



# 機器學習模型

## Support Vector Machine (SVM)

如何利用SVM來預測股票的漲跌？



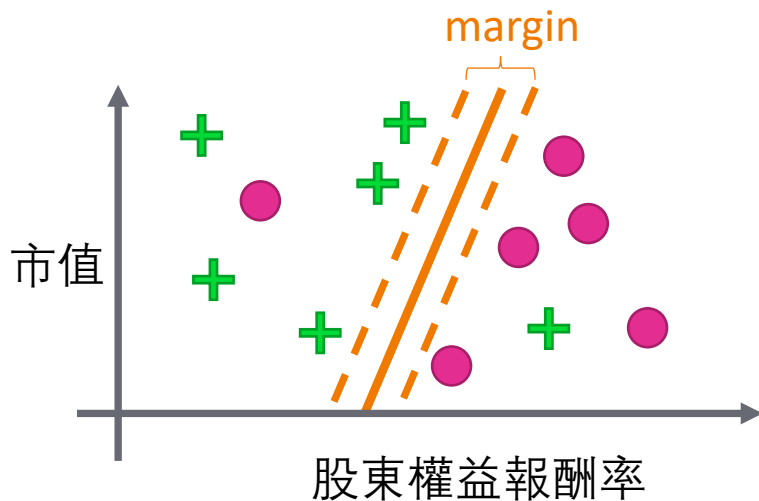
# SVM究竟是Classifier還是Regressor?

- 不論哪種都可以！
  - SVC: Support Vector Classifier
  - SVR: Support Vector Regressor



# SVC: Support Vector Classifier

□ SVC 找出分類的邊界 (hyperplane)



