

Machine Learning

Lecture 12. 推薦系統



全部

switch

您好，登入
帳戶與清單退貨
與訂單

0 購物車

交貨到
臺灣

今日優惠 客戶服務部 禮品卡 註冊 我要開店

Amazon's response to COVID-19

< 返回結果



6 影片



滑動以放大圖片

Nintendo Switch with Neon Blue and Neon Red Joy-Con - HAC-001(-01)

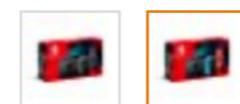
作者 [Nintendo](#)

★★★★★ 26,898等級 | 679 個問題已回覆

1 暢銷書 在 任天堂 Switch 遊戲主機

來自 這些暢銷品 有貨。

樣式: Console

[Console](#)[Console w/ 256GB Memory card](#)顏色: **Neon Blue and Red HAC-001(-01)**

- 3 Play Styles: TV Mode, Tabletop Mode, Handheld Mode
- 6.2-inch, multi-touch capacitive touch screen
- 4.5-9+ Hours of Battery Life *Will vary depending on software usage conditions
- Connects over Wi-Fi for multiplayer gaming; Up to 8 consoles can be connected for local wireless multiplayer
- Model number: HAC-001(-01)

全新和二手 (100), 最低 US\$359.95

◎ 交付至 臺灣

[查看所有購買選項](#)[已新增至清單](#)[加入婚禮註冊表](#)分享 [Email](#) [Facebook](#) [Twitter](#) [Pinterest](#)

以禮品卡出售您的禮品

我們會購買至 \$132.41

[了解更多](#)[立即交易](#)

有要銷售嗎？

[在 Amazon 上銷售](#)

購買此商品的客戶還購買了

頁面 1 / 9



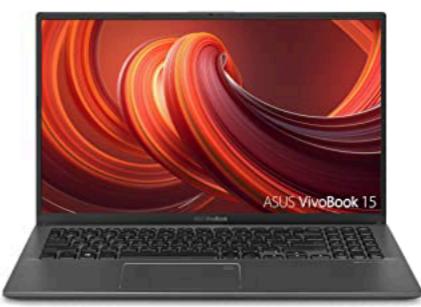
[HP Chromebook 14吋 HD 筆記型電腦, Intel Celeron N4000, 4 GB RAM, 32 GB eMMC, 銀色 \(14a...\)](#)
★★★★★ 673
共 92 筆優惠，售價為 US\$272.94 元起



[Logitech 羅技 C920x Pro HD 網路攝影機](#)
★★★★★ 19,653
共 74 筆優惠，售價為 US\$99.99 元起



[ASUS 華碩 VivoBook L203MA 筆記型電腦, 11.6 吋 HD 顯示器, Intel Celeron 雙核心 CPU, 4GB...](#)
★★★★★ 2,516
US\$287.89
庫存僅剩 3 - 快下單。



[ASUS 華碩 VivoBook 15 輕薄筆電, 15.6 吋 FHD 顯示器, Intel i3-1005G1 CPU, 8GB RAM, 128GB...](#)
★★★★★ 1,953
US\$479.00
庫存僅剩 1 - 快下單。



[Logitech 羅技 HD Pro 網路攝影機 C920 1080p 寬屏視訊和錄影](#)
★★★★★ 3,913
共 36 筆優惠，售價為 US\$99.99 元起



瀏覽本商品的客戶也瀏覽了

頁面 1 / 6



[Nintendo Switch Lite - Turquoise](#)
Nintendo
★★★★★ 1,532
Nintendo Switch
US\$289.00
庫存僅剩 1 - 快下單。



[Nintendo Switch \(Neon Red/Neon blue\)](#)
Nintendo
★★★★★ 2,267
Nintendo Switch
US\$295.13



[Nintendo Switch Console Gray Joy-Con](#)
Nintendo
★★★★★ 1,169
Nintendo Switch
US\$488.00
庫存僅剩 5 - 快下單。



[《精靈寶可夢：Let's Go》](#)
Nintendo
★★★★★ 975
Nintendo Switch
US\$659.00



[Nintendo Switch w/ Gray Joy-Con + Mario Kart 8...](#)
Nintendo
★★★★★ 2,317
Nintendo Switch
共 17 筆優惠，售價為 US\$450.00 元起



Why?

- 使用者角度
- 公司角度

2006



2/3 of the movies
watched are
recommended

2003



Google news
recommendations
generate 38% more
click-throughs

1998

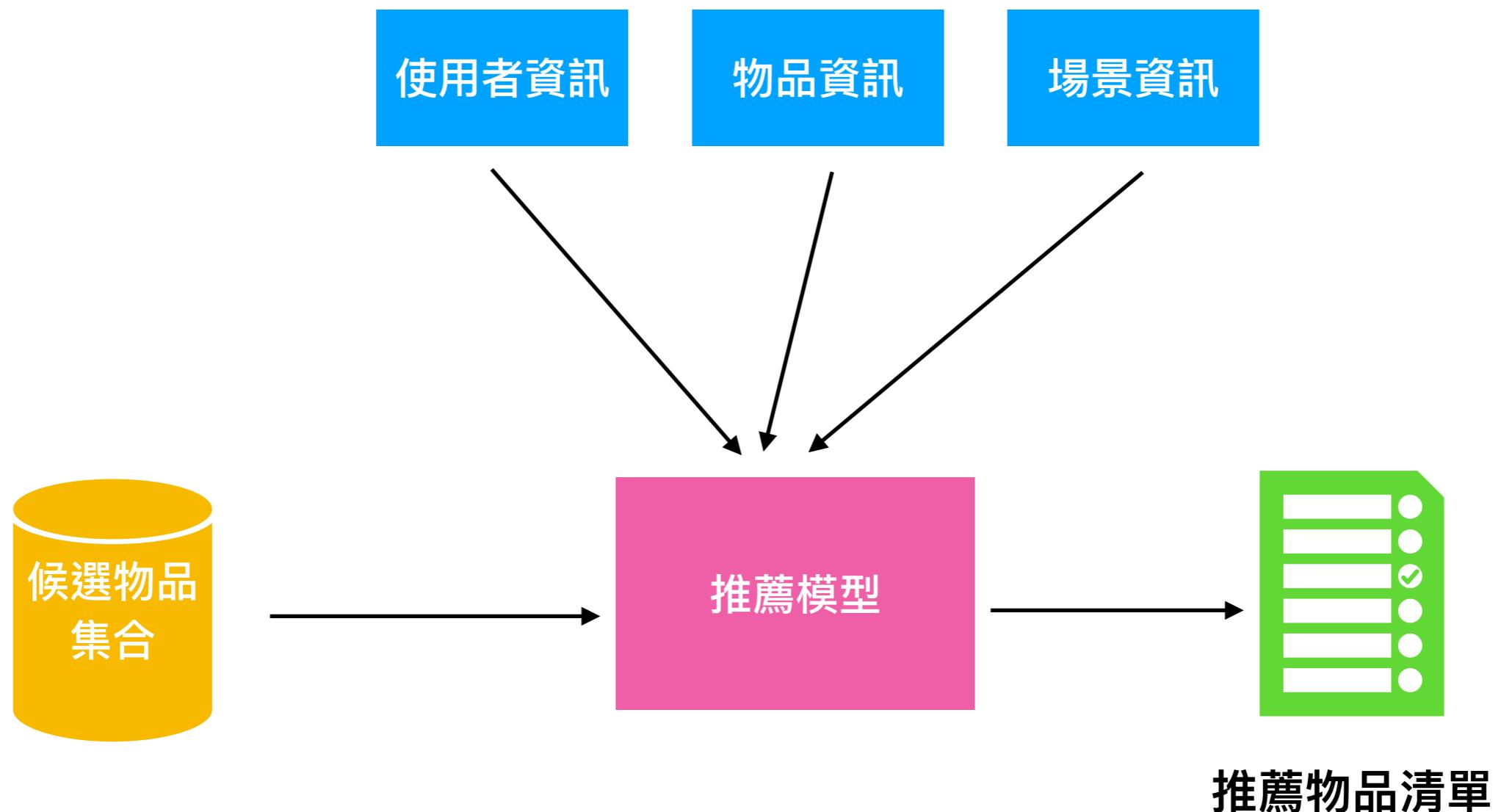


35% sales from
recommendations

Applications

- 電子商務網站，如 Amazon
 - 電影和視訊網站，如 Netflix, YouTube
 - 社群網站，如 Facebook
 - 個性化廣告，如 Google
- •
•
•

