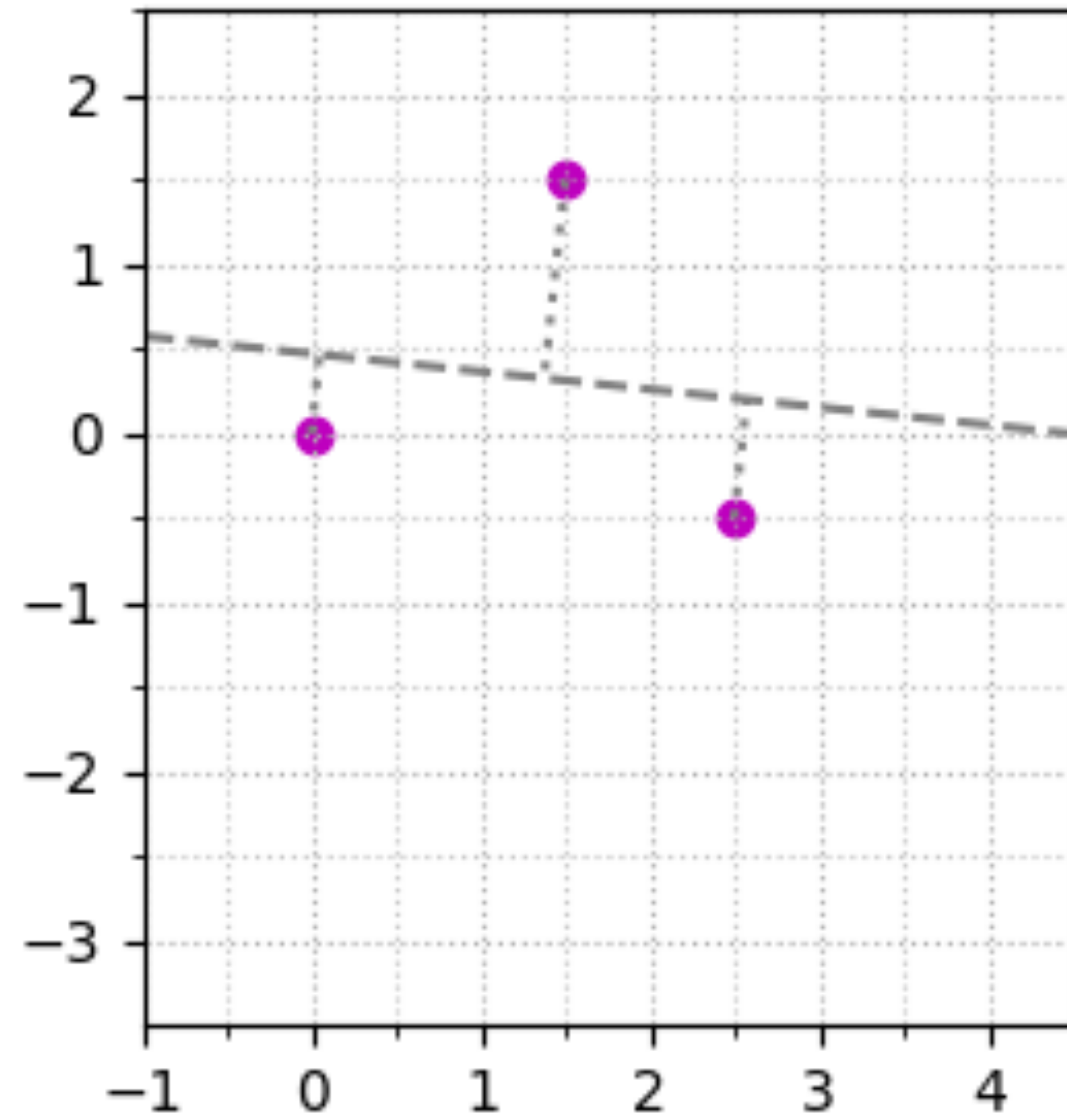
線性回歸
Linear Regression



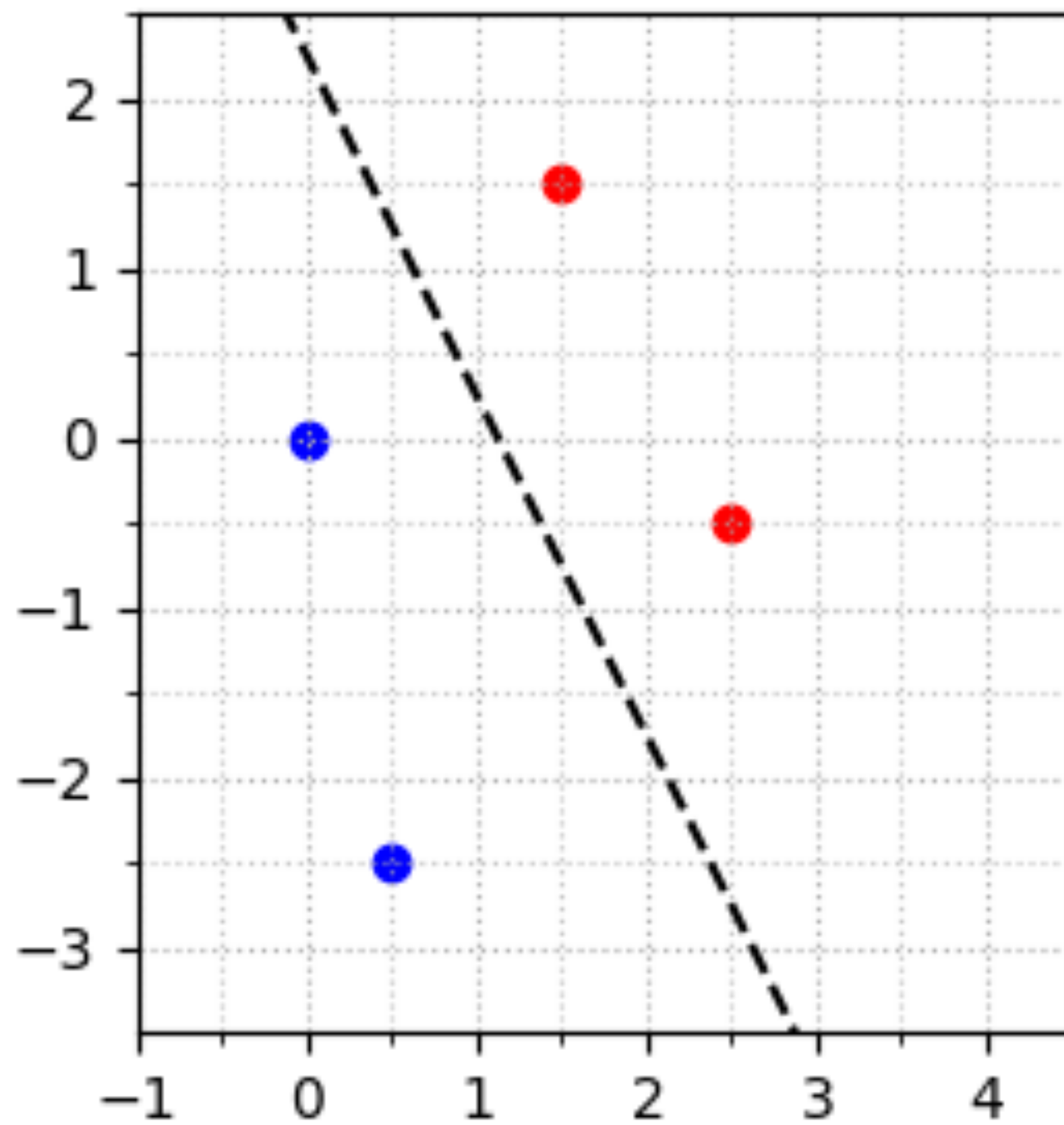
線性支持向量機
Linear Support Vector Machine



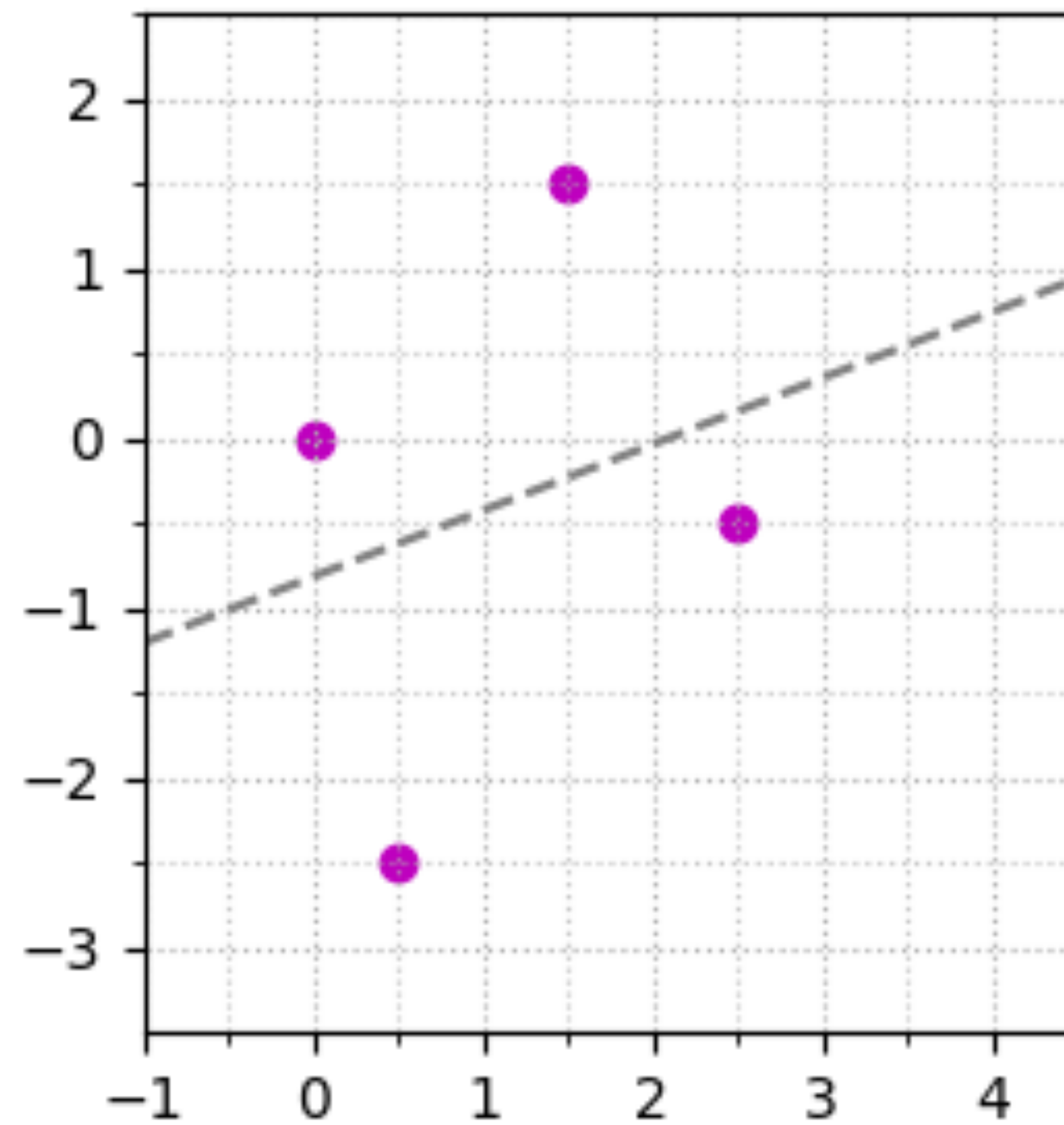