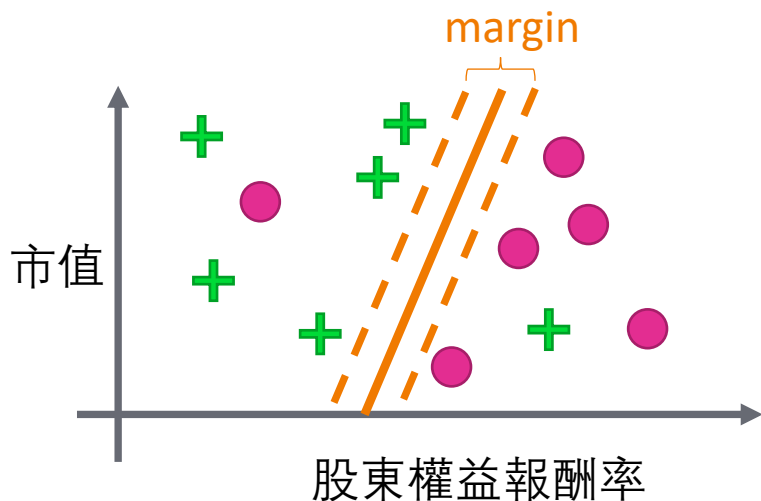
● 一個月後上漲的股票

+ 一個月後下跌的股票



# 加入一點點點數學...

□ 如何找出這條線？



□ 這條線的數學式：

$$a \text{ 股東權益報酬率} + b \text{ 市值} = c$$

□ 我們希望找到a、b、c的值使得：

□ 會漲的股票符合的條件：

$$a \text{ 股東權益報酬率} + b \text{ 市值} - c > 1$$

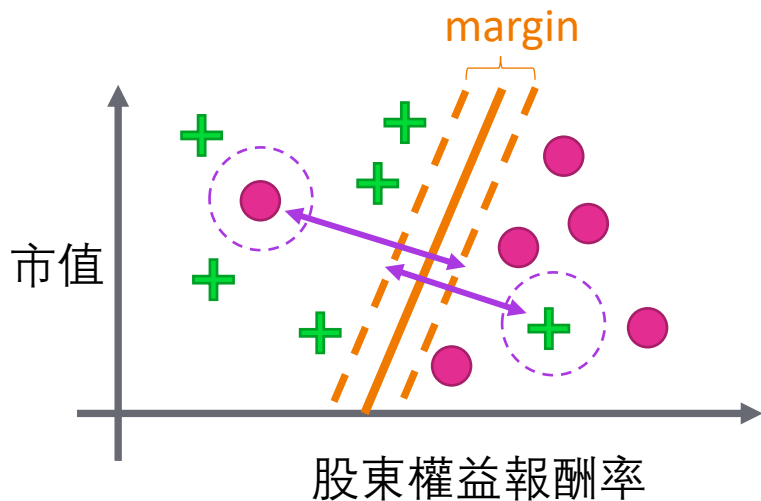
□ 會跌的股票符合的條件：

$$a \text{ 股東權益報酬率} + b \text{ 市值} - c < -1$$



# 一點點的演算法

□ 如何找出這條線？



## 優化演算法

調整、移動這條線，使

1. 分類錯誤最小

紫色線段的長度最小

2. Margin寬度盡可能增加



# 最佳化：Optimization

□ 我們希望縮小： $1/\text{Margin寬度} + \text{分類錯誤}$

□ 究竟  $\text{Margin寬度}$  和  $\text{分類錯誤}$  哪個重要呢？

□ 不一定，所以額外使用「懲罰參數」 $C > 0$

□ 修正後： $1/\text{Margin寬度} + C * \text{分類錯誤}$

□  $C$  越大：分類錯誤懲罰越重（可能會造成overfitting）

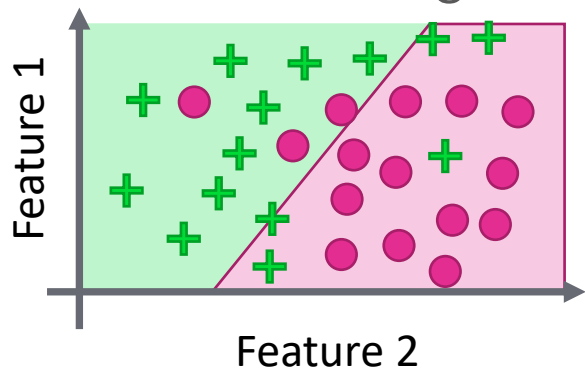
□  $C$  越小：越包容分類錯誤（可能會造成underfitting）



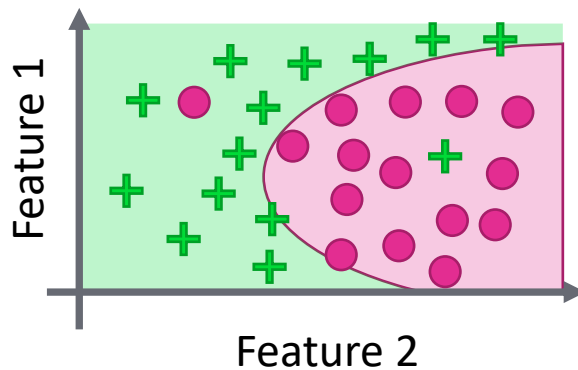
# Overfitting and Underfitting

C

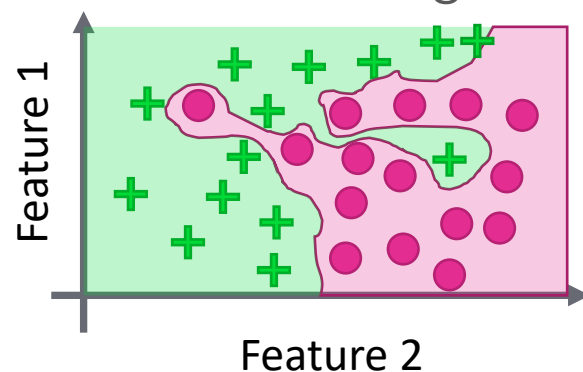
Underfitting



Optimum



Overfitting





# Kernel functions

- 
- Linear

$$a \text{ 股東權益報酬率} + b \text{ 市值} = c$$

- 
- Polynomial

$$[\gamma(a \text{ 股東權益報酬率} + b \text{ 市值}) + r]^d = c$$

- ## Radial bias

$$\exp(-\gamma[(a - \text{股東權益報酬率})^2 + (b - \text{市值})^2]) = c$$

- 
- Sigmoid

$$\tanh((a \text{ 股東權益報酬率} + b \text{ 市值}) + r) = c$$





# Parameter Tuning

## Parameters

## Default

## 意義

C

1

懲罰參數

kernel

rbf (Radial bias )