推薦系統的架構

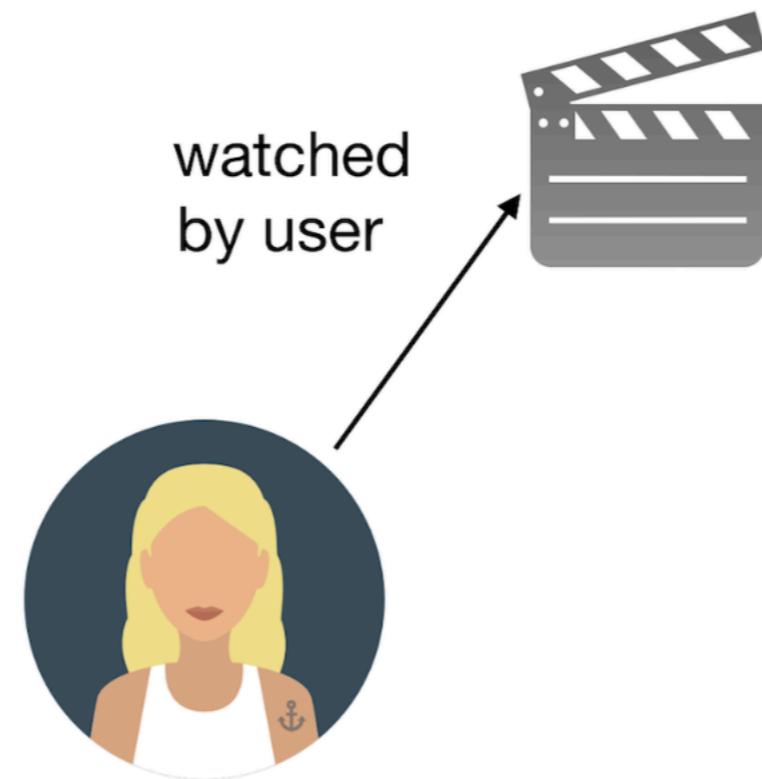


內容過濾
Content-based

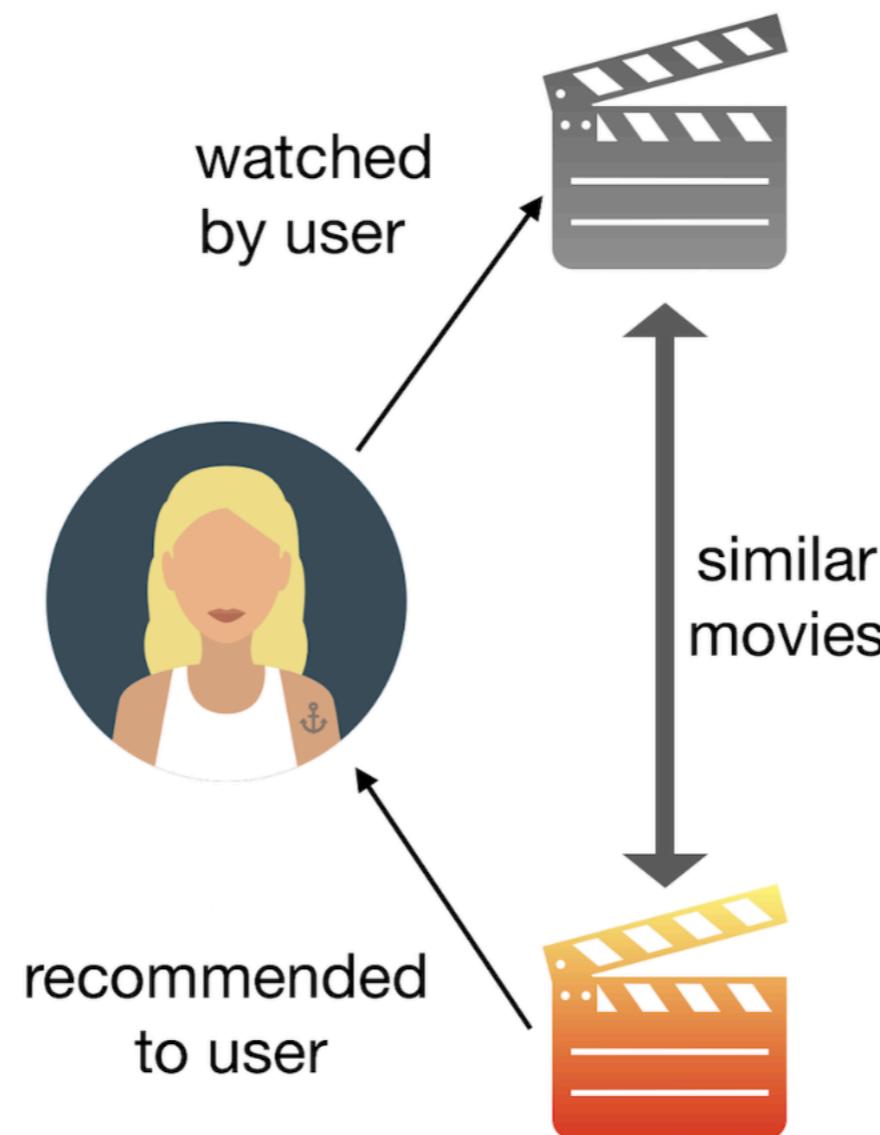
推薦系統
Recommendation
System

協同過濾
Collaborative
Filtering

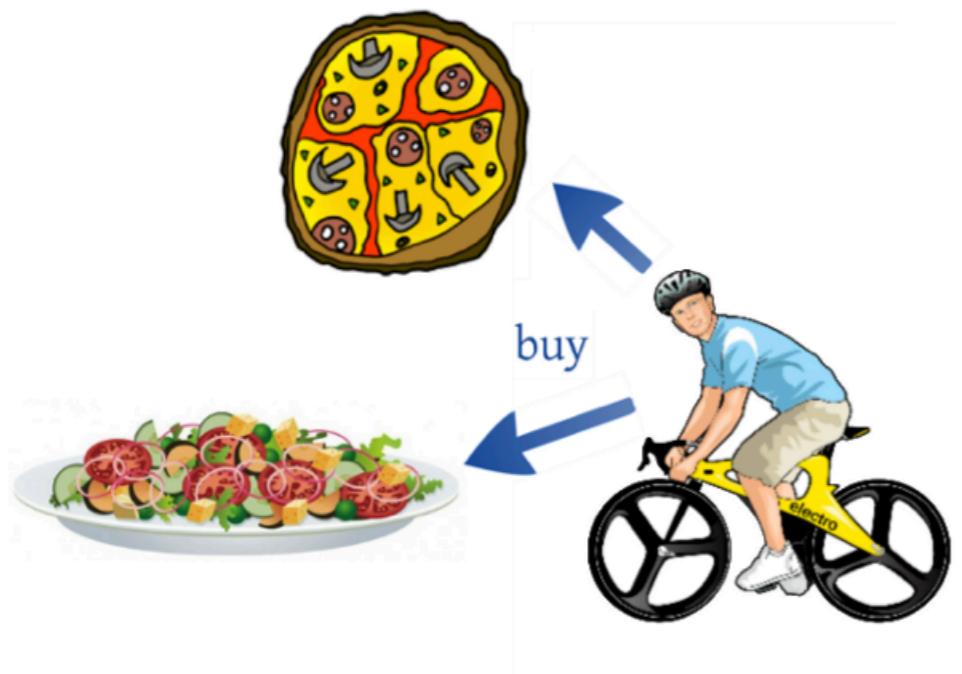
內容過濾 (Content-based filtering)



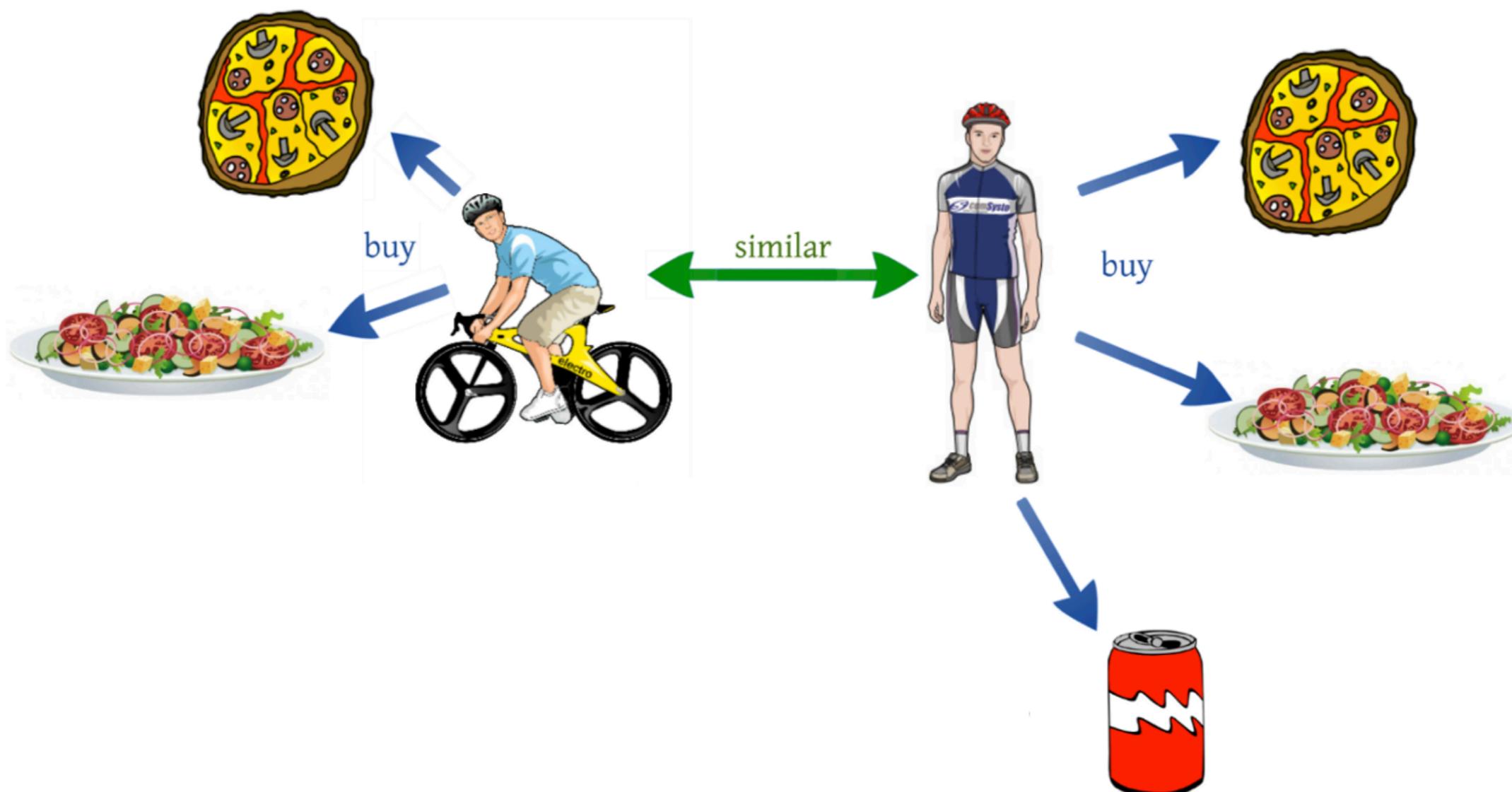
內容過濾 (Content-based filtering)



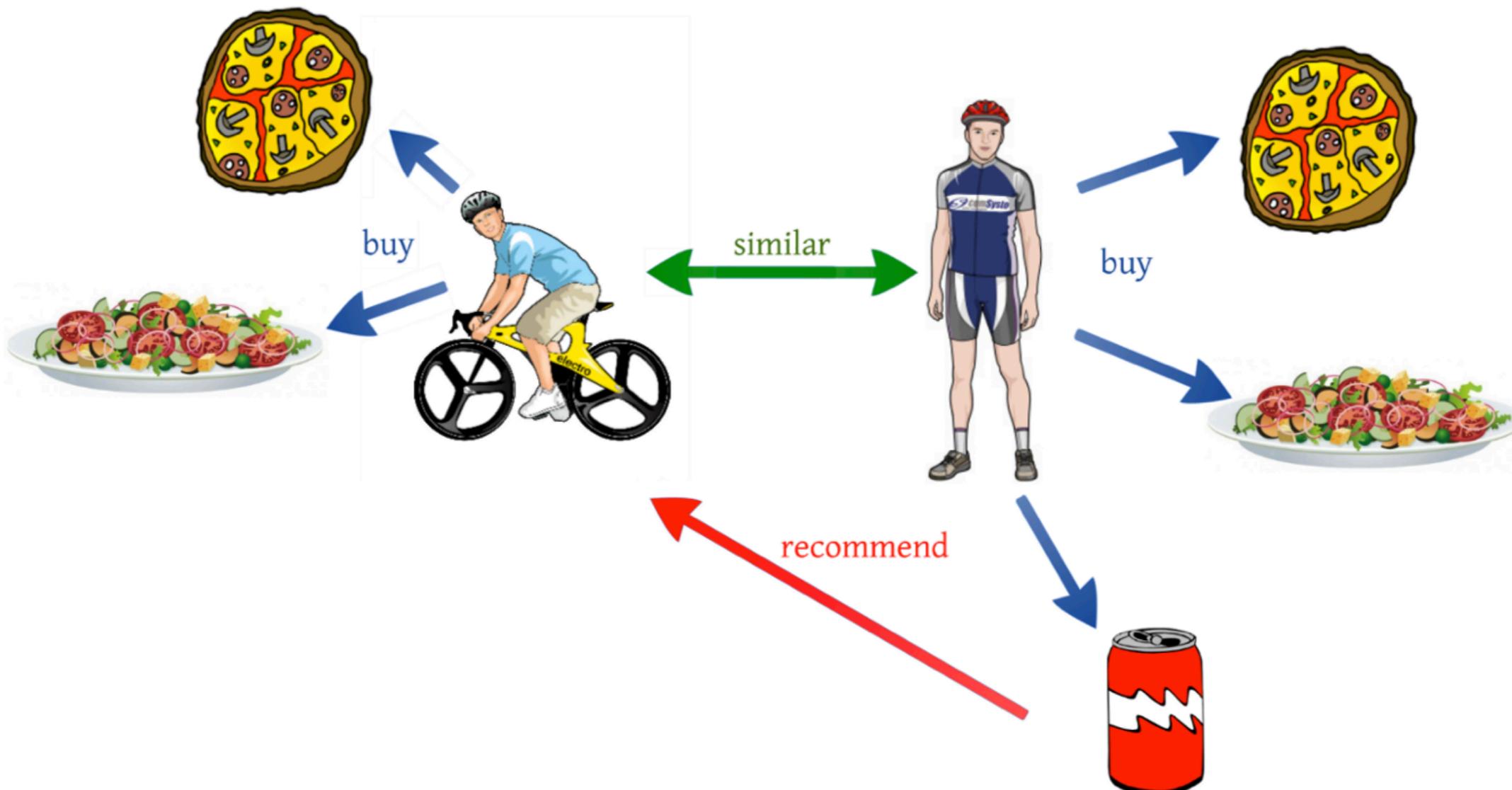
協同過濾 (Collaborative filtering)