線性回歸
Linear Regression



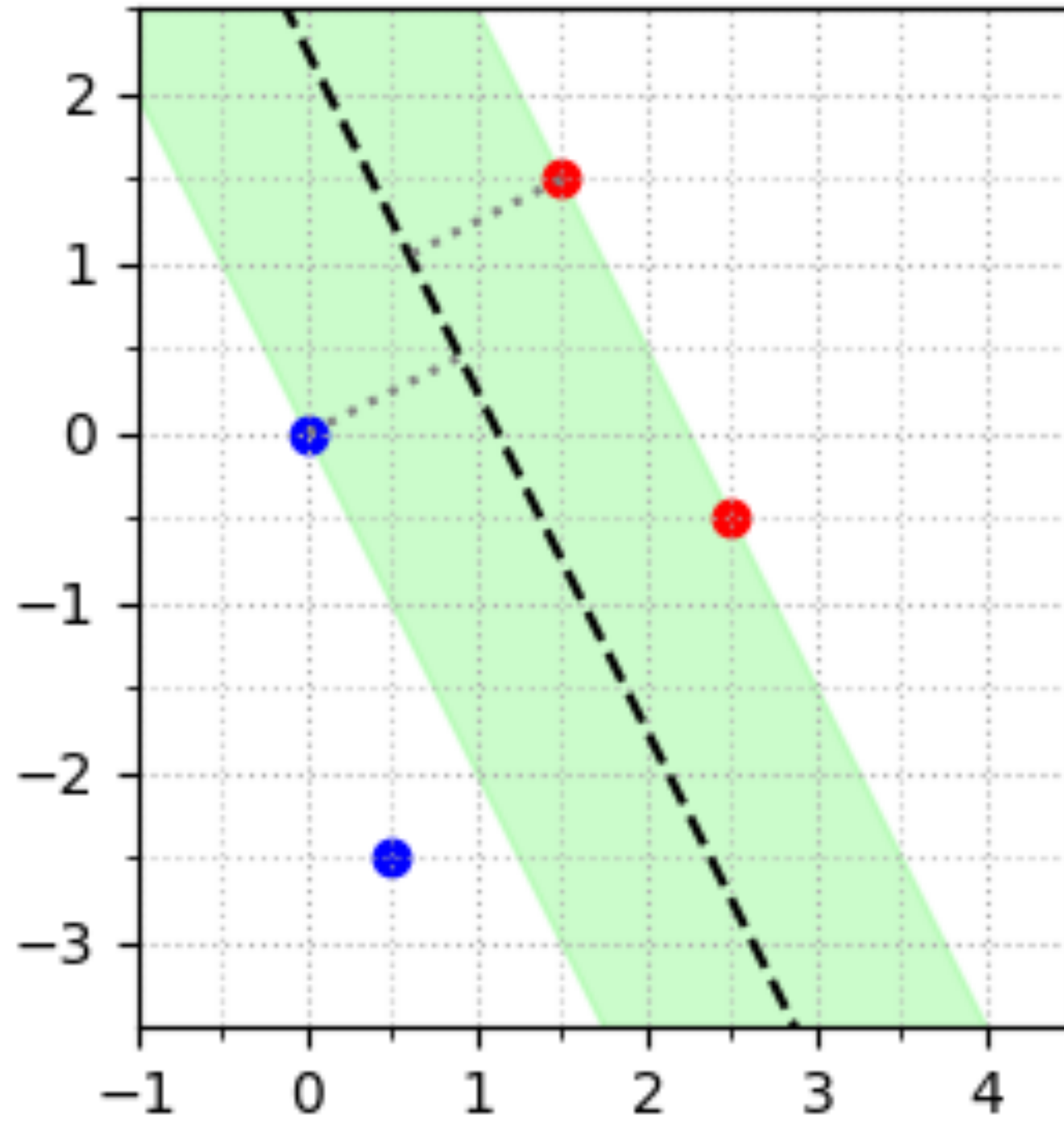
線性支持向量機
Linear Support Vector Machine



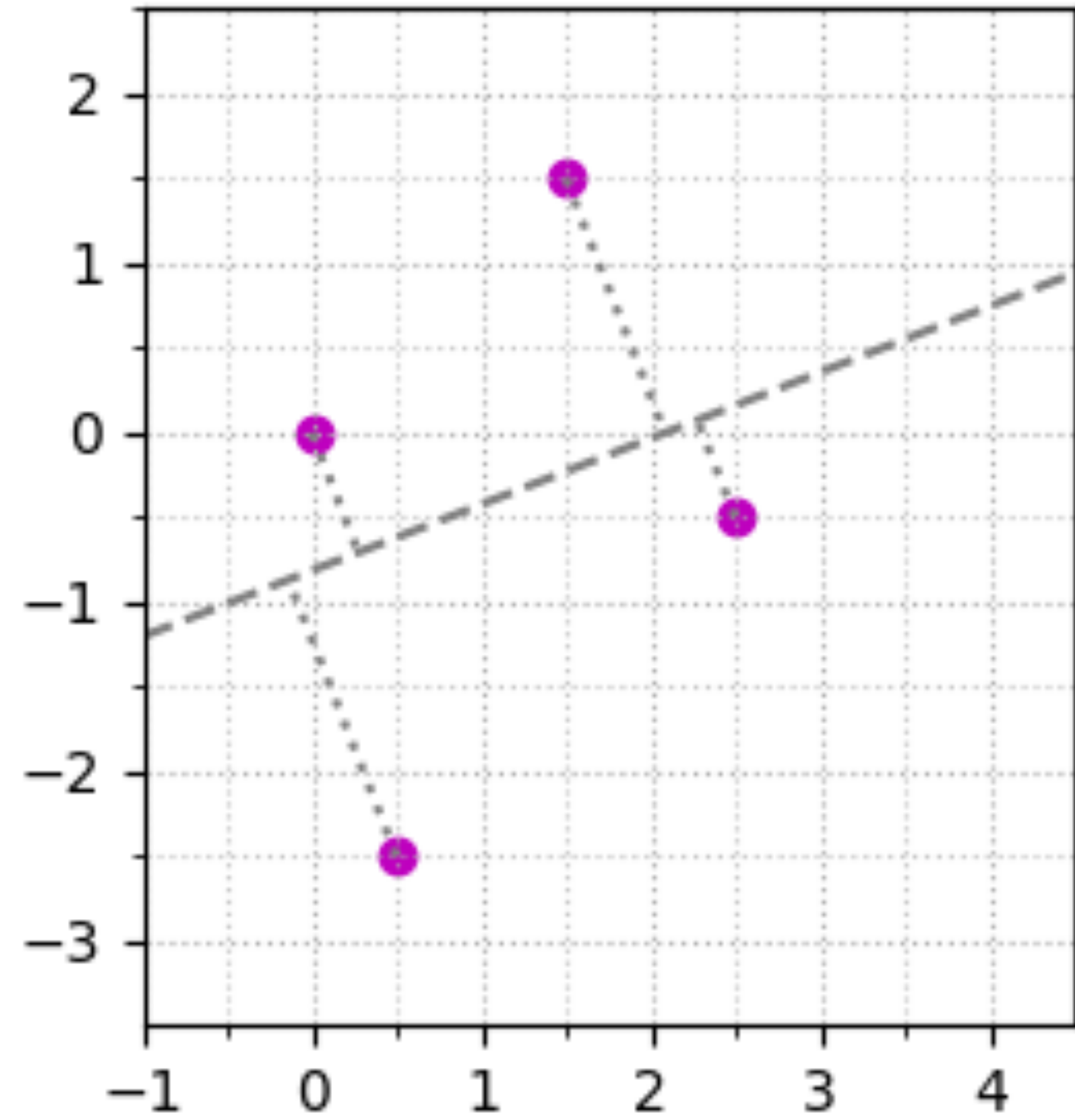
線性回歸
Linear Regression



線性支持向量機
