Semi-supervised Modelling

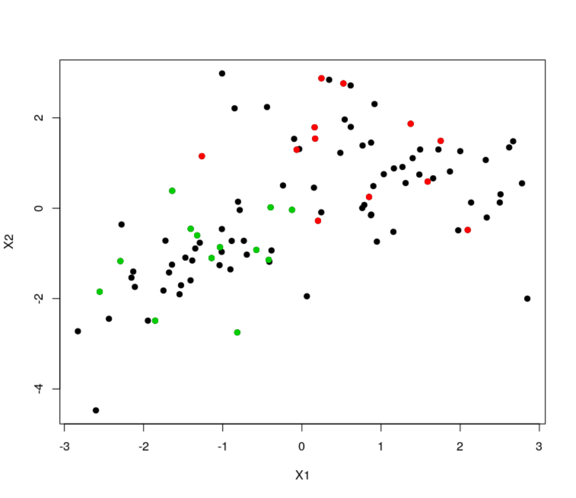
UBCO MDS — DATA 573

Semi-supervised Modelling

- We have focussed on supervised methods from 570 through 572. This includes taking known (labeled) responses and fitting a model — you might then predict for some unknown (unlabeled) responses using that fitted model.
- The bulk of 573 has been focussed on unsupervised methods. This includes taking unlabeled responses and fitting a model.
- Let's motivate an alternative through some examples...

