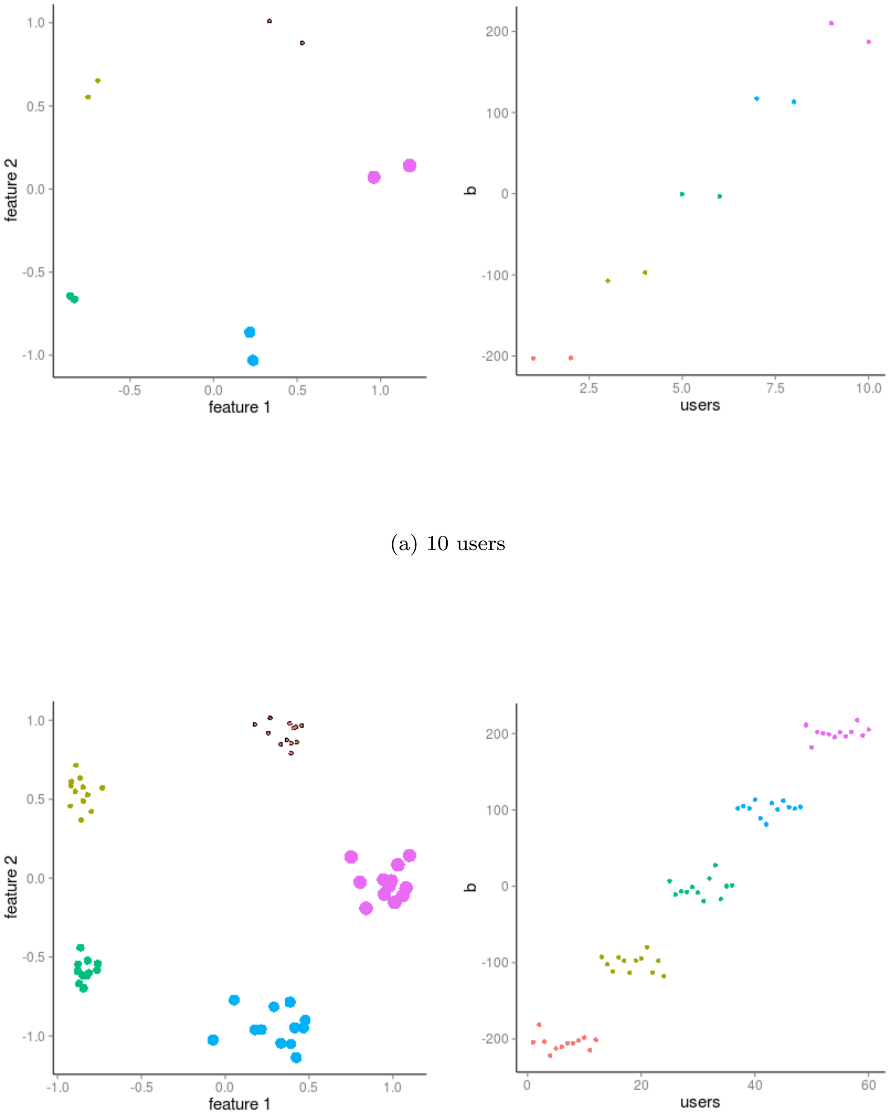
Clustering over user features and latent behavioral functions with dual-view mixture models
-Experiments-

The experiments will be repeated for the settings plotted in below. As usual, the left side of the figures represents the user features and the right side represents the coefficients of each user. Unlike figures in the paper, this time I added the coefficient information also as a size attribute of every user point. [1]

1 Agreement between views

This has already been done and reported in the paper.

[1] Indeed, now the left side of the plots is self-contained. Our task may also be explained as: we have to cluster points considering also size, where size is latent and must be inferred from some other information