Linear Support Vector Machine

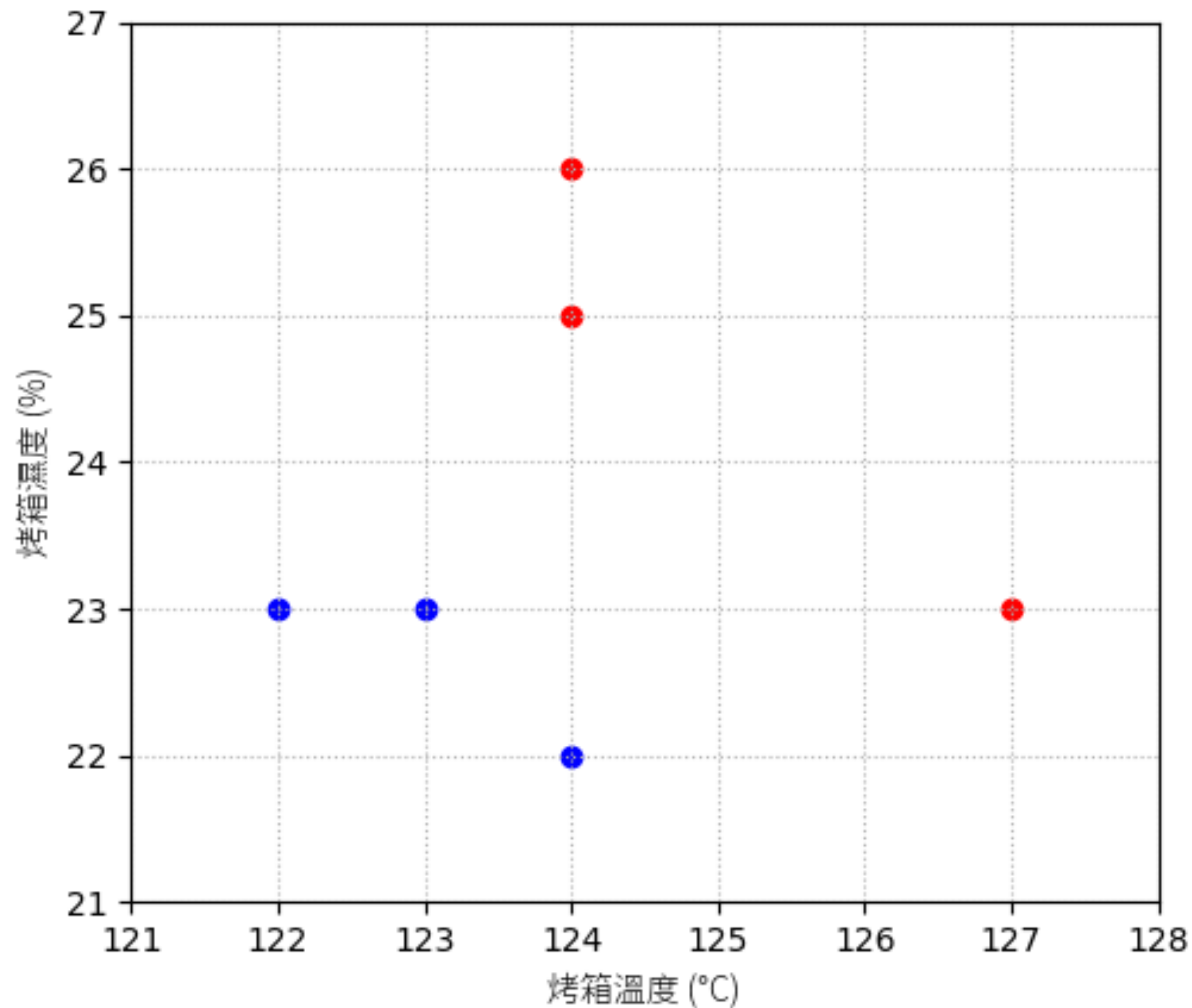


線性回歸
Linear Regression

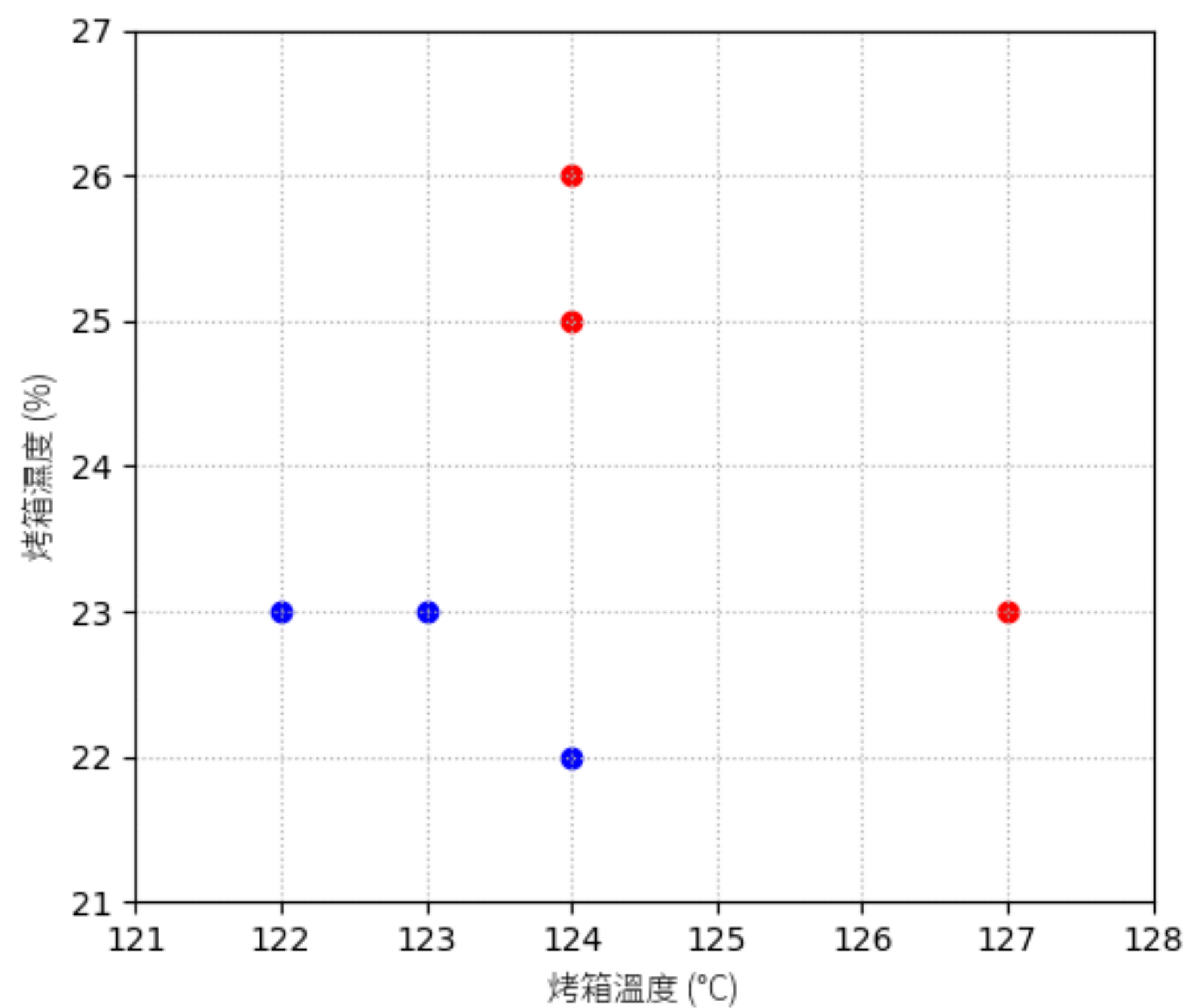
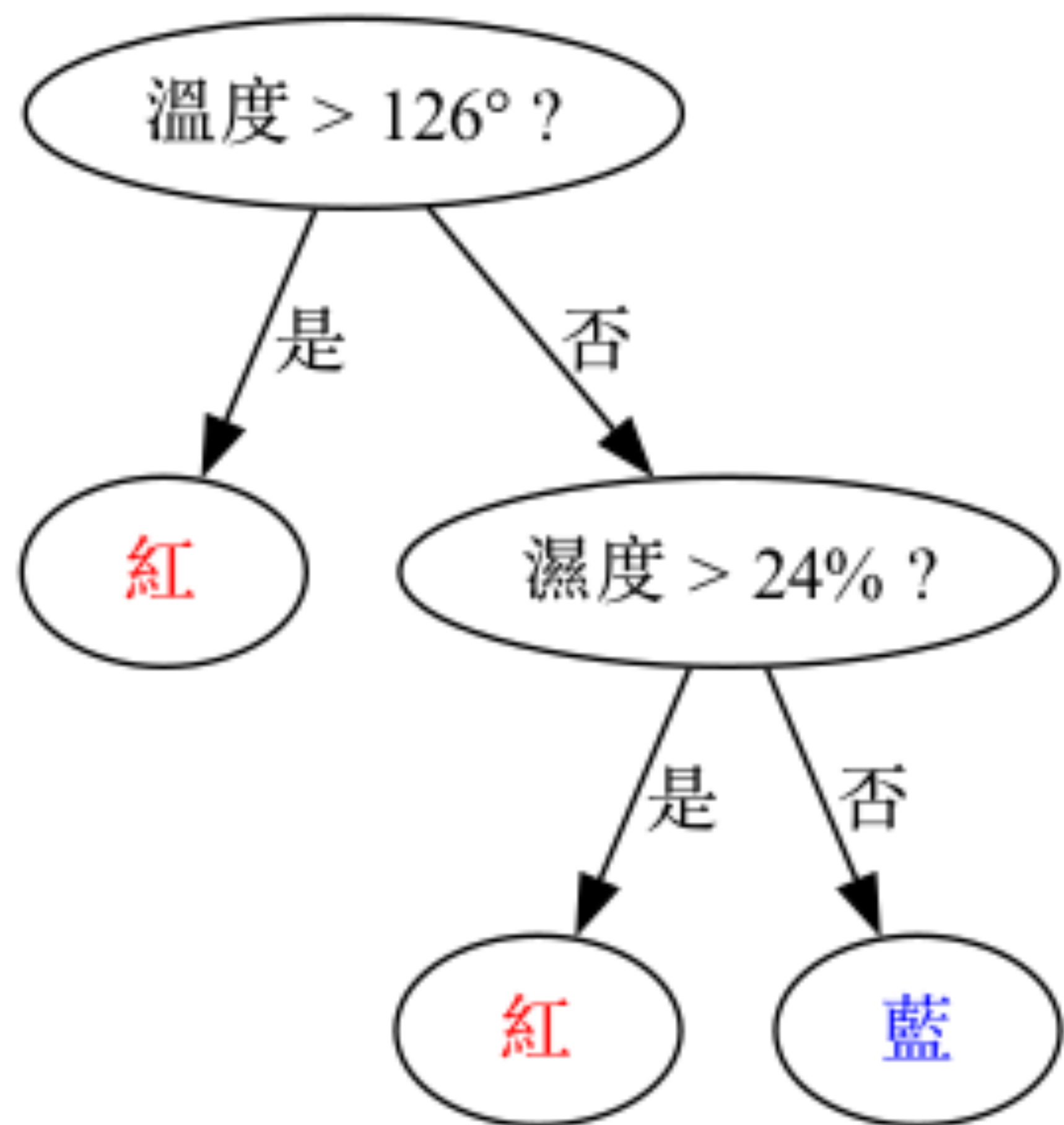


- 最短距離分類器
- kNN (k-最近鄰) 分類器
- 支持向量機
- 決策樹

決策樹 Decision Tree



Pizza 烘烤紀錄				
	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	2	127	23	●
	3	124	25	●
	4	122	23	●
	5	124	26	●
	6	124	22	●
測試集	t	124.55	23.25	?