分類的 gap 的特性

degree

3

曲線參數

gamma

$1/n\_features$

曲線參數

coef0

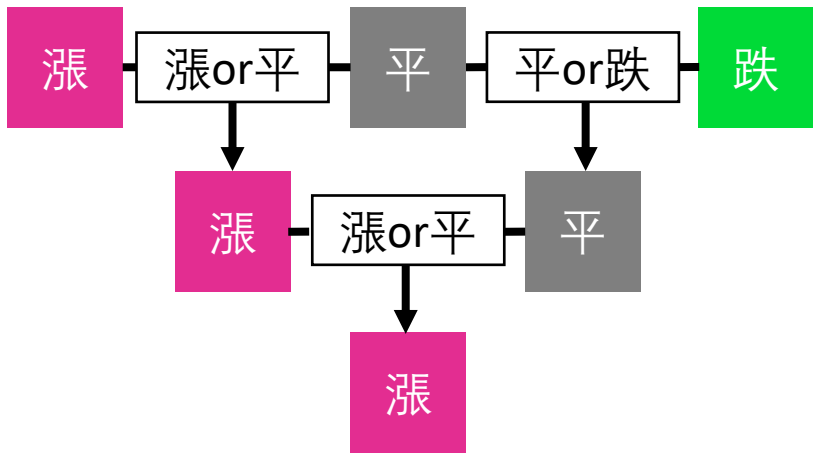
0

曲線參數

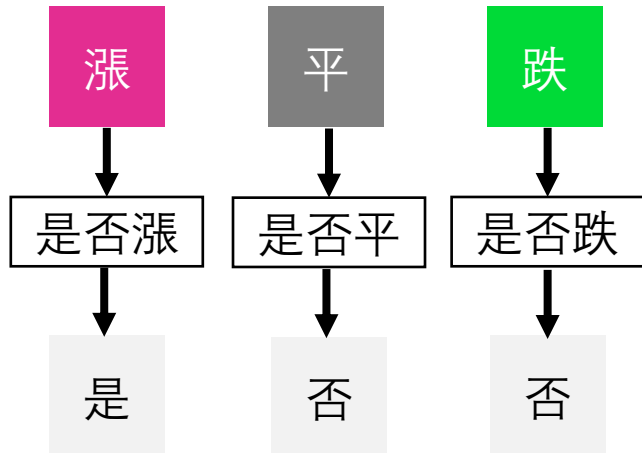


# Multiple Classes

## One against one



## One against all



通常使用scikit-learn或是一般可以找到的SVM 的 package時，都是使用 one against one，原因是因為在使用one against one時資料比較balance



# SVR: Support vector Regressor

□ 有了SVC，我們可以分類上漲和下跌的股票

□ 上漲： $a \text{ 股東權益報酬率} + b \text{ 市值} - c > 0$

□ 下跌： $a \text{ 股東權益報酬率} + b \text{ 市值} - c < 0$

□ 使用SVR，我們想要看上漲或下跌多少：

□ 漲跌數值 =  $a \text{ 股東權益報酬率} + b \text{ 市值} - c$



# 要記得做 Preprocessing

- ❑ 市值跟股東權益報酬率數值分布非常不同，會影響預測結果！
- ❑ 使用SVM時，建議調整 features 到  $[-1, 1]$  之間喔！



# SVM 關鍵名詞解釋

- ❑ Classification problem: SVC
- ❑ Regression problem: SVR
- ❑ Hyperplane
- ❑ Regularization parameter 將Hyperplane做優化，  
懲罰函數C，適當懲罰模型的錯誤
- ❑ Kernel functions
- ❑ One against one
- ❑ 自學：



# 什麼時候用SVM最好？

- ❑ Feature多 且 training data 多（例如：一百萬個）
- ❑ Features 的值有很多是零（sparsity is high）



# SVM 相關資源

## □ 台大資工中文教學

<http://www.cmlab.csie.ntu.edu.tw/~cyy/learning/tutorials/SVM2.pdf>

## □ Scikit Learn 教學

<https://scikit-learn.org/stable/modules/svm.html>

## □ 原始論文

<https://link.springer.com/content/pdf/10.1007%2FBF00994018.pdf>



謝謝您的收看

下個單元見！