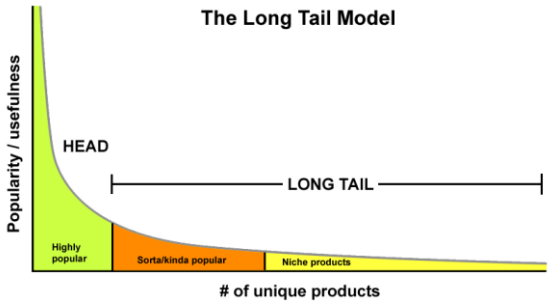
# Content-based RS

## Introduction

*Long tail in commerce*

- Pareto principle (80/20 Principle): roughly 80% of the effects come from 20% of the causes
  - 80% of words we use comes from 20% of dictionary
  - 80% of asset all over the world comes from 20% of people
- Traditional store also applies the principle: 80% of amount of customer buys just 20% of goods in the store, so every store has the warehouse to store other 80% of goods
- Ecommerce can know which 20% of goods that certain customer needs
- Long tail phenomenon also depicts the principle: just 20% of goods (x-axis) are popular




*2 kinds of RS*

- Content-based RS: In one user, based on the existed ratings of the user, RS can predict the ratings the user does not rate
  - If the user rates the fruit high and candy low, RS will recommend the fruit rather than candy
  - The approach needs the **item embedding** which requires to classify the items into categories or find the features of item, but some items cannot classify or find any features
- Collaborative filtering: recommend the goods to the user based on the activity of other users
  - 3 customers A, B, C like the songs of singer NPT. Data shows that customer B, C likes song of singer BT but data doesn't show whether A likes song of singer BT. Based on the similar users like B and C, RS recommend singer BT to A

## Utility Matrix

*Construct Utility Matrix*

- Without Utility matrix, RS cannot do anything, so utility matrix is vital, but it got several difficulties, we have 2 popular approach to construct the matrix
  - Explicit feedback: but the user rarely rated directly the items and also got the bias feedback
  - Implicit feedback: If the user buy item on Amazon, watch the video on Youtube, we can tick 1 for point (item, user), and sometimes, we can record the `dislike` by -1

## Content-based Recommendation

*Construct Item Profile – Item embedding*

| | A | B | C | D | E | F | item's feature vectors |
|---|---|---|---|---|---|---|---|
| Mưa nửa đêm | 5 | 5 | 0 | 0 | 1 | ? | $\mathbf{x}_1 = [0.99, 0.02]^T$ |
| Cỏ úa | 5 | ? | ? | 0 | ? | ? | $\mathbf{x}_2 = [0.91, 0.11]^T$ |
| Vùng lá me bay | ? | 4 | 1 | ? | ? | 1 | $\mathbf{x}_3 = [0.95, 0.05]^T$ |
| Con cò bé bé | 1 | 1 | 4 | 4 | 4 | ? | $\mathbf{x}_4 = [0.01, 0.99]^T$ |
| Em yêu trường em | 1 | 0 | 5 | ? | ? | ? | $\mathbf{x}_5 = [0.03, 0.98]^T$ |
| User's models | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\theta_5$ | $\theta_6$ | ← **need to optimize** |

•We can base on the categories like singer, musician, year composed, but in this case, to make it simple, we just consider how level of Bolero the song is

- After item profile, we will have
  - Item profile $X = \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \dots & \mathrm{x}_M \end{bmatrix} \in \mathbb{R}^{d \times M}$ ; M: number of item; d: number of features of each items. **Note**: in the exercise, little reverse, $X \in \mathbb{R}^{M \times d}$

- Utility Matrix: $Y \in \mathbb{R}^{M \times N}$ ; M: number of item; N: number of users
- But the Utility Matrix Y miss lots of point, so we apply Linear Regression to find the weight $W \in \mathbb{R}^{d \times N}$ and bias $b \in \mathbb{R}^{1 \times N}$
- Rated or not matrix: $R \in \mathbb{R}^{M \times N}$ ; $r_{mn} = 1$ if user #n rate item #m

## Model

- $Y = X^T W + b \in \mathbb{R}^{M \times N}$
- For each user, We only get the item profile the user already rated $\hat{X}_n \in \mathbb{R}^{s_n \times d}$, $w_n \in \mathbb{R}^d$ : column #n in $W$ , the rating user rated $\hat{y}_n \in \mathbb{R}^{s_n}$ , $b_n \in \mathbb{R}$, $s_n$: number of items the user rated: $L(w_n, b_n) = \frac{1}{s_n}\left(\frac{1}{2}\left\|\hat{\mathrm{X}}_n^T \mathrm{w}_n + b_n \mathrm{e}_n - \hat{\mathrm{y}}_n\right\|_2^2\right) + \frac{\lambda}{2}\left\|\mathrm{w}_n\right\|_2^2$
- For example, consider *user* D with $\hat{\mathrm{y}}_4 = \begin{bmatrix} 0 & 0 & 4 \end{bmatrix}^T$, $\mathrm{e}_4 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$, $\hat{\mathrm{X}}_4 = \begin{bmatrix} 0.99 & 0.91 & 0.95 \\ 0.02 & 0.11 & 0.99 \end{bmatrix}$

  $\rightarrow L(\mathrm{w}_4, b_4) = \frac{1}{3}\left(\frac{1}{2}\left\|\hat{\mathrm{X}}_4^T \mathrm{w}_4 + b_4 \mathrm{e}_4 - \hat{\mathrm{y}}_4\right\|_2^2\right) + \frac{\lambda}{2}\left\|\mathrm{w}_4\right\|_2^2$