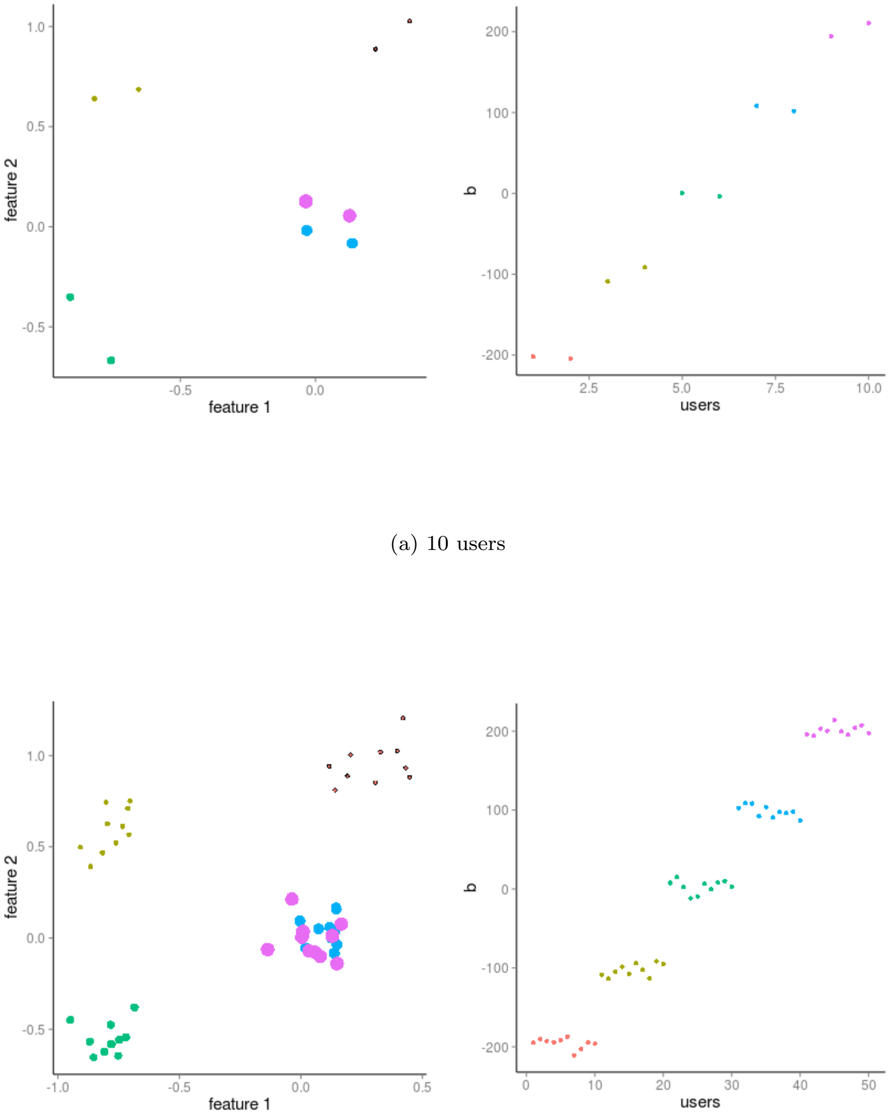
(a) 10 users

(b) 50 users

Figure 1: Examples of datasets for the first setting.

2 Disagreement between views

This has been already done but will be further documented in the paper. There are two ground truths corresponding to 4 and 5 clusters. Accuracy of the models will be reported with respect to these two ground truths.



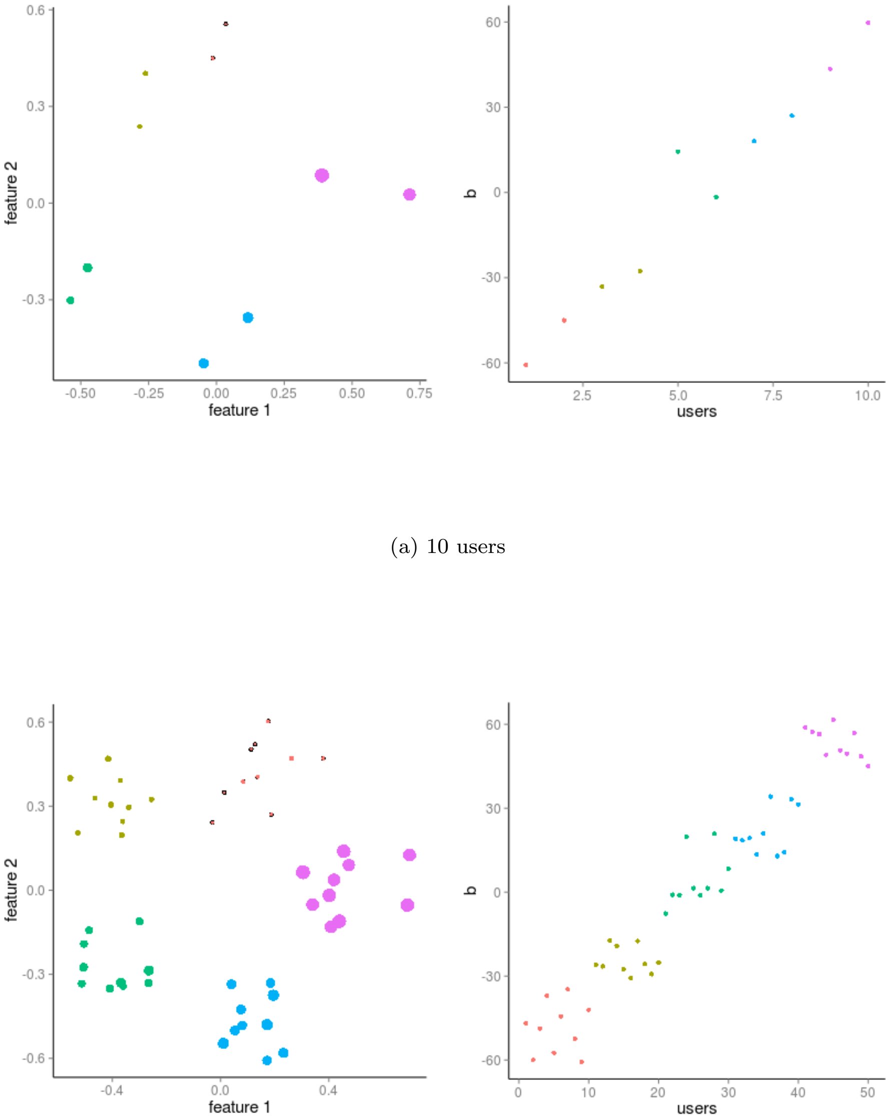
(a) 10 users

(b) 50 users

Figure 2: Examples of datasets for the second setting.