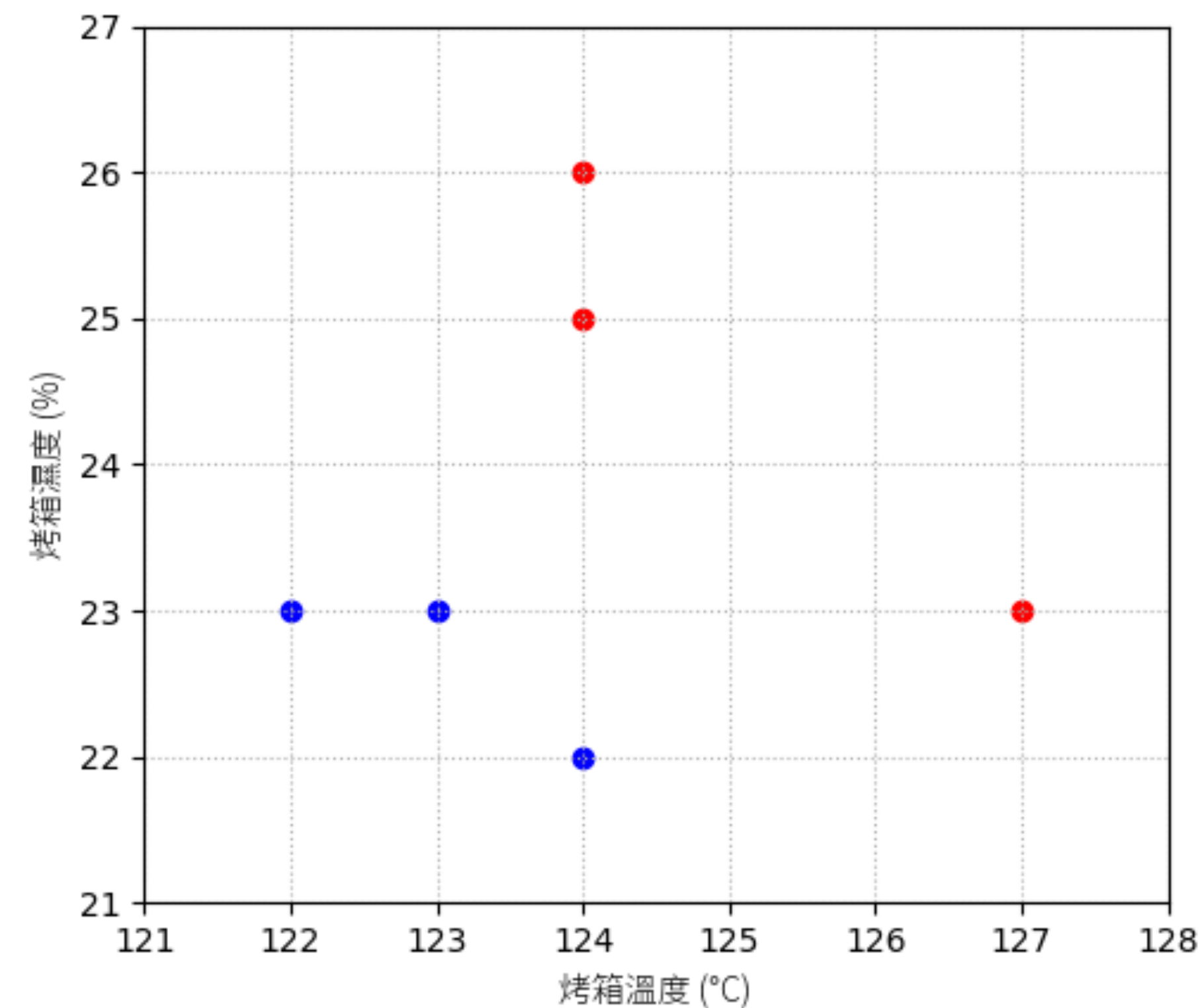
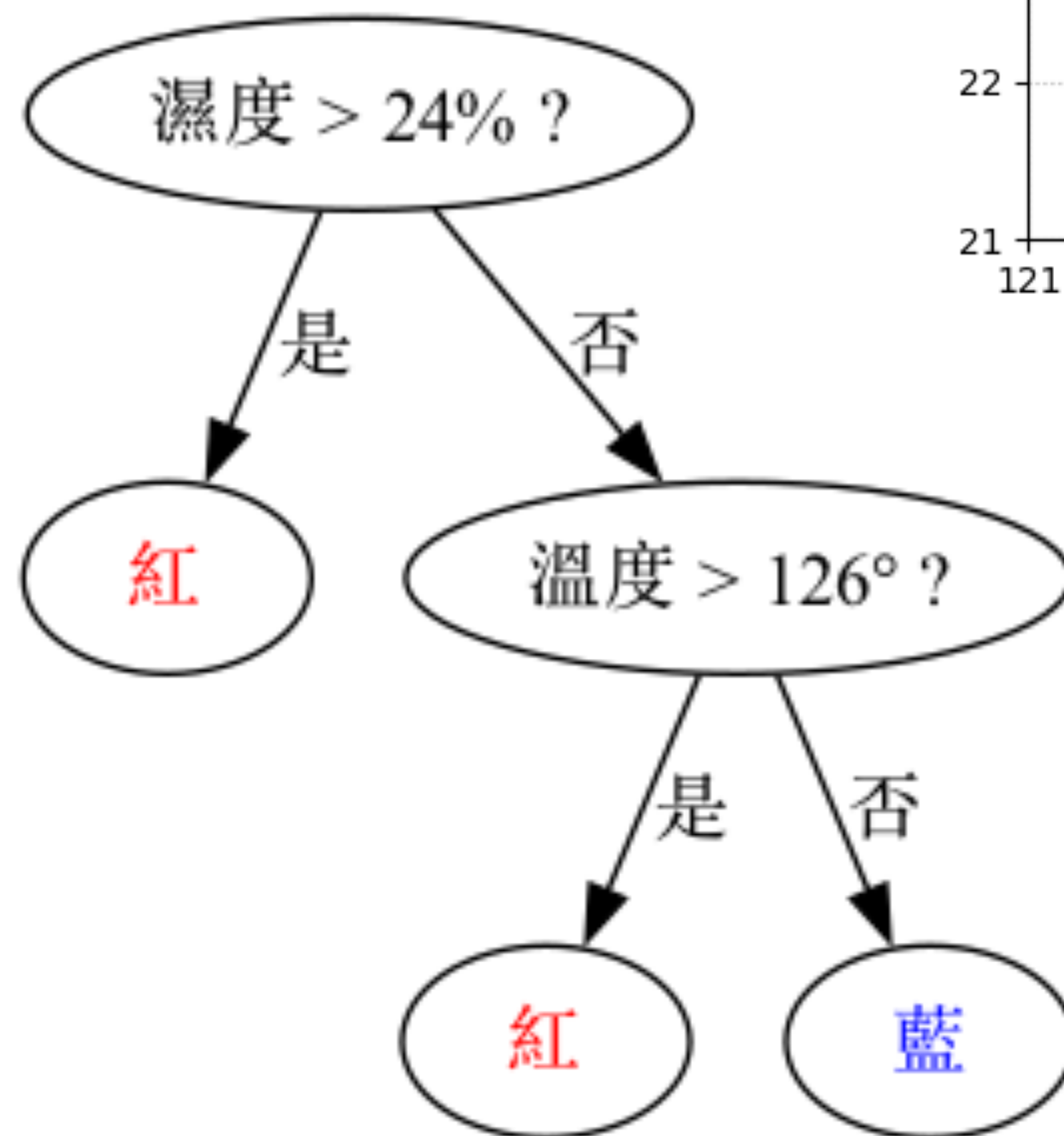
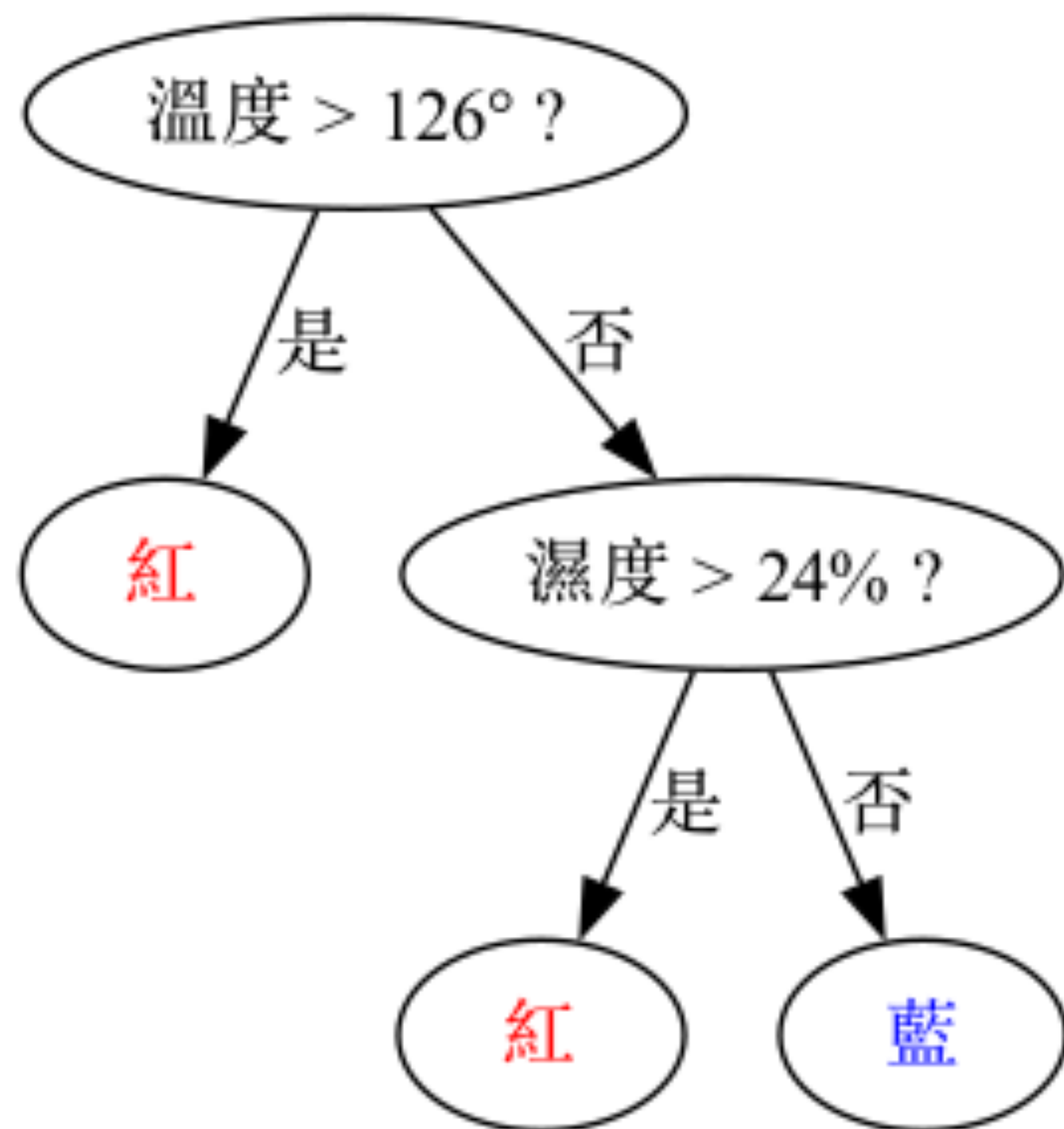