協同過濾 (Collaborative filtering)



協同過濾 (Collaborative filtering)



推薦系統

- Content-based filtering
- Collaborative filtering
- Hybrid recommendation
- Other solutions

協同過濾 (Collaborative filtering)

協同過濾

- Input: 用戶商品評分矩陣 (large and sparse)
- Output :
 - 評分預測
 - Top-k 商品 (或用戶)

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	7	6	7	4	5
User 2	6	7	?	4	3
User 3	?	3	3	1	1
User 4	1	2	2	3	3
User 5	1	?	1	2	3

Rating

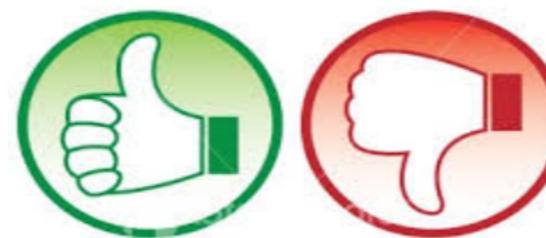
- Interval-based ratings
- Ordinal ratings
- Binary ratings
- Unary ratings



Recommender systems the textbook by Charu C. Aggarwal



Like



Feedback

顯式評分矩陣

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS	SCARFACE	SPARTACUS
U_1	1			5		2
U_2		5			4	
U_3	5	3		1		
U_4			3			4
U_5				3	5	
U_6	5		4			

(a) Ordered ratings

Feedback

客戶的行為，
如購買、瀏覽

顯式評分矩陣

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS	SCARFACE	SPARTACUS
U_1	1			5		2
U_2		5			4	
U_3	5	3		1		
U_4			3			4
U_5				3	5	
U_6	5		4			

(a) Ordered ratings

隱式評分矩陣

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS	SCARFACE	SPARTACUS
U_1	1			1		1
U_2		1			1	
U_3	1	1			1	
U_4			1			1
U_5				1	1	
U_6	1		1			

(b) Unary ratings

Collaborative Filtering

- 基於鄰里的協同過濾（Neighborhood-based CF），又稱為基於內存的協同過濾（Memory-based CF）
 - 基於用戶鄰里（User-based models）
 - 基於商品鄰里（Item-based models）

Neighborhood-Based Methods

User-based models



	Item 1	Item 2	Item 3	Item 4
User 1	7	6	7	4
User 2	6	7	?	4
User 3	?	3	3	1

相似用戶會顯示出相似的評分行為

Neighborhood-Based Methods

User-based models



	Item 1	Item 2	Item 3	Item 4
User 1	7	6	7	4
User 2	6	7	?	4
User 3	?	3	3	1

相似用戶會顯示出相似的評分行為

Item-based models



	Item 1	Item 2	Item 3	Item 4
User 1	7	6	7	4
User 2	6	7	?	4
User 3	?	3	3	1

相似的商品會獲得相似的評分

Collaborative Filtering

- 基於鄰里的協同過濾（Neighborhood-based CF），又稱為基於內存的協同過濾（Memory-based CF）
 - 基於用戶鄰里（User-based models）
 - 基於商品鄰里（Item-based models）
- 基於模型的協同過濾（Model-based CF）

基線 (Baseline)

- 可以用來與個性化推薦系統比較
 1. 平均評分 (global average)
 2. 該用戶 (或商品) 的平均評分
 3. 結合 該用戶的平均評分 與 其他用戶對特定商品的評分
偏差

Rating matrix

- 如何預測 (User 3, Item 1) 的分數？

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	7	6	7	4	5	4
User 2	6	7	?	4	3	4
User 3	?	3	3	1	1	?
User 4	1	2	2	3	3	4
User 5	1	?	1	2	3	3

Rating matrix

Global Average rating = 3.9

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	7	6	7	4	5	4
User 2	6	7	?	4	3	4
User 3	?	3	3	1	1	?
User 4	1	2	2	3	3	4
User 5	1	?	1	2	3	3

Rating matrix

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	平均
User 1	7	6	7	4	5	4	5.5
User 2	6	7	?	4	3	4	4.8
User 3	?	3	3	1	1	?	2
User 4	1	2	2	3	3	4	2.5
User 5	1	?	1	2	3	3	2

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

基線 (Baseline)

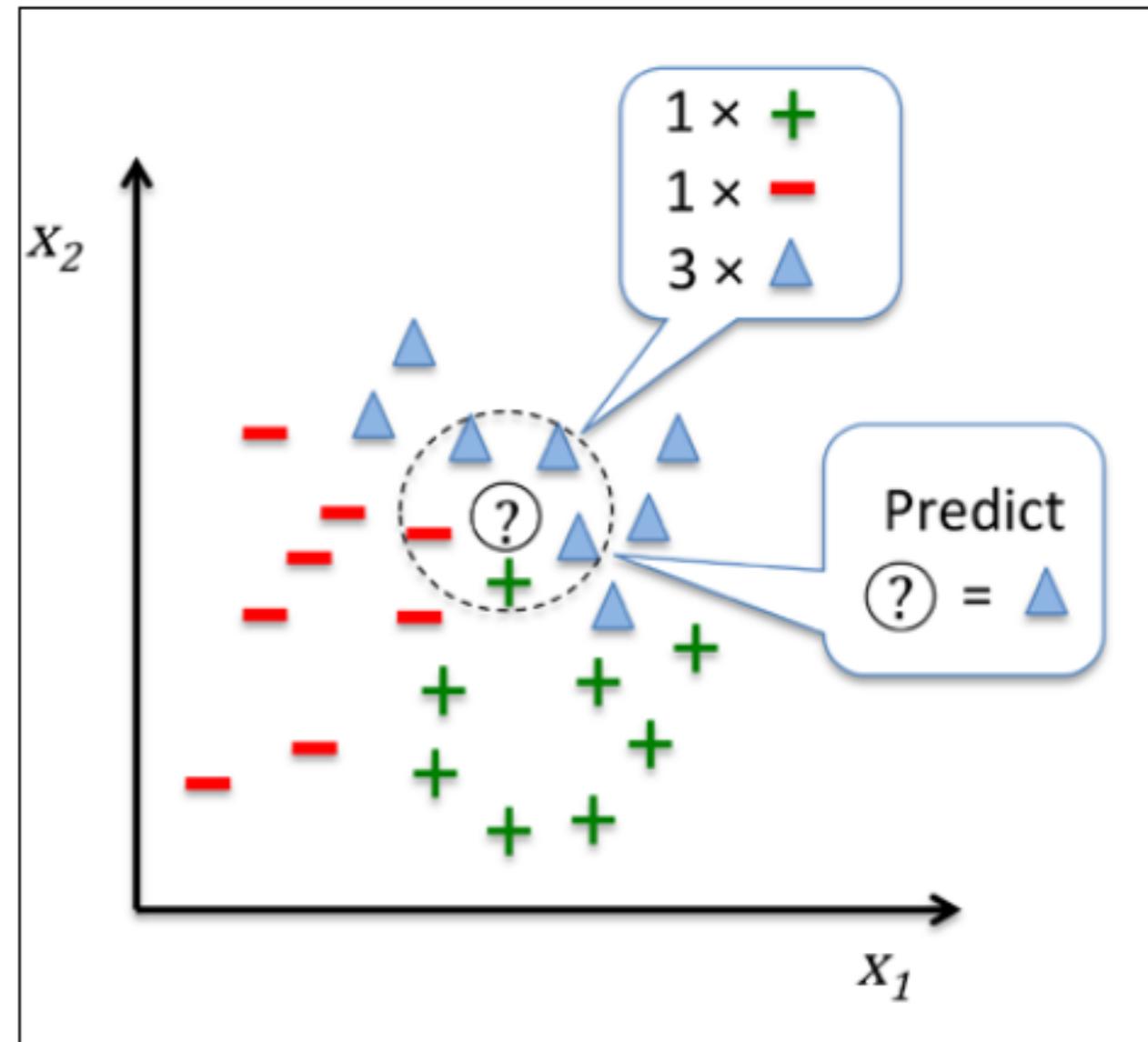
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	平均
User 1	7	6	7	4	5	4	5.5
User 2	6	7	?	4	3	4	4.8
User 3	?	3	3	1	1	?	2
User 4	1	2	2	3	3	4	2.5
User 5	1	?	1	2	3	3	2

- Item 1 的平均偏差 $= \frac{(7 - 5.5) + (6 - 4.8) + (1 - 2.5) + (1 - 2)}{4} = 0.05$
- Item 1 的基線預測 $= 2 + 0.05 = 2.05$

基於用戶鄰里的協同過濾 (User-based models)

K-NN (K-nearest neighbor classifier)

1. 確定k大小和距離度量($k = 5$)。
2. 對於測試集中的一個樣本，找到訓練集中和它最近的k個樣本。
3. 將這k個樣本的投票結果作為測試樣本的類別 (多數決)。



<https://ljalphabeta.gitbooks.io/python-/content/knn.html>

基於鄰里的協同過濾

- 相似用戶會顯示出相似的評分行為，相似的商品會獲得相似的評分
- 方法
 1. 建立用戶（或商品）的相似度矩陣
 2. 選擇相似度最高的 k 個用戶（或商品）
 3. 根據相似性計算評分的加權平均

