Problem 1

It is often stated, that pure content-based recommendation models provide very low level of personalization to users.

Prove this claim using a standard regression-based formulation for the case when a single global model is learned in the form:

$$r \approx \theta^T x + \epsilon$$

where vector x encodes some features of both users and items (e.g., user attributes and item characteristics), and θ are the corresponding learnable weights of the regression model. Recall that personalization task is formulated as the ranking problem of the top-n best-matching items:

$$\text{toprec}(u, n) = \arg\max_i{}^n r_{ui}$$

where r_{ui} is the relevance score assigned by the model to item i for user u.

$$\arg\max_i{}^n r_{ui} = \arg\max_i{}^n (\theta^T x + \epsilon)$$

Let's clarify what we have:

θ^T - some learnable weights for both user and item features represented like:

$$\theta^T = (\theta_U^T, \theta_I^T), \text{ where } U \text{ stands for user and } I \text{ for item correspondigly.}$$

r_{ui} - predicted rating of item i for user u:

Thus, $r_{ui} \approx \theta^T x_{ui} + \epsilon = (\theta_U^T, \theta_I^T) \cdot \begin{pmatrix} x_u \\ y_i \end{pmatrix} = b_u + \theta_U^T x_u + b_i + \theta_I^T y_i$

According to our model our predictions for items i and j will be ranked as follows:

$$\Rightarrow \Delta r = r_{ui} - r_{uj} = \underbrace{b_u - b_u}_{0} + \theta_U^T x_u - \underbrace{\theta_U^T x_u}_{0} + \underbrace{b_i - b_j}_{0} + \theta_I^T y_i - \theta_I^T y_j =$$

$$= \theta_I^T (y_i - y_j)$$

Well, now we can see, that our ranking approach doesn't really rely on user preferences.

Problem 2

You're given the matrix of interactions between 3 users and 6 items:

$$M = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Is it possible to build a personalized recommendation model with this data? Explain your answer.

I suppose it's not possible, because there are no overlaps between users and items.