ECON 31703: Assignment 3

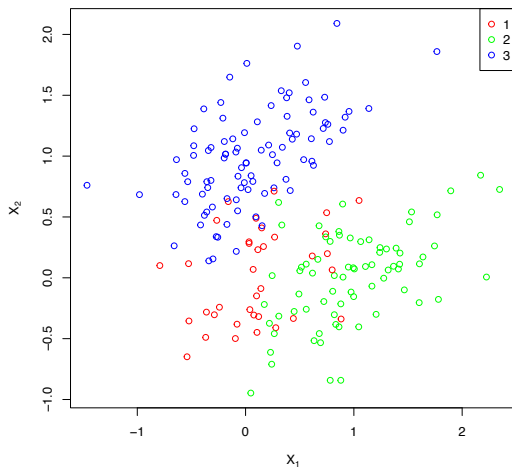sasha petrov

June 7, 2022

(e)

Figure 1: true clustering
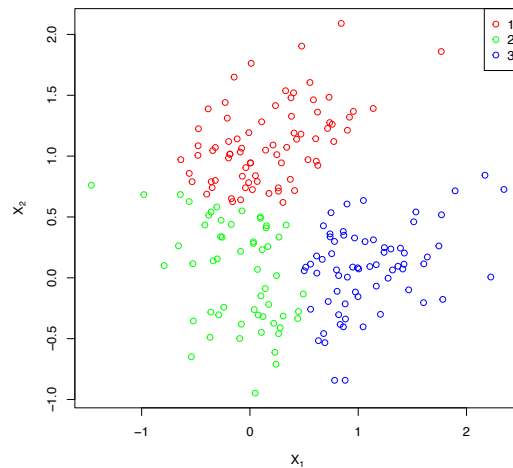
Figure 2: best k-means

Figure 2 captures the best partitioning of X_i-space into 3 clusters. Overall, it does a fair job of revealing the true hierarchical model (that is highlighted in figure 1). However, importantly, k-means fails to reflect a) the possibility of 'overlap' among domains of models corresponding to different hyper-parameters; b) the difference in 'weights' of different models. That is, k-means partitions the given 'cloud' of data points into 3 roughly equally-sized and non-intersecting sections on the plane.

Factors that seem important to me for how well k-means works:

- distance between means of different hyper-models

    - in our case, it seems that distances are large enough to separate 3 distinct areas

- variance within each of the hyper-models

– in our case, variances are quite low, which makes the risk of 'overlapping' quite low and, thus, leads k-means to do a good job

- correlation structure between covariates; shapes of 'clouds' corresponding to different hyper-models matter: given the same variance, if there's higher correlation we expect less risk of 'overlapping' (intuitively, ellipsoids that are prolongated in the same direction are harder to intersect than circles)

– in our case, there is some positive correlation between X_1 and X_2, which probably helps the algorithm to avoid building clusters that have a 'wrong' shape

(f)

Table 1: trying different $\tilde{K}$

$\tilde{K}$	2	3	4	5
bias	0.47	0.18	0.07	0.04
variance	0.01	0.02	0.03	0.03

Table 1 reports average biases and variances (as defined in the problem set) for 4 different values of $\tilde{K}$. We see something that resembles a classic bias-variance trade-off. As the number of suggested clusters increases, the average bias goes down dramatically but at the cost of the increasing variance of estimates. Interestingly, the shifts along this trade-off become less pronounced as the number of clusters increases. The reason for this is probably that once the number of clusters becomes sufficiently high, some of them are negligibly different (because they capture points from the same hyper-model); therefore, we shouldn't expect to see a change in patterns of $\mu_{\theta_i 1}$ across some subsets of clusters; in other words, both $\bar{\mu}_{k_i} - \mu_{\theta_i 1}$ and $Y_i - \mu_{\theta_i 1}$ won't change much when we 'split' certain clusters.