Pizza 烘烤紀錄				
	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	2	127	23	●
	3	124	25	●
	4	122	23	●
	5	124	26	●
	6	124	22	●
測試集	t	124.55	23.25	?

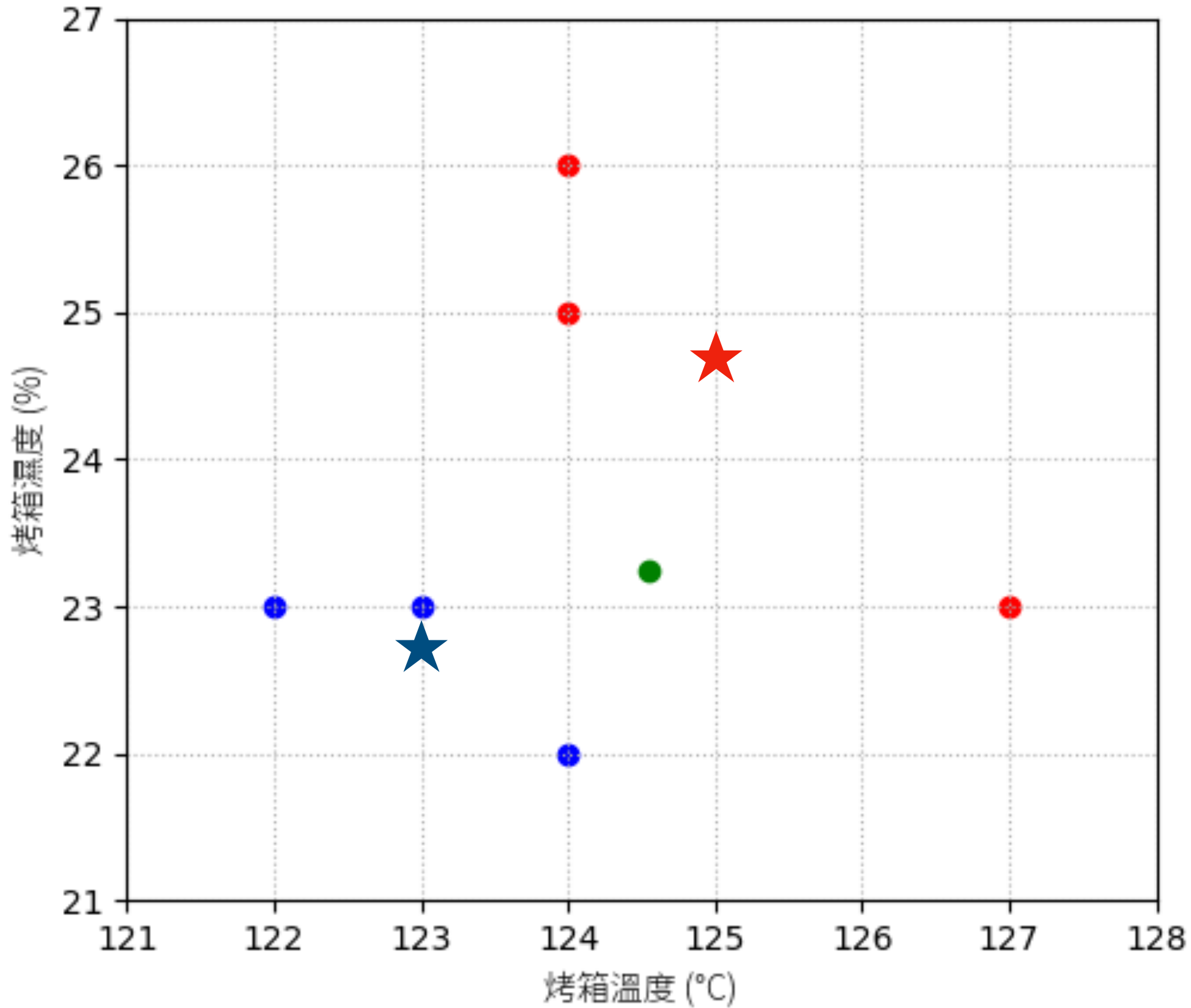
$$\frac{2}{3} \times 2 + \frac{1}{3} \times 1 = \frac{5}{3}$$

