Econ 900 Problem set 1 Report

Haibin Jiang

April 23, 2019

In this report I look at the relationship between the price of the game and the users'rating of the game.

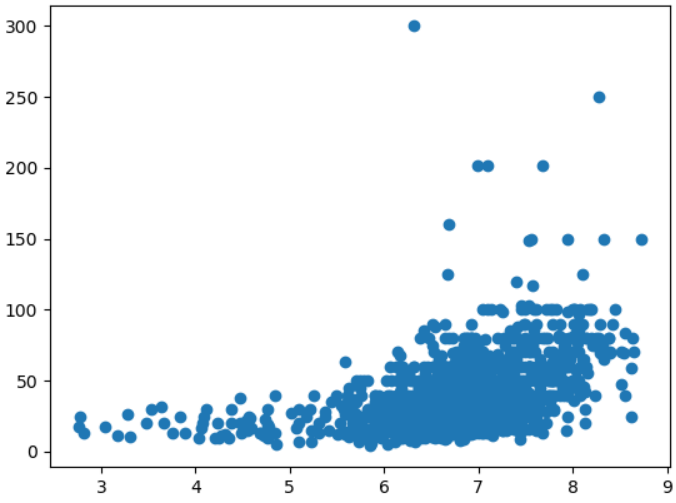
I first plot the relationship between list price of the game and the customer rating of the game. As shown in Figure 1, the list price of the game is positively associated with the customer player rating. It seems that the better game is well charged.

But this is just a general trend. In order to have a better understanding of the relationship between the list price and the customer rating, I use the GaussianMixture to separate the data in clusters. This is an unsupervised learning, with no priori hypothesis, just using the data driven algorithm. To see how many cluster is appropriate I report the silhouette score for each clustering process. The results of the silhouette score are shown in Table 1. From the silhouette score, it seems that the data is better separated into 4 clusters.

Table 1: Number of components and Silhouette score

number of components	Silhouette score
2	0.39952376133117234
3	0.44701307212489155
4	0.5083493836471616
5	0.36574401476811047
6	0.2884292281841645

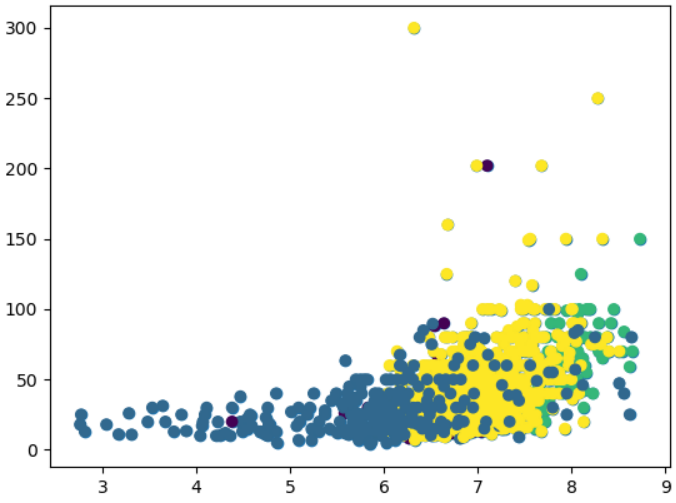
Figure 1: Correlation between list price and customer rating



I also tried higher number of clusters, but the Silhouette scores are lower. It seem the games can be clustered into three or four groups. Figure 2 shows the clusters with component equal 4. Within each cluster, the price and the rating seems positively associated. We can tell there are some differences among the different clustered groups with different colored dots. The group with blue dots form a group that have relatively low price and relatively low rating. The group with yellow dots form a group that list relatively high price and receive relatively high rating. And the group with green dots form a group that receives very high rating but still charges a relatively low price. The pattern of the group with purple dot is not quite clear.

If the rating reflect the real customer satisfaction, and if the customer satisfaction is related to the effort of game design, then the yellow group has the good return for the game designers. For it charge good price for their good quality. The green group can be under priced, for it generate high customer

Figure 2: Gaussian Cluster with 4 components



satisfaction but charges a low price.

More information about the game will be needed to test the hypothesis.