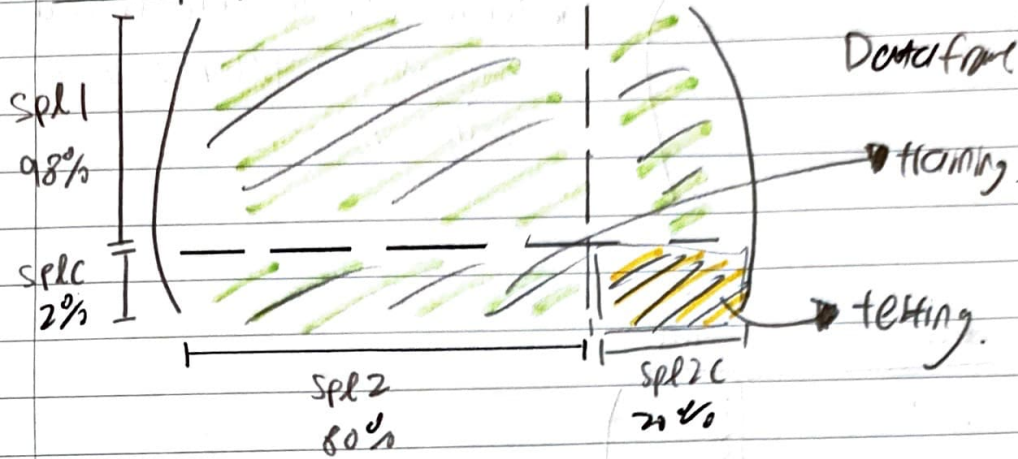


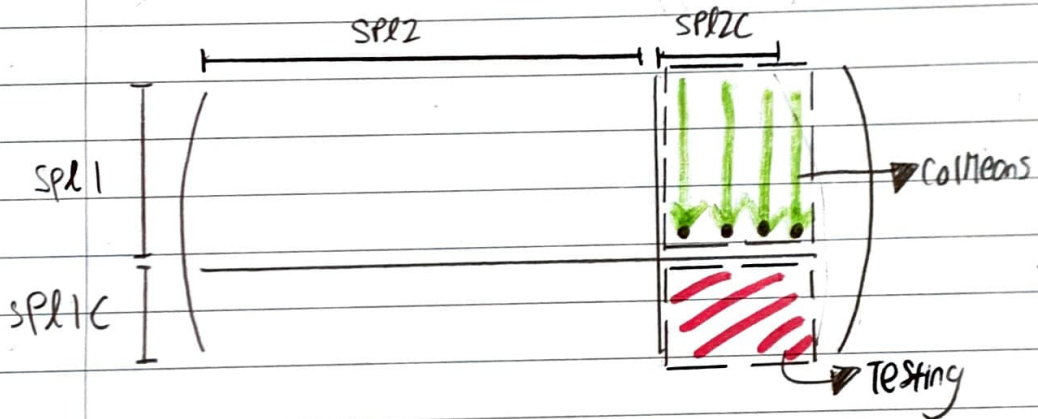
James G8M

SMA WK 11 - collaborative filtering

Data preparation



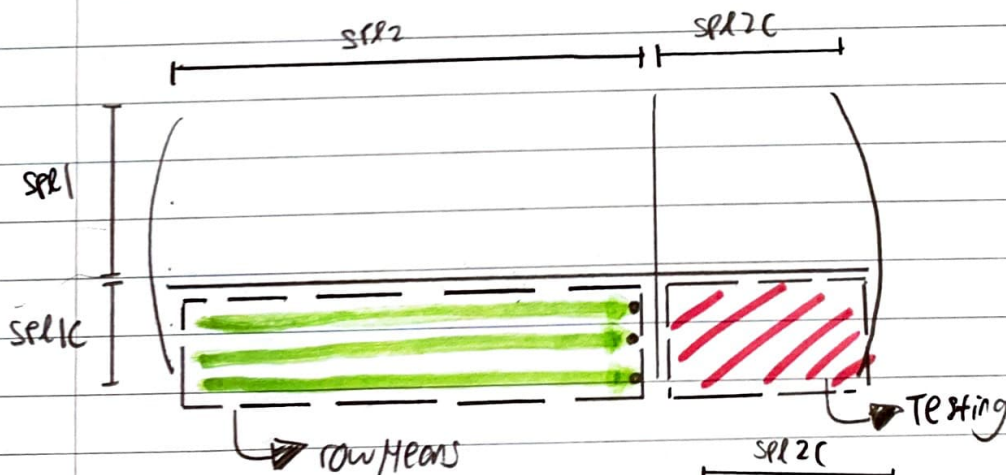
A) Baseline model 1



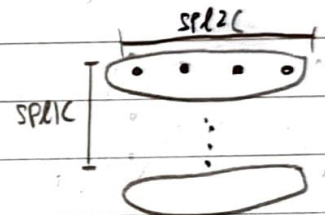
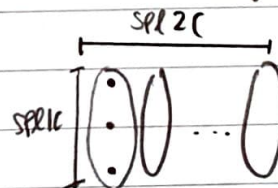
Interpretation:
Calculate the **average rating**
to each movie in **spl2c**
given by users in **spl1**

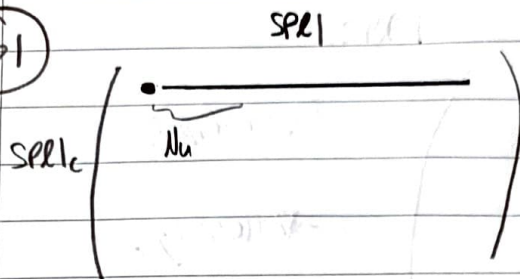
The dots are the column
mean that are applied
equally to all values
in the **TRAIN** set.

B) Baseline model 2



Interpretation: **average**
Calculate the **average rating** given by each
user in **spl1c** to the movies
in **spl2**



User-based model**Step 1**

→ For every user in S_{k1} , we find which users in S_{k1} has the highest correlation

Step 2