$$\frac{2}{3} \times 1 + \frac{1}{3} \times 2 = \frac{4}{3}$$

勝



- 最短距離分類器
- kNN (k-最近鄰) 分類器
- 支持向量機
- 決策樹

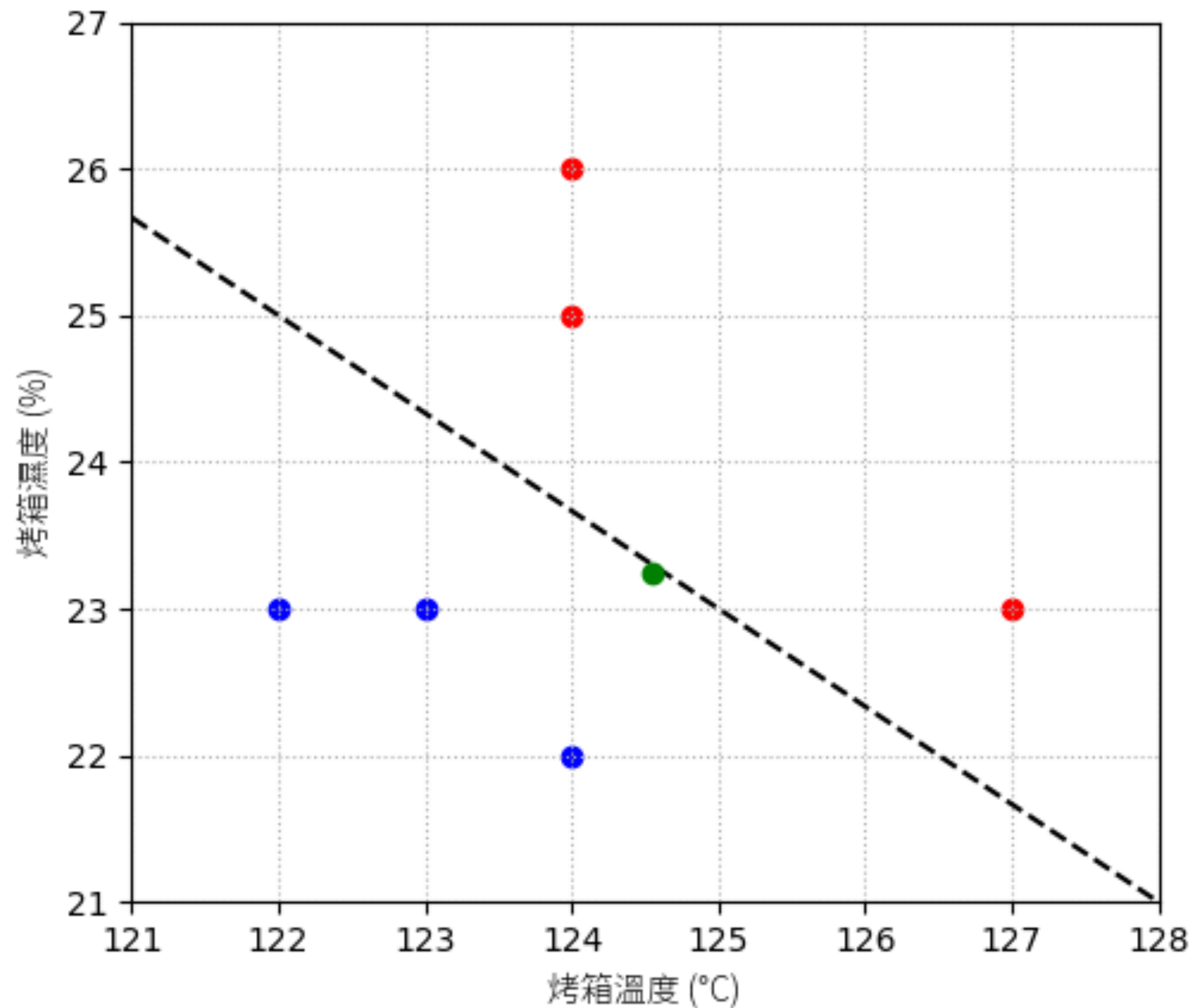


	編號	和★距離 ²	和★距離 ²	推論分類
測試集	t	2.7	2.2	●

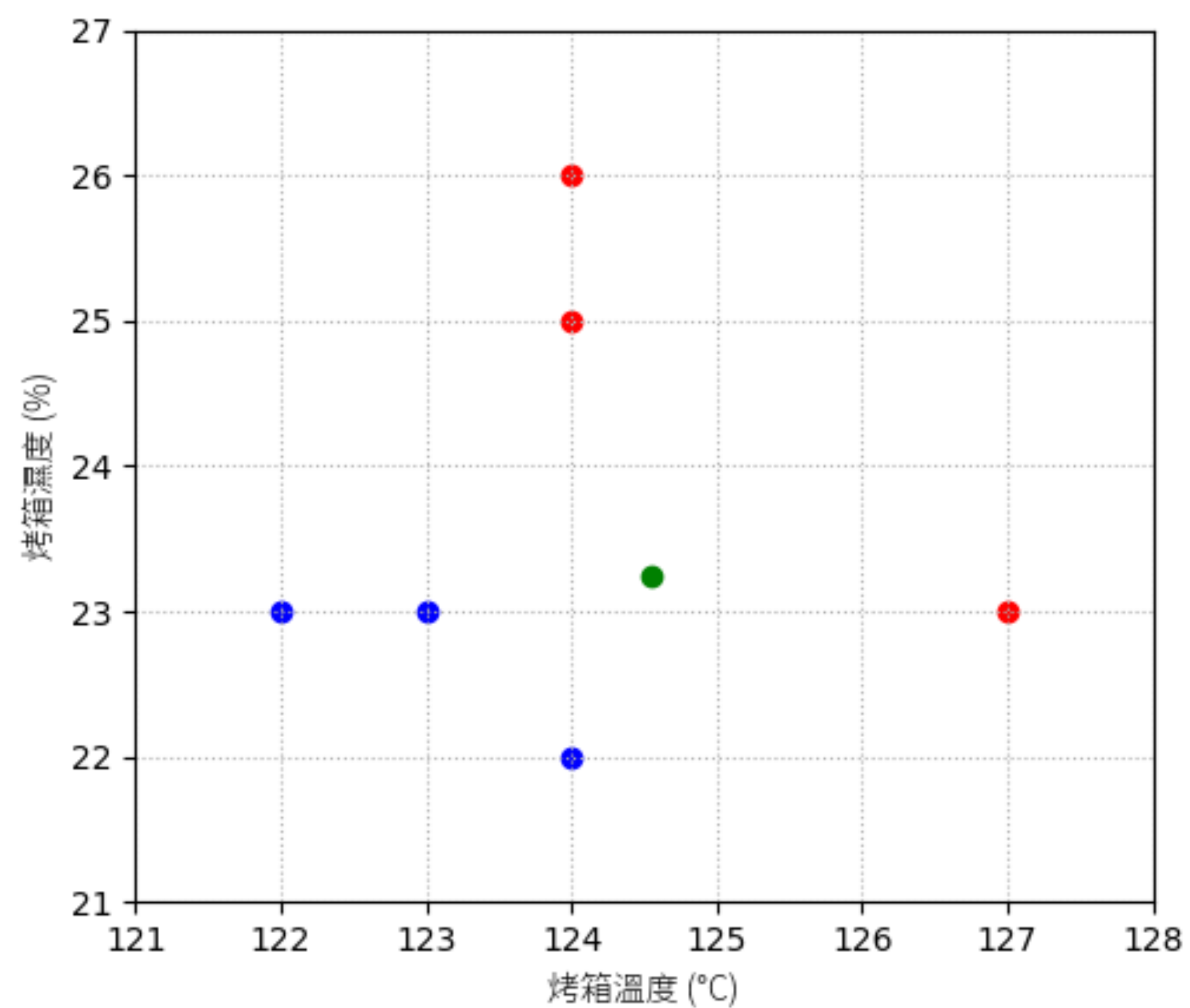
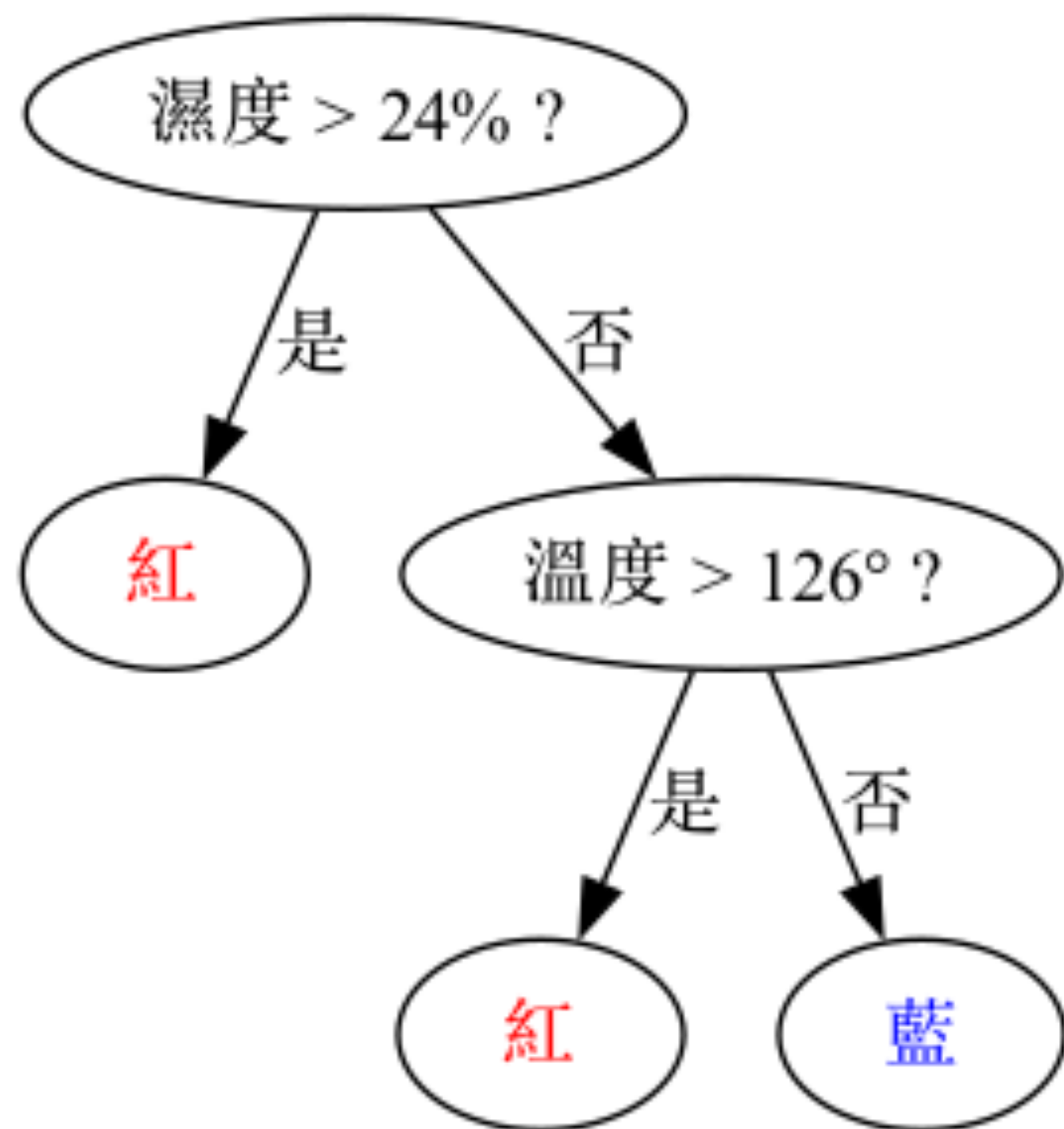
	編號	溫度 °C	濕度 %	評價
訓練集	2	127	23	●
	3	124	25	●
	5	124	26	●
平均值		125.0	24.7	★

