

2.a.

設 $R \in \{0, 1\}^{m \times n}$ 為「使用者-節目」的觀看矩陣，表示在三個月內某使用者是否觀看某節目。每一列是使用者，每一欄是節目。

我們定義：

$$T = RR^T$$

則 T 是一個 $m \times m$ 的矩陣，代表所有使用者兩兩之間的「互動程度」。

解釋：

1. 對角線元素 T_{ii} ：

- $T_{ii} = \sum_{j=1}^n R_{ij}^2$
- 因為 $R_{ij} \in \{0, 1\}$ ，平方後不變 $\Rightarrow T_{ii} = \sum_j R_{ij}$
- 即：使用者 i 總共看過的節目數 (degree)

2. 非對角元素 T_{ij} (當 $i \neq j$)：

- $T_{ij} = \sum_{k=1}^n R_{ik} \cdot R_{jk}$
- 表示使用者 i 和使用者 j 「共同看過的節目數」

這可以視為在「使用者-節目」的二分圖 (Bipartite Graph) 中，從 i 經由某節目走到 j 的所有長度為 2 的路徑條數。

2.b.1

◆ Cosine similarity 定義 (向量版)：

對於任兩節目 i, j ，其 cosine 相似度為：

$$\cos(i, j) = \frac{R_{*i} \cdot R_{*j}}{\|R_{*i}\| \cdot \|R_{*j}\|}$$

- 分子：共同被哪些使用者看過 $\Rightarrow R^T R$
- 分母：各節目被觀看的總人數 (平方根) 乘積 $\Rightarrow \|R_{*i}\| = \sqrt{|U_i|}$

◆ 用矩陣寫法表示：

我們希望將所有節目間的 cosine similarity 計算出來，即：

$$S_I = Q^{-\frac{1}{2}} \cdot R^T \cdot P^{-1} \cdot R \cdot Q^{-\frac{1}{2}}$$

◆ 各部分定義：

- $R^T R \in \mathbb{R}^{n \times n}$ ：每一項目間被哪些使用者共同觀看，稱為共現矩陣
 - $(R^T R)_{ij} = \sum_u R_{ui} R_{uj}$
- $Q \in \mathbb{R}^{n \times n}$ ：節目的度數對角矩陣
 - $Q_{ii} = \sum_u R_{ui} = |U_i|$ ：看過節目 i 的使用者數
- P^{-1} ：為中間加權，考慮 user 的活躍度，可省略或當成單位矩陣
- $Q^{-\frac{1}{2}}$ ：取平方根後反比，為了執行 cosine 分母歸一化

◆ 推導步驟總結：

1. 共現次數（節目對）： $R^T R$
2. 除以節目觀看人數的平方根：兩側乘上 $Q^{-\frac{1}{2}}$
3. 即為 cosine similarity：

$$S_I = Q^{-\frac{1}{2}} \cdot R^T \cdot R \cdot Q^{-\frac{1}{2}}$$

2.b.2

◆ 設定：

- 每個使用者 u 可視為一個向量（ R 的一列）：

$$R_{u*} \in \{0, 1\}^n$$

表示「看過哪些節目」

- 任兩使用者 u, v 的 cosine 相似度為：

$$\cos(u, v) = \frac{R_{u*} \cdot R_{v*}}{\|R_{u*}\| \cdot \|R_{v*}\|}$$

◆ 矩陣推導：

同樣方式，用矩陣表示：

$$S_U = P^{-\frac{1}{2}} \cdot R \cdot Q^{-1} \cdot R^T \cdot P^{-\frac{1}{2}}$$

◆ 各部分定義：

- $R \cdot R^T \in \mathbb{R}^{m \times m}$ ：使用者之間共同看過的節目數
- $P \in \mathbb{R}^{m \times m}$ ：使用者的節目觀看數對角矩陣
 - $P_{uu} = \sum_j R_{uj} = |I_u|$ ：使用者 u 看過的節目數
- Q^{-1} ：節目活躍度的加權 (optional)
- $P^{-\frac{1}{2}}$ ：平方根倒數，執行 cosine denominator normalization

◆ 簡化版本 (若不考慮節目加權)：

若不考慮節目加權 (即將 Q^{-1} 當作單位矩陣)，可簡化為：

$$S_U = P^{-\frac{1}{2}} \cdot R \cdot R^T \cdot P^{-\frac{1}{2}}$$

這個版本常用於純粹比較使用者行為模式，不考慮節目本身的權重。

2.c.1

◆ 核心觀念：

- 使用者 i 的推薦分數來自「與 i 相似的其他使用者」，這些人對節目 j 是否看過。
- 公式邏輯：

$$\Gamma_U = S_U \cdot R$$

其中：

- $R \in \mathbb{R}^{m \times n}$ ：使用者-節目互動矩陣 (0/1)
- $S_U \in \mathbb{R}^{m \times m}$ ：使用者-使用者的相似度矩陣

◆ 使用 cosine similarity 推導 S_U

為了用 cosine similarity 計算 S_U ，我們需要：

- P ：使用者觀看節目數的對角矩陣， $P_{ii} = \sum_j R_{ij} = |I_i|$
- Q ：節目觀看人數的對角矩陣， $Q_{jj} = \sum_i R_{ij} = |U_j|$

則：

$$S_U = P^{-\frac{1}{2}} \cdot R \cdot Q^{-1} \cdot R^T \cdot P^{-\frac{1}{2}}$$

◆ 帶入計算推薦分數：

$$\Gamma_U = S_U \cdot R = P^{-\frac{1}{2}} \cdot R \cdot Q^{-1} \cdot R^T \cdot P^{-\frac{1}{2}} \cdot R$$