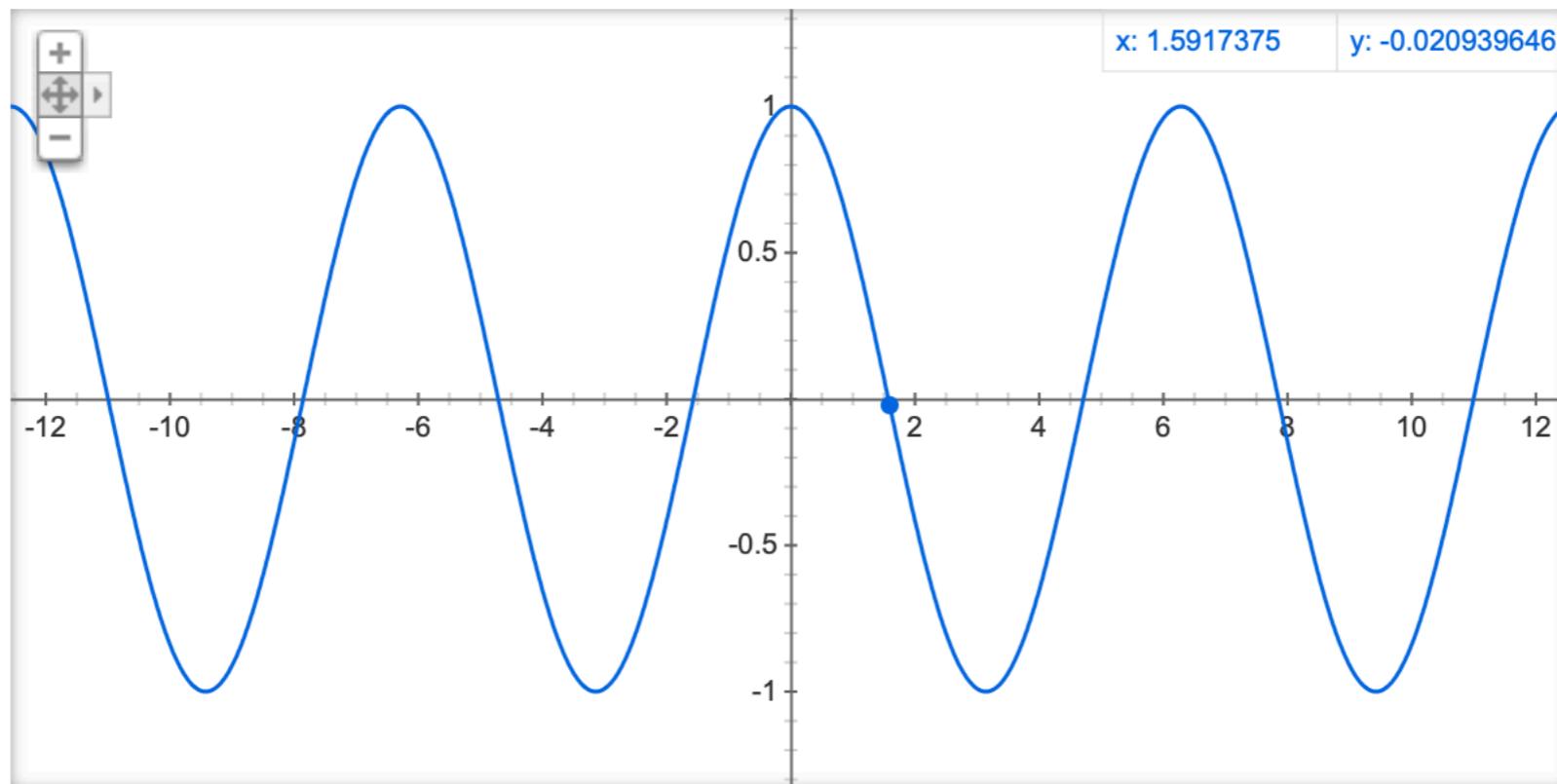
相似性

	1	2	3	4	5	6
a	+	?	-	-	?	-
b		-		+		+
c	+	+		-	-	-
d			+	+	-	
e	-		-		+	+

相似度 1：餘弦(cosine) 相似係數

$$\text{Cosine}(u, v) = \frac{\sum_{k \in I_u \cap I_v} r_{uk} \cdot r_{vk}}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk})^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk})^2}}$$

cos(x) 的圖表



Rating matrix

- 如何預測 (User 3, Item 1) 的分數？

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	7	6	7	4	5	4
User 2	6	7	?	4	3	4
User 3	?	3	3	1	1	?
User 4	1	2	2	3	3	4
User 5	1	?	1	2	3	3

相似度 1：餘弦(cosine) 相似係數

Q. 計算 User 1 和 3 的 Cosine(1, 3)

Step 1. 找出User 1 和 3 共同評分過的商品集
 $\{2, 3, 4, 5\}$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	7	6	7	4	5	4
User 2	6	7	?	4	3	4
User 3	?	3	3	1	1	?
User 4	1	2	2	3	3	4
User 5	1	?	1	2	3	3

相似度 1：餘弦(cosine) 相似係數

Q. 計算 User 1 和 3 的 Cosine(1, 3)

Step 2. 利用這些商品集計算 Cosine(1, 3)

$$\text{Cosine}(1,3) = \frac{6 \times 3 + 7 \times 3 + 4 \times 1 + 5 \times 1}{\sqrt{6^2 + 7^2 + 4^2 + 5^2} \cdot \sqrt{3^2 + 3^2 + 1^2 + 1^2}} = 0.956$$

	Item 2	Item 3	Item 4	Item 5
User 1	6	7	4	5
User 2				
User 3	3	3	1	1

Rating matrix

假設 $k = 2$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Cosine($i, 3$)
User 1	7	6	7	4	5	4	0.956
User 2	6	7	?	4	3	4	0.981
User 3	?	3	3	1	1	?	1.0
User 4	1	2	2	3	3	4	0.789
User 5	1	?	1	2	3	3	0.645

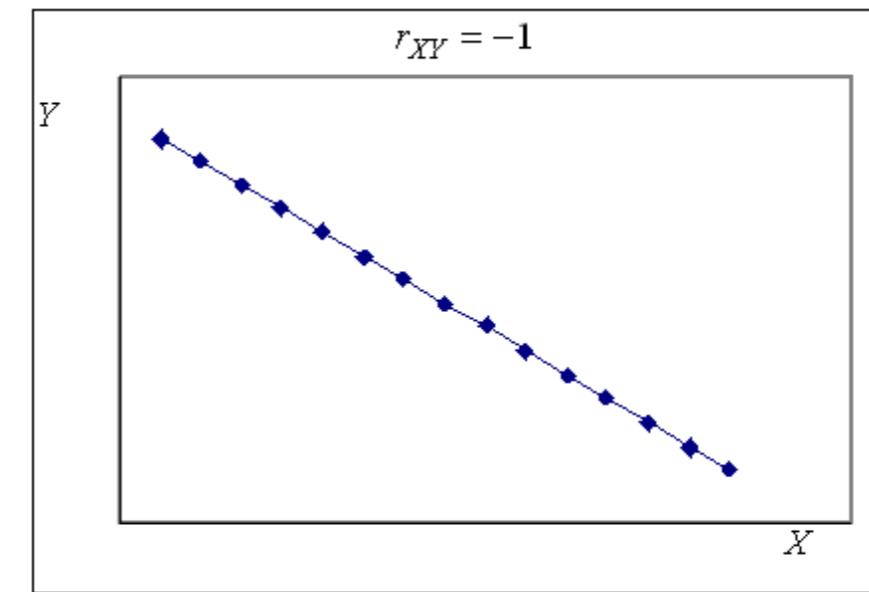
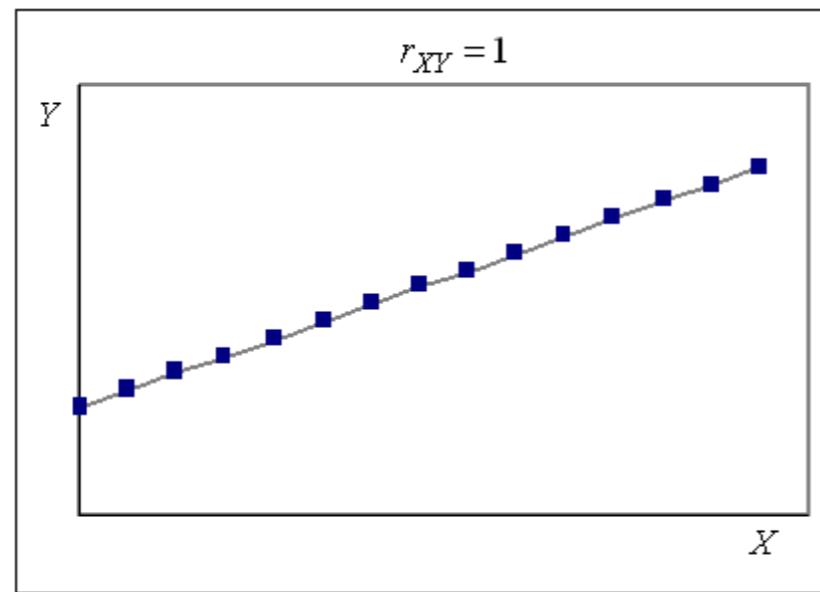
預測

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Cosine($i, 3$)
User 1	7	6	7	4	5	4	0.956
User 2	6	7	?	4	3	4	0.981
User 3	?	3	3	1	1	?	1.0

$$\hat{r}_{31} = \frac{7 \times 0.956 + 6 \times 0.981}{0.956 + 0.981} \approx 6.49$$

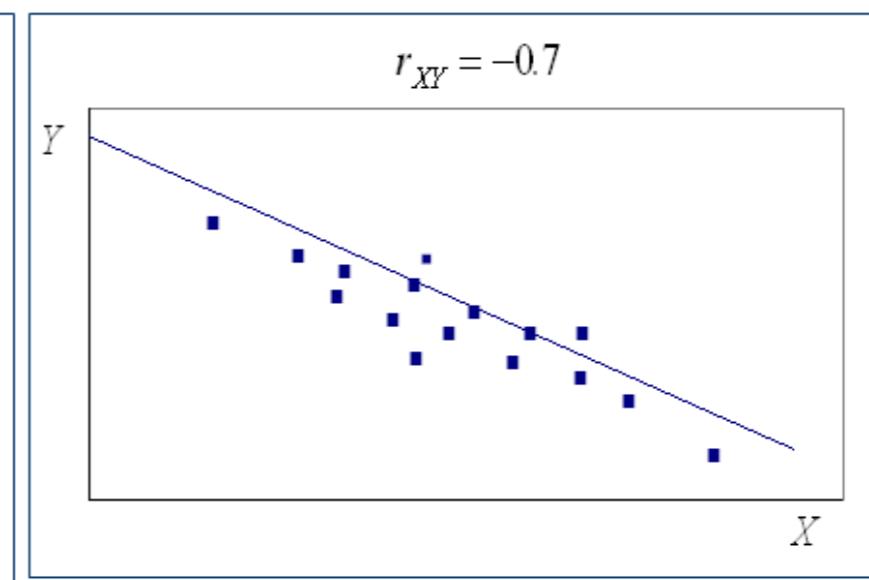
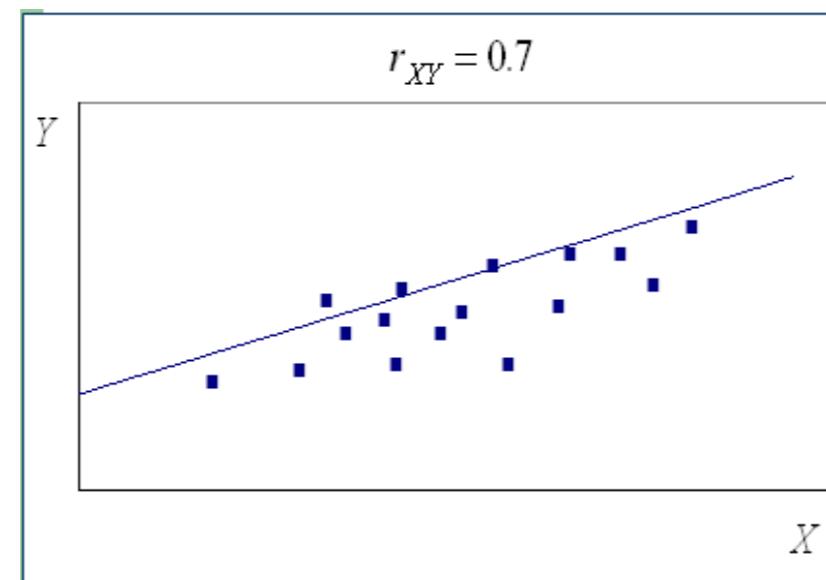
相似度 2：Pearson 相關係數

$$\text{Sim}(u, v) = \text{Pearson}(u, v) = \frac{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u) \cdot (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}}$$



正相關

負相關



相似度 2：Pearson 相關係數

Q. 計算 User 1 和 3 的 Pearson 相關係數 Pearson(1, 3)

Step 1. 找出User 1 和 3 共同評分過的商品集
 $\{2, 3, 4, 5\}$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	平均
User 1	7	6	7	4	5	4	5.5
User 2	6	7	?	4	3	4	4.8
User 3	?	3	3	1	1	?	2
User 4	1	2	2	3	3	4	2.5
User 5	1	?	1	2	3	3	2

相似度 2：Pearson 相關係數

Q. 計算 User 1 和 3 的 Pearson 相關係數 Pearson(1, 3)

Step 2. 利用這些商品集計算 Pearson 相關係數

Pearson(1,3)

$$\begin{aligned} &= \frac{(6 - 5.5) \times (3 - 2) + (7 - 5.5) \times (3 - 2) + (4 - 5.5) \times (1 - 2) + (5 - 5.5) \times (1 - 2)}{\sqrt{(6 - 5.5)^2 + (7 - 5.5)^2 + (4 - 5.5)^2 + (5 - 5.5)^2} \cdot \sqrt{(3 - 2)^2 + (3 - 2)^2 + (1 - 2)^2 + (1 - 2)^2}} \\ &= 0.894 \end{aligned}$$