	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	4	122	23	●
	6	124	22	●
平均值		123.0	22.7	★

	編號	溫度 °C	濕度 %	評價	和t距離 ²	順序	1NN 推論	3NN 推論	5NN 推論	7NN 推論
訓練集	1	123	23	●	2.5	2nd		✓	✓	✓
	2	127	23	●	6.1	4th			✓	✓
	3	124	25	●	3.4	3rd		✓	✓	✓
	4	122	23	●	6.6	5th			✓	✓
	5	124	26	●	7.9	6th				✓
	6	124	22	●	1.9	1st	✓	✓	✓	✓
	7	122	21	●	11.6	8th				
	8	126	26	●	9.7	7th				✓
測試集	t	124.55	23.25	?			●	●	●	●



Pizza 烘烤紀錄				
	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	2	127	23	●
	3	124	25	●
	4	122	23	●
	5	124	26	●
	6	124	22	●
測試集	t	124.55	23.25	?



Pizza 烘烤紀錄				
	編號	溫度 °C	濕度 %	評價
訓練集	1	123	23	●
	2	127	23	●
	3	124	25	●
	4	122	23	●
	5	124	26	●
	6	124	22	●
測試集	t	124.55	23.25	?

機器學習

2023 (C) Elton Huang

訓練：
training/fitting

訓練集

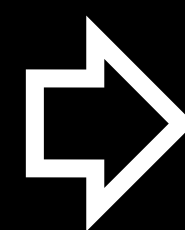


模型

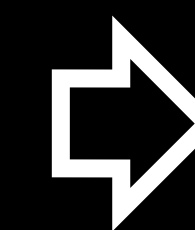


推論/預測：
inference/
prediction

輸入



模型

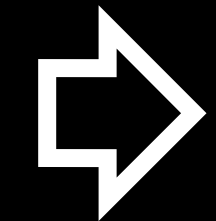
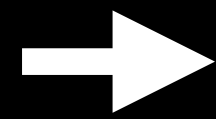


輸出

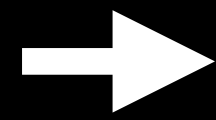
機器學習文獻變數使用習慣

2023 (C) Elton Huang

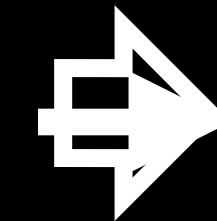
溫度 x



濕度 y



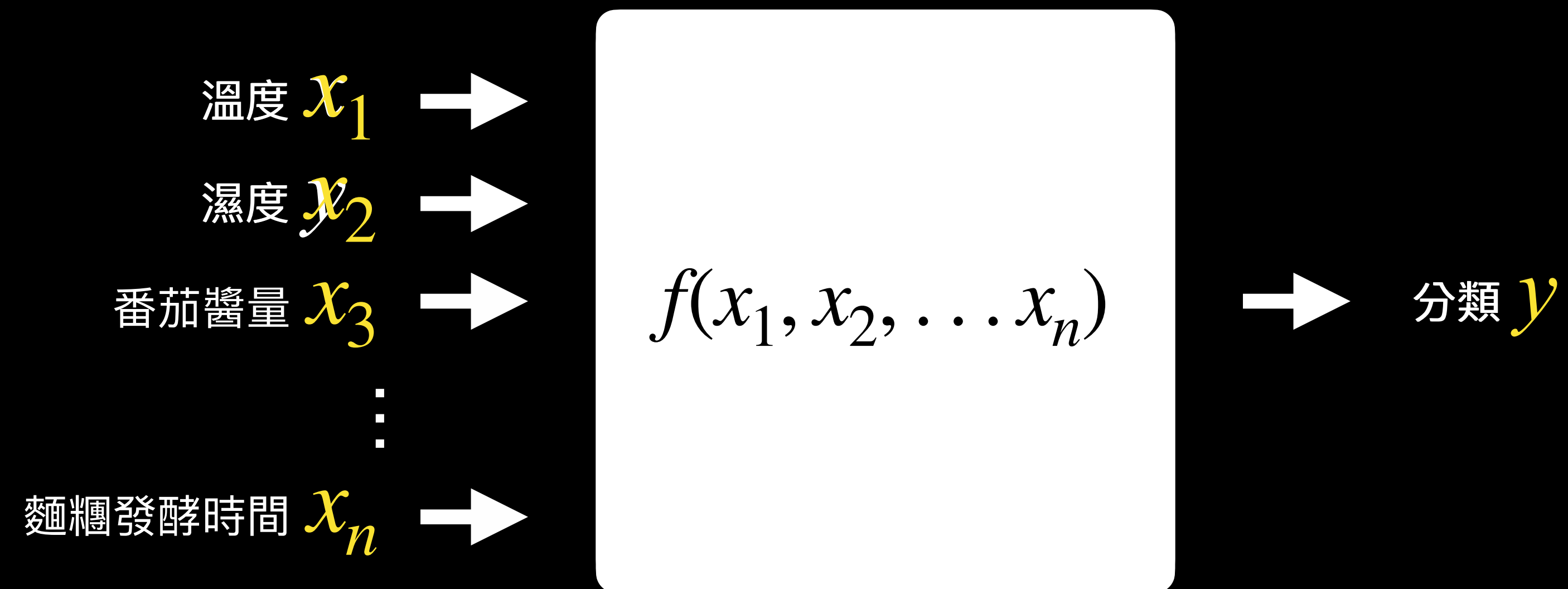
$f(x, y)$



分類

機器學習文獻變數使用習慣

2023 (C) Elton Huang



參考資料

- 鴻海教育基金會《人工智慧導論》全華圖書 2019/07/01
- 周志华《机器学习》清华大学出版社 2016 年 1 月
- 涂益郎等《和 AI 做朋友 - 相知篇》教育部電子書民國 108 年 8 月
(第 3 頁頁首關於圖靈測試的敘述或已須修改)
- Dustin Boswell, Introduction to Support Vector Machines, August 6, 2002
- J. R. Quinlan, Induction of Decision Trees, 1986
 - Khan Academy, Information Entropy, 2014
- ChatGPT 諮詢

延伸學習

- 蔡炎龍老師 2021 北一女 Python 人工智慧工作坊
- How to Train TensorFlow Lite Models Using Google Colab
- DeepMind AI 參賽寫程式，AlphaCode 擊敗近半數工程師！為何專家說取代人類還很遙遠？

ChatGPT

- 大規模視覺基礎模型之機會與挑戰
- 李宏毅老師 ChatGPT (可能)是怎麼煉成的 - GPT 社會化的過程
- 李宏毅老師【生成式AI】 ChatGPT 原理剖析 (1/3)一對 ChatGPT 的常見誤解
- 陳縉儂老師 OpenAI ChatGPT 驚驗眾人的對話互動式 AI
- What Is ChatGPT Doing ... and Why Does It Work?

「機器學習」 評量測驗考試規則：

1. 不可討論、不可使用手機
2. 可以使用的電腦工具：
Geogebra、Desmos、Google Sheet、Excel、
Google、電腦中的計算機。如須使用其它工具
請提出詢問。

使用電腦須全程錄影。學校電腦錄影方法：
 - 2.1. 左下角 App 目錄 →
Cyberlink ScreenRecorder →
開始錄影，開始操作作答
 - 2.2. 結束作答後左下角 App 圖示打開 →
停止錄影 → 繳交彈出視窗的 mp4 錄影檔
3. 得到答案的過程須交代清楚：
 - 3.1. 紙筆演算交演算過程，
 - 3.2. 用 Desmos 或 Geogebra 交分享連結，
試算表交檔案，
 - 3.3. 使用電腦交全程錄影。

有用到什麼就要交什麼。
4. 範例試題
範例解答