✓ 最終矩陣表示式：

$$\Gamma_U = P^{-\frac{1}{2}} \cdot R \cdot Q^{-1} \cdot R^T \cdot P^{-\frac{1}{2}} \cdot R$$

◆ 推導說明：

1. $R \cdot Q^{-1} \cdot R^T$ ：衡量使用者彼此的共現程度（共同看過哪些節目），並考慮節目人氣的權重（避免熱門節目造成扭曲）
2. $P^{-\frac{1}{2}} \cdot (\cdot) \cdot P^{-\frac{1}{2}}$ ：將使用者度數正規化 \Rightarrow 就是 cosine similarity
3. 最後乘上 R ，代表以「相似使用者的行為」來預測每個節目是否適合這位使用者

2.c.2

◆ 基本公式：

- 每位使用者看過某些節目，我們假設他也可能會喜歡「與這些節目相似」的其他節目
- 推薦分數矩陣公式為：

$$\Gamma_I = R \cdot S_I$$

其中 S_I 是節目-節目的相似度矩陣（基於 cosine）

◆ 推導節目-節目相似度矩陣 S_I

使用 cosine similarity 計算：

- $R^T \cdot R$ ：節目之間的共現次數
- Q ：節目觀看人數的對角矩陣（ $Q_{jj} = |U_j|$ ）

則：

$$S_I = Q^{-\frac{1}{2}} \cdot R^T \cdot R \cdot Q^{-\frac{1}{2}}$$

◆ 帶入推薦矩陣：

$$\Gamma_I = R \cdot S_I = R \cdot Q^{-\frac{1}{2}} \cdot R^T \cdot R \cdot Q^{-\frac{1}{2}}$$

✓ 最終矩陣表示式：

$$\Gamma_I = R \cdot Q^{-\frac{1}{2}} \cdot R^T \cdot R \cdot Q^{-\frac{1}{2}}$$

◆ 推導說明：

1. $R^T \cdot R$ ：節目間共現的 raw count (哪些節目常被同一批使用者一起看)
2. 兩側乘上 $Q^{-\frac{1}{2}}$ ：將節目觀看人數歸一化，構成 cosine similarity
3. 最後用 $R \cdot S_I$ ：讓使用者過去看過的節目，去加權與其他節目的相似度，以預測其他節目的推薦分數

3.a.1

◆ 第一小題：對 q_i 、 p_u 推導梯度 (各 3 分，共 6 分)

✓ 先定義所有符號：

符號	意義
R	評分矩陣， R_{iu} 是使用者 u 對項目 i 的評分
q_i	項目 i 的 latent vector，為 Q 的第 i 列
p_u	使用者 u 的 latent vector，為 P 的第 u 列
k	latent factor 維度數
λ	正則化參數，防止過擬合
\hat{r}_{iu}	預測分數， $\hat{r}_{iu} = q_i \cdot p_u^T$

◆ 目標函數重寫為兩項：

1. 誤差平方和：

$$E_{\text{error}} = \sum_{(i,u)} (R_{iu} - q_i \cdot p_u^T)^2$$

2. 正則化項：

$$E_{\text{reg}} = \lambda \left(\sum_u \|p_u\|^2 + \sum_i \|q_i\|^2 \right)$$

◆ 對 p_u 求偏導數：

我們只對使用者 u 相關的項目求導：

$$\frac{\partial E}{\partial p_u} = \sum_{i \in \mathcal{I}_u} \frac{\partial}{\partial p_u} (R_{iu} - q_i \cdot p_u^\top)^2 + \lambda \cdot \frac{\partial}{\partial p_u} \|p_u\|^2$$

第一項使用鏈式法則：

$$\frac{\partial}{\partial p_u} (R_{iu} - q_i \cdot p_u^\top)^2 = -2(R_{iu} - q_i \cdot p_u^\top)q_i$$

第二項為：

$$\frac{\partial}{\partial p_u} \|p_u\|^2 = 2p_u$$

✅ 所以得到：

$$\frac{\partial E}{\partial p_u} = -2 \sum_{i \in \mathcal{I}_u} (R_{iu} - q_i \cdot p_u^\top)q_i + 2\lambda p_u$$

◆ 對 q_i 求偏導數（同理）：

$$\frac{\partial E}{\partial q_i} = -2 \sum_{u \in \mathcal{U}_i} (R_{iu} - q_i \cdot p_u^\top)p_u + 2\lambda q_i$$

✅ 最終表達式：

$$\frac{\partial E}{\partial q_i} = -2 \sum_{u \in \mathcal{U}_i} (R_{iu} - q_i \cdot p_u^\top)p_u + 2\lambda q_i$$

3.a.2

- ◆ **SGD 精神**：針對單筆樣本 (i, u, R_{iu}) 隨機抽樣、即時更新

先計算預測誤差：

$$e_{iu} = R_{iu} - q_i \cdot p_u^\top$$

則 SGD 的更新公式為：

✓ 對 p_u ：

$$p_u \leftarrow p_u + \eta \cdot (e_{iu} \cdot q_i - \lambda p_u)$$

✓ 對 q_i ：

$$q_i \leftarrow q_i + \eta \cdot (e_{iu} \cdot p_u - \lambda q_i)$$

✓ 所有符號總整理：

符號	意義
η	學習率 (learning rate)，控制每步更新的幅度
e_{iu}	當前預測誤差： $R_{iu} - \hat{r}_{iu}$
λ	正則化係數，防止參數過大導致過擬合
q_i, p_u	項目 / 使用者的 latent 向量