3 Overlapping

This is a new scenario. The idea is to reproduce a structure where two views agree but where the clustering structure (in either view) is not so strong.



(a) 10 users

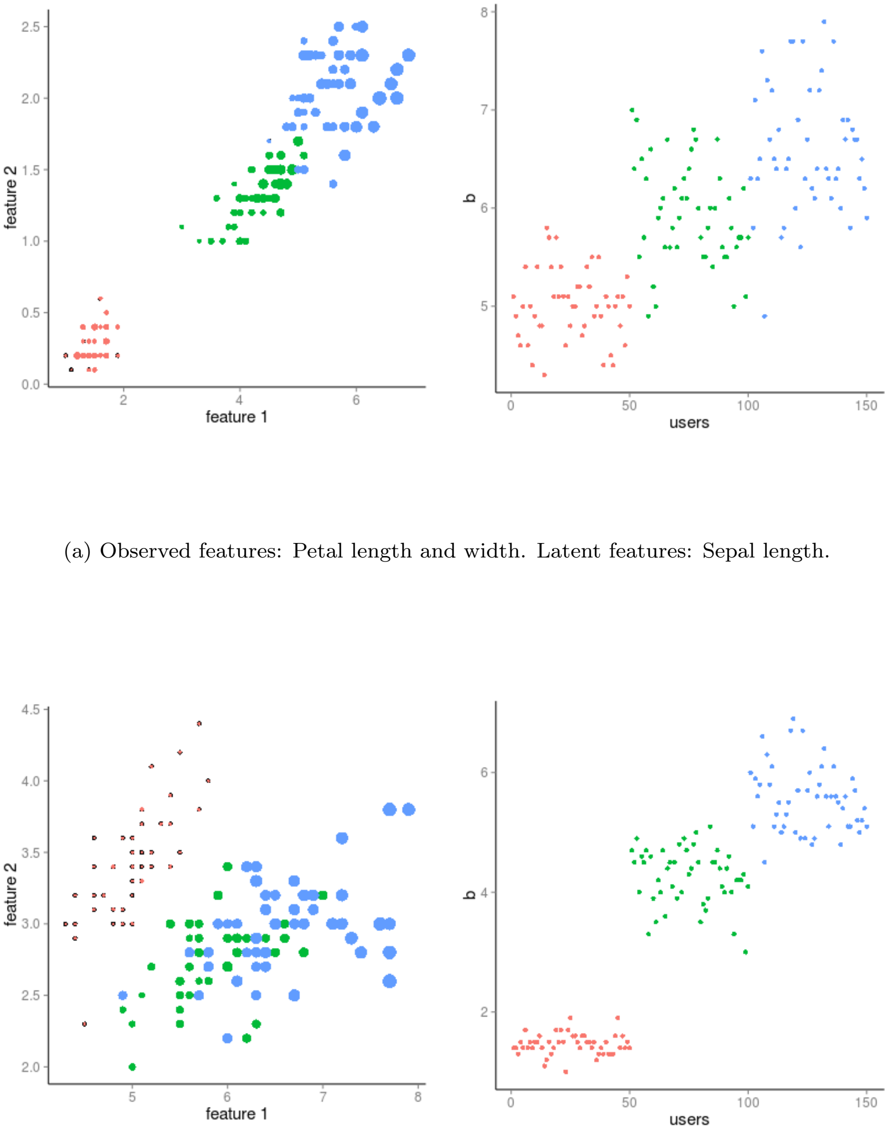
(b) 50 users

Figure 3: Examples of datasets for the third setting.

4 Overlapping (real data)

This is a new scenario. The idea is to reproduce the structure of a real dataset (`iris`). We will consider that users are flowers and with two observed features and a latent one taken from the real features (`sepal length, sepal width, petal length, petal width`).

As in the other scenarios, the latent coefficient is used to generate some observation. For instance, the thread length given a participation matrix, that is $y_t = P^t b$) The more threads we have, the easier it will be to infer the hidden coefficients (e.g.: the sepal length). At the extreme case, the coefficients can be perfectly recovered and then the clustering problem is equivalent to the original one where all dimensions where observed.



(a) Observed features: Petal length and width. Latent features: Sepal length.

(b) Observed features: Sepal length and width. Latent features: Petal length.

Figure 4: Examples of datasets for the iris setting.