	Item 2	Item 3	Item 4	Item 5	平均
User 1	6	7	4	5	5.5
User 2					
User 3	3	3	1	1	2

	Item 2	Item 3	Item 4	Item 5
User 1	0.5	1.5	-1.5	-0.5
User 2				
User 3	1	1	-1	-1

Rating matrix

假設 $k = 2$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Pearson($i, 3$)
User 1	7	6	7	4	5	4	0.894
User 2	6	7	?	4	3	4	0.939
User 3	?	3	3	1	1	?	1.0
User 4	1	2	2	3	3	4	-1.0
User 5	1	?	1	2	3	3	-0.817

預測

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	平均	Pearson($i, 3$)
User 1	7	6	7	4	5	4	5.5	0.894
User 2	6	7	?	4	3	4	4.8	0.939
User 3	?	3	3	1	1	?	2	1.0

$$\hat{r}_{uj} = \mu_u + \frac{\sum_{v \in P_u(j)} \mathbf{Sim}(u, v) \cdot (r_{vj} - \mu_v)}{\sum_{v \in P_u(j)} |\mathbf{Sim}(u, v)|}$$

$$\hat{r}_{31} = 2 + \frac{1.5 \times 0.894 + 1.2 \times 0.939}{0.894 + 0.939} \approx 3.35$$

$$\hat{r}_{36} = 2 + \frac{-1.5 \times 0.894 + 0.8 \times 0.939}{0.894 + 0.939} \approx 0.86$$

基於物品鄰里的協同過濾 (Item-based models)

Neighborhood-Based Methods

User-based models



	Item 1	Item 2	Item 3	Item 4
User 1	7	6	7	4
User 2	6	7	?	4
User 3	?	3	3	1

相似用戶會顯示出相似的評分行為

Item-based models



	Item 1	Item 2	Item 3	Item 4
User 1	7	6	7	4
User 2	6	7	?	4
User 3	?	3	3	1

相似的商品會獲得相似的評分

Rating matrix

- 如何預測 (User 3, Item 1) 的分數？

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	平均
User 1	7	6	7	4	5	4	5.5
User 2	6	7	?	4	3	4	4.8
User 3	?	3	3	1	1	?	2
User 4	1	2	2	3	3	4	2.5
User 5	1	?	1	2	3	3	2

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	1.5	0.5	1.5	-1.5	-0.5	-1.5
User 2	1.2	2.2	?	-0.8	-1.8	-0.8
User 3	?	1	1	-1	-1	?
User 4	-1.5	-0.5	-0.5	0.5	0.5	1.5
User 5	-1	?	-1	0	1	1
cosine(1, j)	1	0.735	0.912	-0.848	-0.813	-0.990

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	7	6	7	4	5	4
User 2	6	7	?	4	3	4
User 3	?	3	3	1	1	?
User 4	1	2	2	3	3	4
User 5	1	?	1	2	3	3
cosine(1, j)	1	0.735	0.912	-0.848	-0.813	-0.990

$$\hat{r}_{31} = \frac{3 \times 0.735 + 3 \times 0.912}{0.735 + 0.912} = 3$$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	1.5	0.5	1.5	-1.5	-0.5	-1.5
User 2	1.2	2.2	?	-0.8	-1.8	-0.8
User 3	?	1	1	-1	-1	?
User 4	-1.5	-0.5	-0.5	0.5	0.5	1.5
User 5	-1	?	-1	0	1	1

cosine(1, j)	1	0.735	0.912	-0.848	-0.813	-0.990
cosine(6, j)	-0.990	-0.622	-0.912	0.829	0.730	1

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	7	6	7	4	5	4
User 2	6	7	?	4	3	4
User 3	?	3	3	1	1	?
User 4	1	2	2	3	3	4
User 5	1	?	1	2	3	3
cosine(1, j)	1	0.735	0.912	-0.848	-0.813	-0.990
cosine(6, j)	-0.990	-0.622	-0.912	0.829	0.730	1

$$\hat{r}_{36} = \frac{1 \times 0.829 + 1 \times 0.730}{0.829 + 0.730} = 1$$

	基於用戶	基於商品
準確性		通常較好，因為使用自己的評分
多樣性、新穎性、驚奇性 (Diversity, Novelty, Serendipity)	通常較好，因為使用其他用戶的評分	
解釋性		通常可以提供具體的解釋
穩定性		通常較好，因為通常用戶比商品多得多，所以每個商品有較多的評分

實作

- m users, n items (通常 $m \gg n$)
- 離線階段 (offline phase)
 - 計算相似度

	user 1	...	user m
user 1			
:			
user m			

user-based: m^2

	item 1	...	item n
item 1			
:			
item n			

item-based: n^2

- 在線階段 (online phase)
 - 利用 Top - k 計算評分

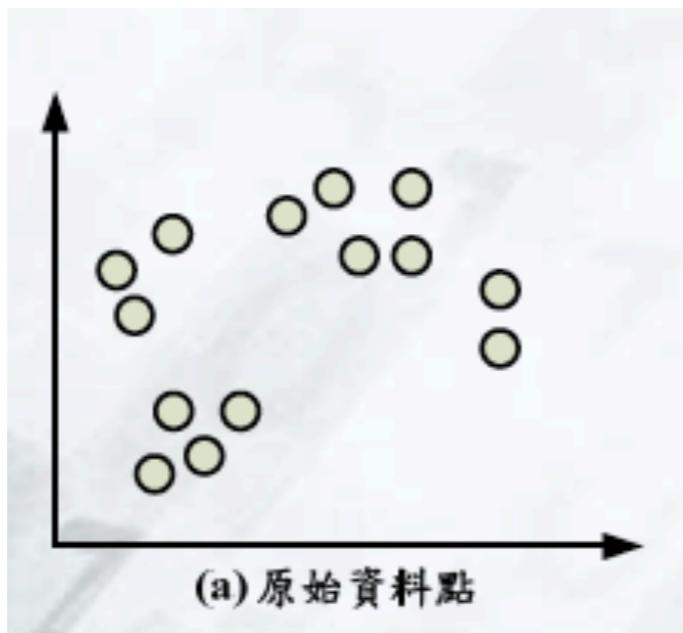
基於鄰里的方法

優點	缺點
簡單、直覺	離線階段在大規模環境中有時 不太實際 (m^2)
易於實現和調試	評分矩陣的稀疏性
易於解釋為何一個商品被推薦	冷啟動問題
對增加新的用戶和商品時，表現相對穩定	

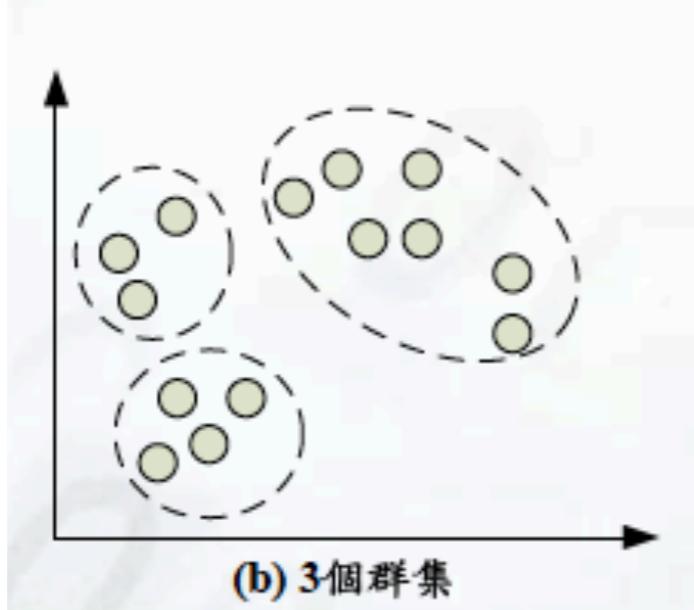
聚類和基於鄰里的方法

- 利用離線聚類階段（offline clustering phase）取代離線最近鄰居階段

鄰居



聚類

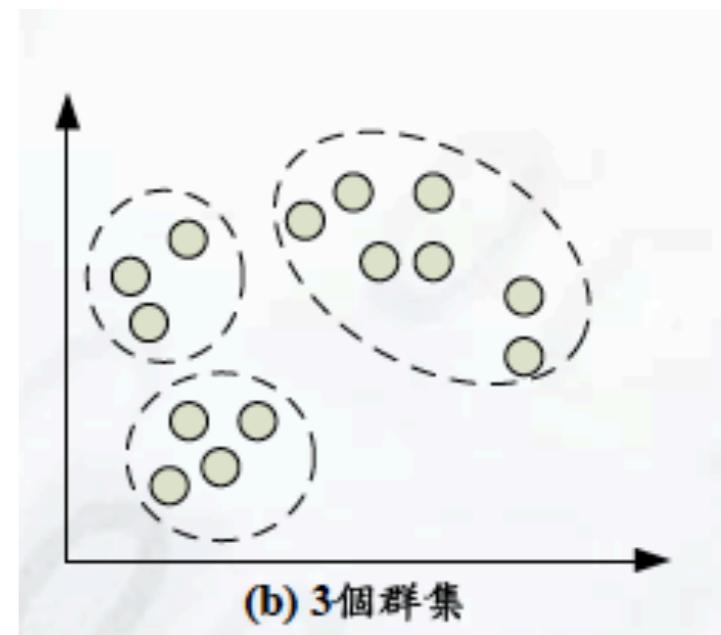


	user 1	...	user m
user 1			
:			
user m			

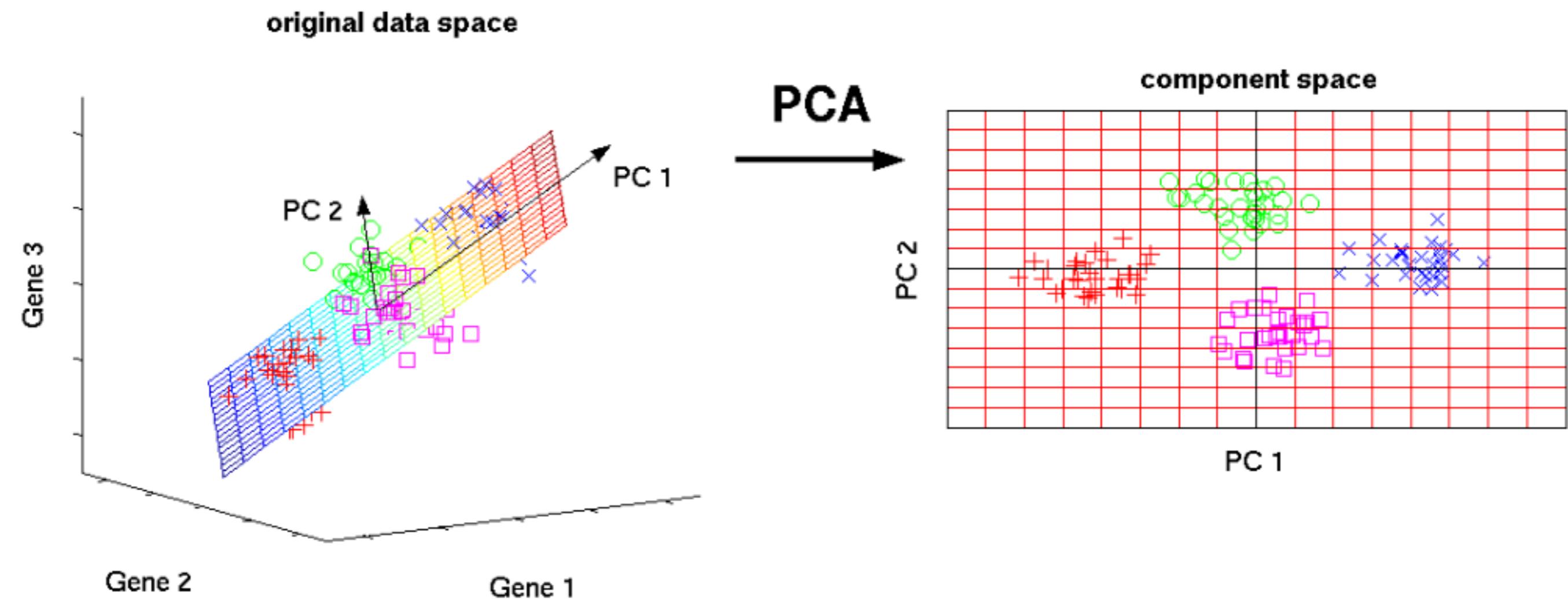
	cluster 1	cluster 2	cluster 3
cluster 1			
cluster 2			
cluster 3			

聚類和基於鄰里的方法

- 利用離線聚類階段（offline clustering phase）取代離線最近鄰居階段
- 計算時間及儲存空間較有效率
- 在線階段預測評分時，只使用同一個聚類裡的 Top-k
- 準確性稍差，但效率大幅提高



維度縮減和基於鄰里的方法



潛在因子 (Latent Factor)

	不可能的任務	牠	哭聲	BJ 單身日記	即刻救援
Alice	2	4	3	5	2
Bob	4	4	5	1	3
Tom	2	5	5	2	3



	動作	恐怖	愛情
Alice	0.1	0.3	0.6
Bob	0.4	0.5	0.1
Tom	0.2	0.7	0.1

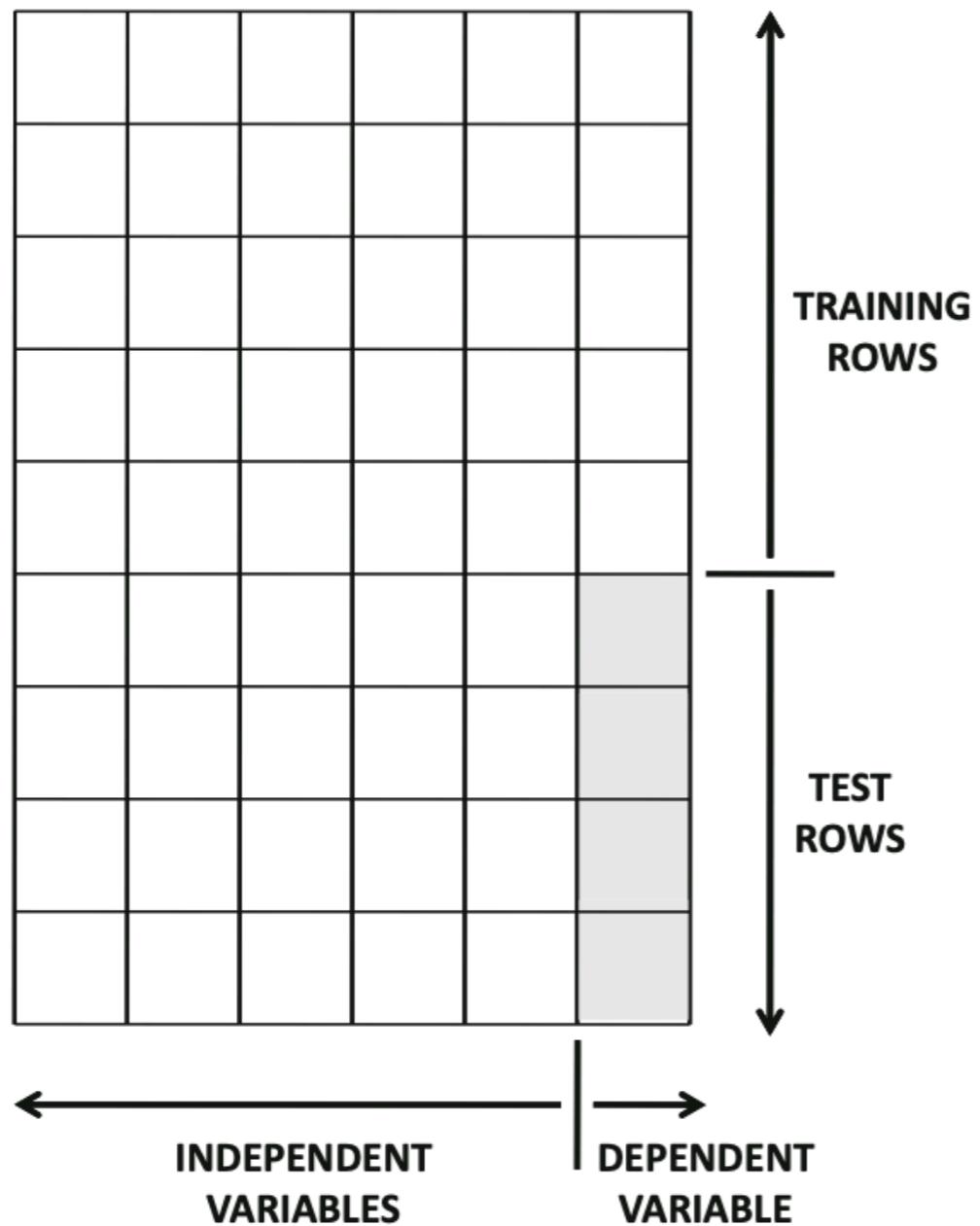
基於模型的協同過濾

(Model-based CF)

基於模型的協同過濾

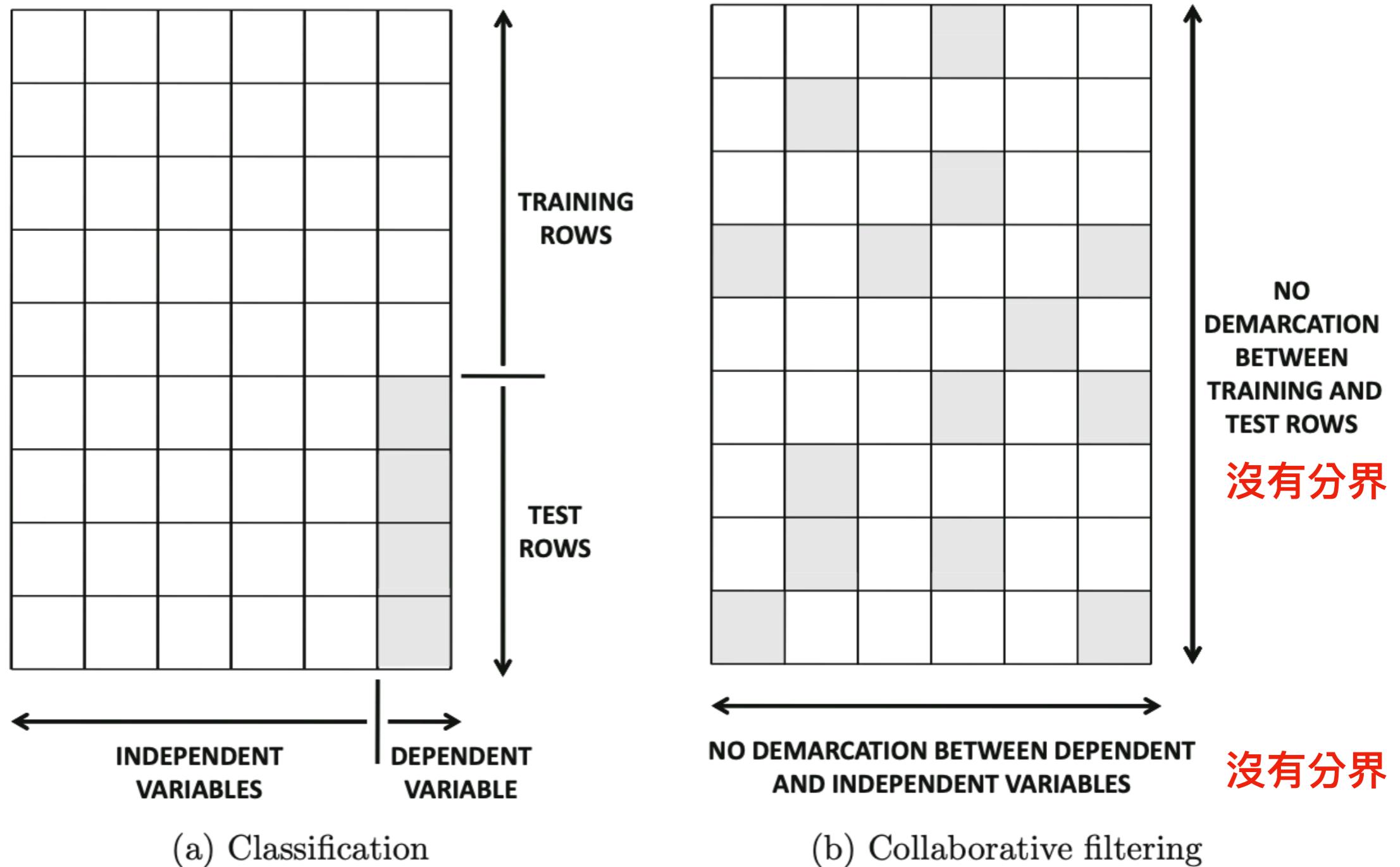
- 關聯規則
- 決策樹
- 貝葉斯 (Bayesian) 分類
- 支持向量機 (SVM)
- 類神經網絡
-
-
-

分類與協同過濾



(a) Classification

分類與協同過濾



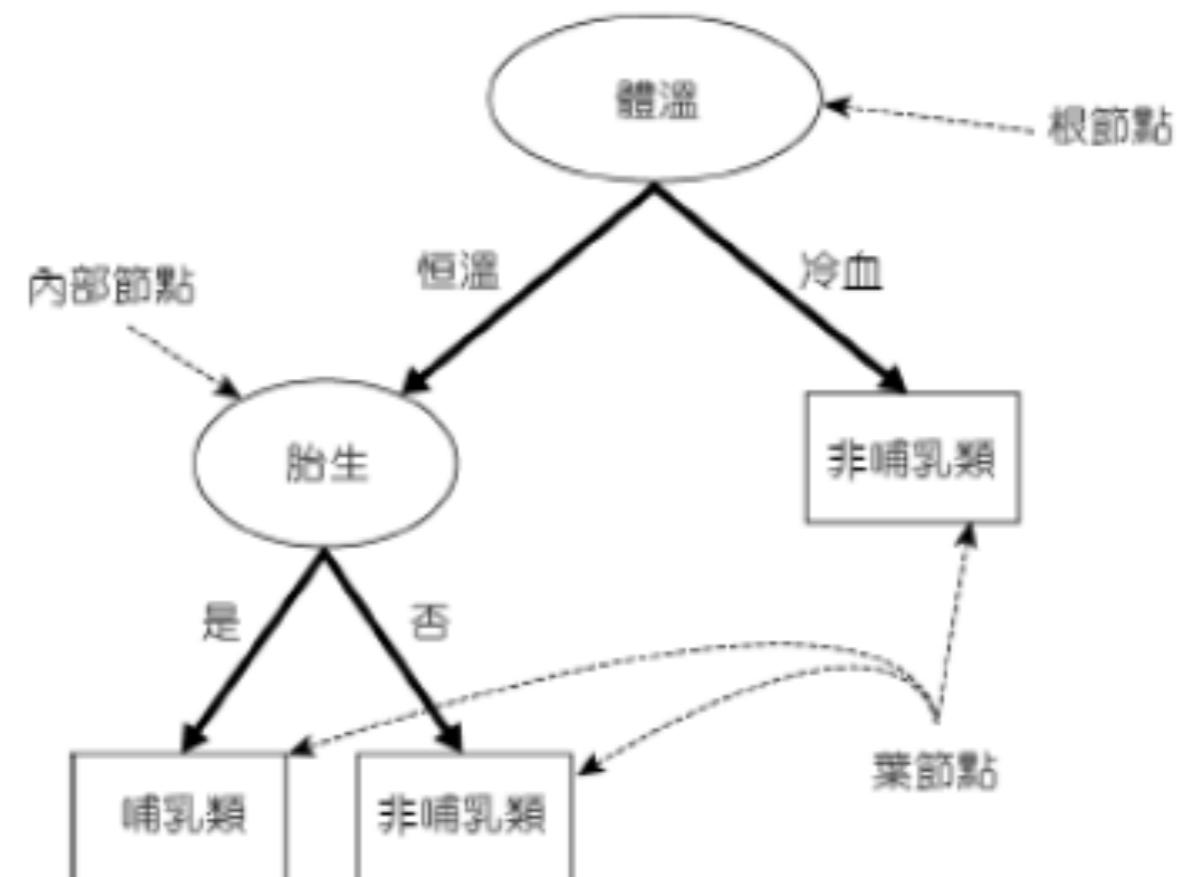
決策樹

● 什麼是決策樹(Decision Tree)？

- 用來處理分類問題的**樹狀**結構
- 每個內部節點表示一個**評估欄位**
- 每個分枝代表一個可能的**欄位輸出結果**
- 每個樹葉節點代表不同分類的**類別標記**

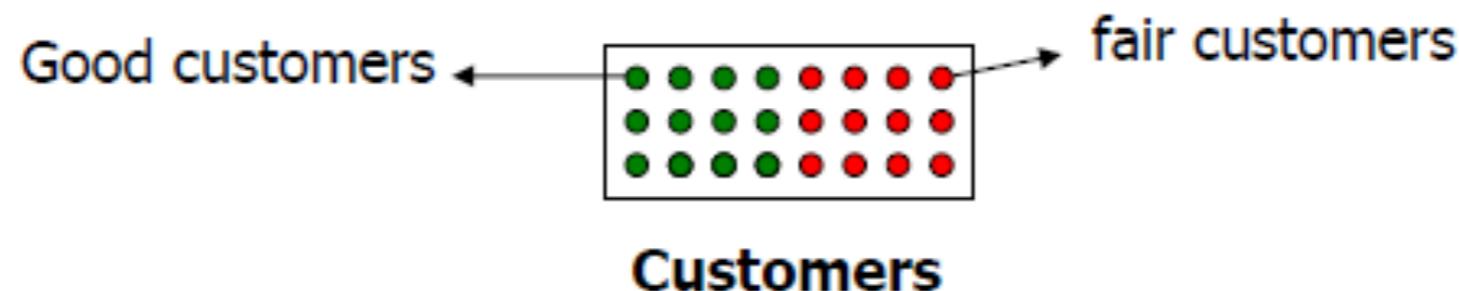
● 決策樹範例

- 哺乳類動物分類的問題
(鴨嘴獸和食蟻獸不適用!!)

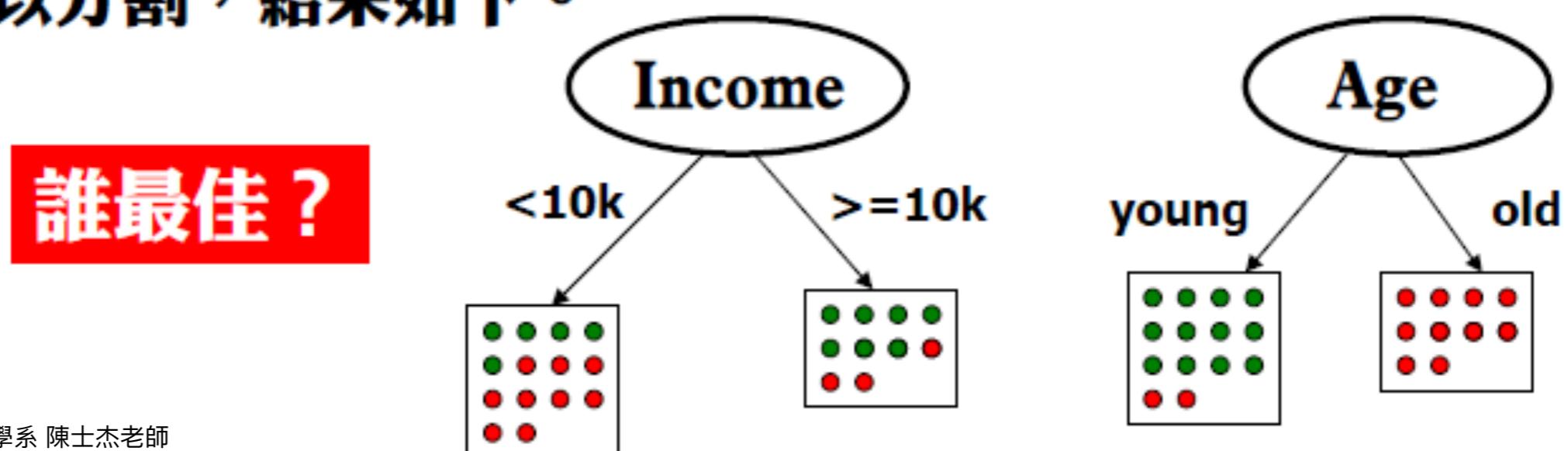


決策樹

- 假設有一個表格共有24筆顧客資料。其類別欄位為“Customers”，可分成“好客人 Good Customers”與“一般客人 Fair Customer”兩類。

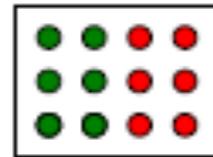


- 分別用Income和Age兩個欄位，對這24筆顧客資料加以分割，結果如下。



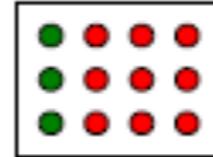
決策樹

- 如何決定分割結果何者較佳：
 - 分割結果中，若具有較高**同質性**(Homogeneous)類別的節點，則該分割結果愈佳。
- 因此，需要檢驗節點的**不純度**(Node Impurity):
 - 不純度愈低愈好。



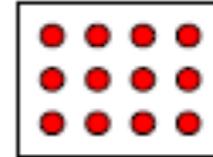
50% red
50% green

High degree
of impurity



75% red
25% green

Low degree
of impurity



100% red
0% green

pure

常用的屬性選擇指標

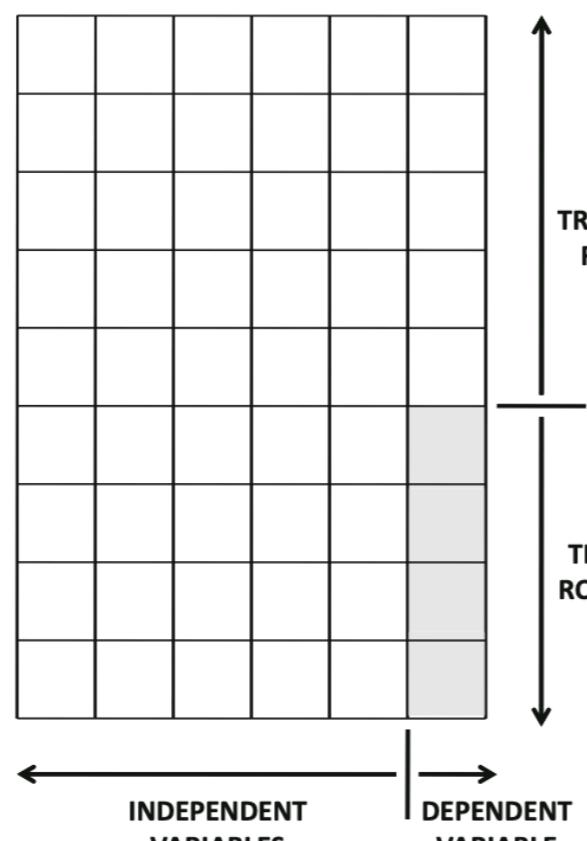
- 資訊獲利 (Information Gain) : ID3
- 獲利比例 (Gain ratio) : C4.5
- 吉尼係數 (Gini Index) : CART

建立決策樹

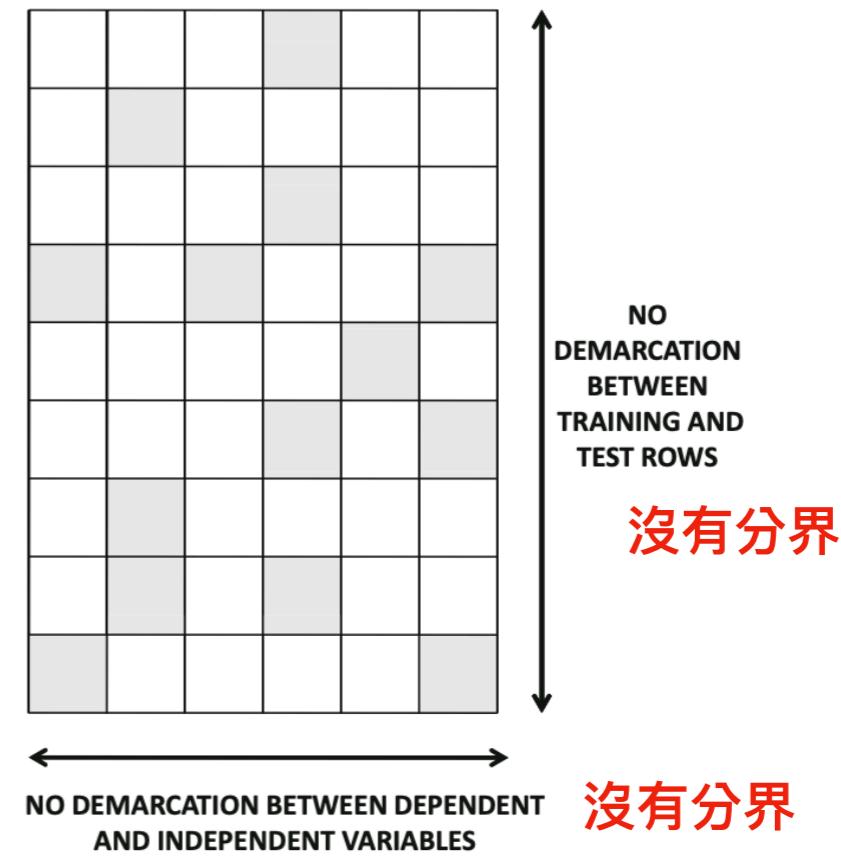
- Step 1: 找出最能將資料點均勻區分的問題作為樹的內部節點，並將節點分割以產生對應的分支
- Step 2: 在每一個葉節點重複Step 1，直到達到停止條件

Extending Decision Trees to Collaborative Filtering

1. The predicted entries and the observed entries are not clearly separated as feature and class variables.
2. The rating matrix is very sparsely.



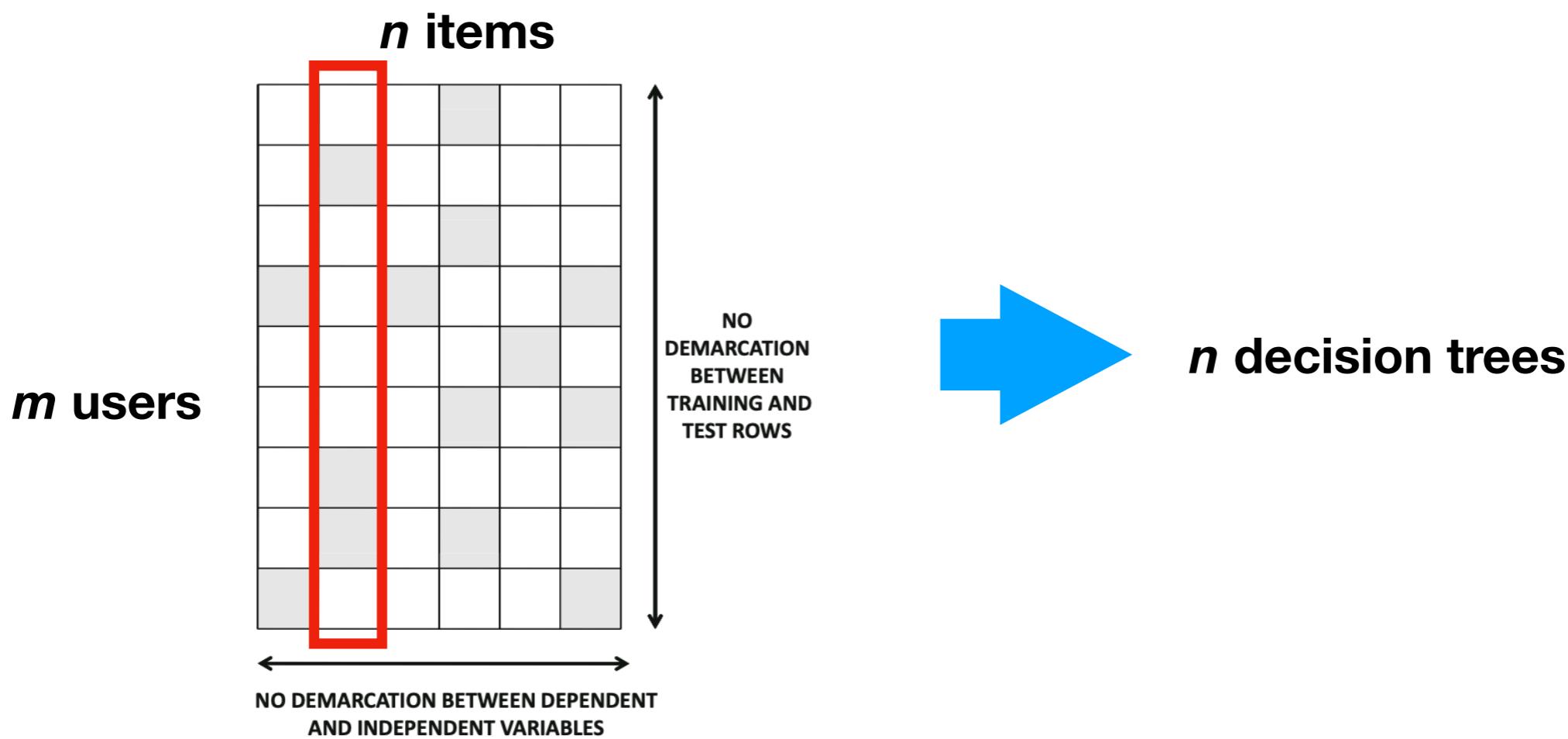
(a) Classification



(b) Collaborative filtering

Extending Decision Trees to Collaborative Filtering

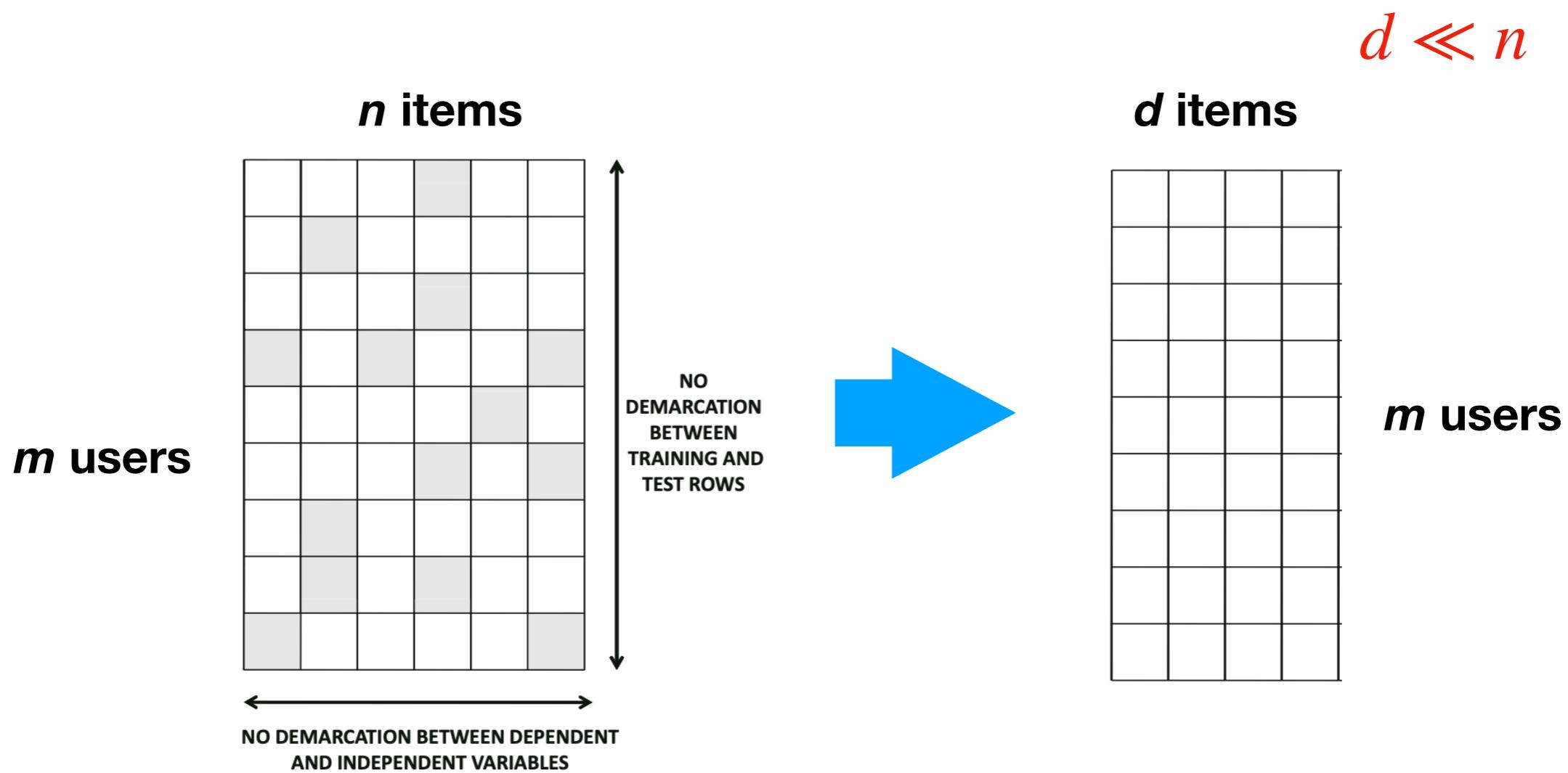
1. The predicted entries and the observed entries are not clearly separated as feature and class variables.
 - ★ constructing separate decision trees to predict the rating of each item.



Extending Decision Trees to Collaborative Filtering

2. The rating matrix is very sparsely..

★ creating a lower-dimensional representation of the data with fully populated ratings.



貝葉斯分類

- 基本想法
 - 給定用戶 / 商品 評分矩陣
 - 基於用戶 Alice 的歷史評分，計算她喜歡一個特定商品的機率
 - 利用這些機率提出推薦

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

問題：天氣評估系統

	H ₁	H ₂	H ₃	H ₄	B
Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

{H₁ = Sunny, H₂ = Cool, H₃ = High, H₄ = Strong, B = ?}

問題：天氣評估系統

	H₁	H₂	H₃	H₄	B
Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Let **H = (H₁ ∩ H₂ ∩ H₃ ∩ H₄)**

P(B=Yes | H) v.s. P(B=No | H)

Example

$$P(B = Yes | H) = P(B = Yes) \prod_{i=1}^4 P(H_i | B = Yes)$$

$$P(B = No | H) = P(B = No) \prod_{i=1}^4 P(H_i | B = No)$$

Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	2/9	4/9	3/9
No	3/5	0/5	2/5

P(Sunny|Yes)

	Temp.		
	Hot	Mild	Cool
Yes	2	4	3
No	2	2	1
Yes	2/9	4/9	3/9
No	2/5	2/5	1/5

	Humidity			Wind	
	High	Normal		Weak	Strong
Yes	3	6	Yes	6	3
No	4	1	No	2	3
Yes	3/9	6/9	Yes	6/9	3/9
No	4/5	1/5	No	2/5	3/5

Example

	Item1	Item2	Item3	Item4	Item5
Alice	1	3	3	2	?
User1	2	4	2	2	4
User2	1	3	3	5	1
User3	4	5	2	3	3
User4	1	1	5	2	1

$X = (\text{Item1} = 1, \text{Item2} = 3, \text{Item3} = \dots)$

$$P(X|Item5 = 1)$$

$$\begin{aligned}
 &= P(\text{Item1} = 1 | Item5 = 1) \times P(\text{Item2} = 3 | Item5 = 1) \\
 &\times P(\text{Item3} = 3 | Item5 = 1) \times P(\text{Item4} = 2 | Item5 = 1) \\
 &= \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \approx 0.125
 \end{aligned}$$

Example

	Item1	Item2	Item3	Item4	Item5
Alice	1	3	3	2	?
User1	2	4	2	2	4
User2	1	3	3	5	1
User3	4	5	2	3	3
User4	1	1	5	2	1

$X = (\text{Item1} = 1, \text{Item2} = 3, \text{Item3} = \dots)$

$$P(X|Item5 = 1)$$

$$\begin{aligned}
 &= P(\text{Item1} = 1 | Item5 = 1) \times P(\text{Item2} = 3 | Item5 = 1) \\
 &\times P(\text{Item3} = 3 | Item5 = 1) \times P(\text{Item4} = 2 | Item5 = 1) \\
 &= \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \approx 0.125
 \end{aligned}$$

$$P(X|Item5 = 2)$$

$$\begin{aligned}
 &= P(\text{Item1} = 1 | Item5 = 2) \times P(\text{Item2} = 3 | Item5 = 2) \\
 &\times P(\text{Item3} = 3 | Item5 = 2) \times P(\text{Item4} = 2 | Item5 = 2) \\
 &= \frac{0}{0} \times \dots \times \dots \times \dots = 0
 \end{aligned}$$

Laplace Estimator

- 使用Laplace Estimator評估條件機率的作法如下：
 - 分子：加上1。
 - 分母：加上q，其中 q 為該屬性內的不同資料個數。
- 在此使用Laplace Estimator，對類別為No所有Outlook情況的條件機率做修正：

		Outlook		
		Sunny	Overcast	Rain
Yes	Sunny	2	4	3
	Overcast	3	0	2
No	Sunny	2/9	4/9	3/9
	Overcast	3/5	0/5	2/5

↓ ↓ ↓

4/8 1/8 3/8

Slope One

- Example

	Item 1	Item 2
Alice	2	?
User 1	1	2

Slope One

- Example

	Item 1	Item 2
Alice	2	?
User 1	1	2

$2 + (2 - 1) = 3$

+1 →

- In general: Find a function of the form $f(x) = x + b$
 - that is why the name is “Slope One”

Slope One

- Example

	Item 1	Item 2	Item 3
Alice	2	5	?
User 1	3	2	5
User 2	4		3

Slope One

- Example

Alice
User 1
User 2

Item 2	Item 3
5	?
2	5
	3

Slope One

- Example

	Item 1	
Alice	2	
User 1	3	
User 2	4	

$\xrightarrow{+2}$
 $\xrightarrow{-1}$

	Item 3
	?
	5
	3

- Based on Item 1,

$$2 + \frac{2 + (-1)}{2} = 2.5$$

Slope One

- Example

	Item 1	Item 2	Item 3
Alice	2	5	?
User 1	3	2	5
User 2	4		3

- Rating: weighted average

$$\frac{2 \times 2.5 + 1 \times 8}{2 + 1} = 4.33$$