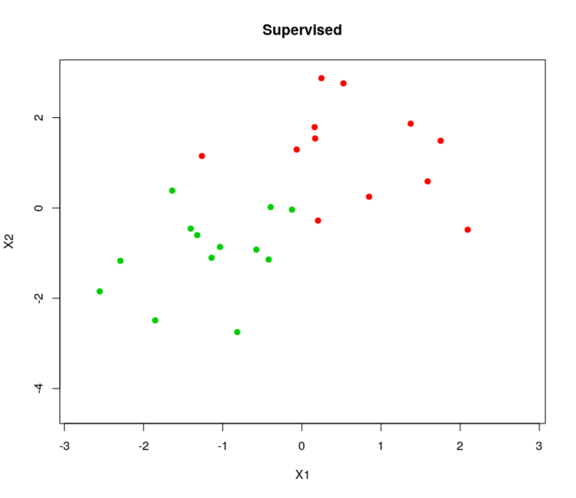
Supervised vs Unsupervised

- Suppose that 25% of cases have labeled response from the following data



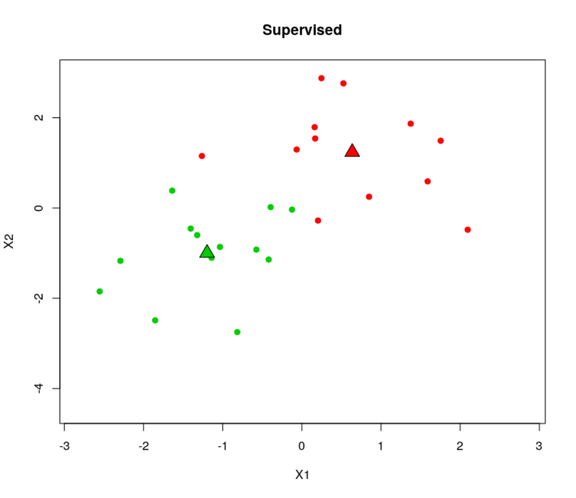
Supervised estimation

- Let's calculate the means from each group in a supervised manner



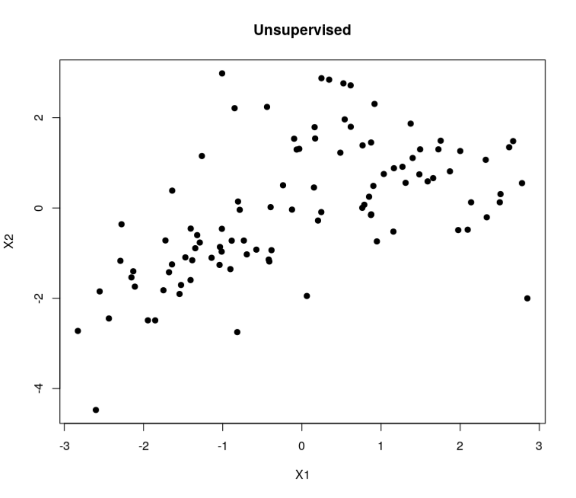
Supervised estimation

- Let's calculate the means from each group in a supervised manner



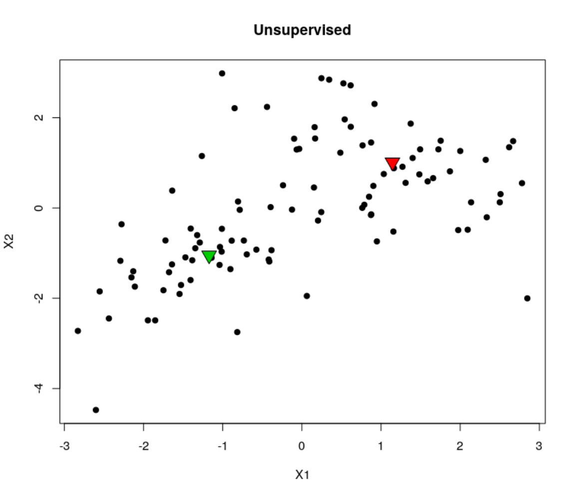
Unsupervised estimation

- Let's calculate the means for estimated groups in an unsupervised manner (using mclust)



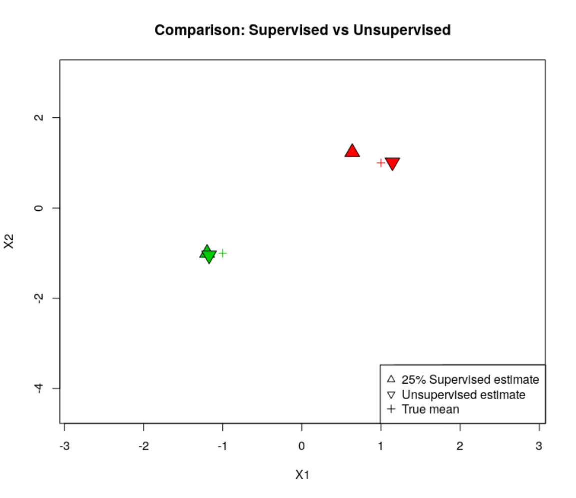
Unsupervised estimation

- Let's calculate the means for estimated groups in an unsupervised manner (using mclust)



Supervised vs Unsupervised

- Here are the group means on the same plot

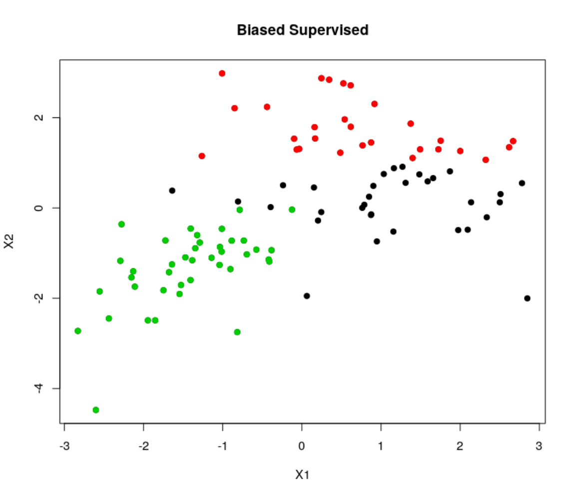


Comments

- And that was with 25% randomly selected to have known labels.
- Suppose instead that your labeling process was biased...
- One natural bias that can happen in realistic scenarios is that clear cases are labeled, and less clear cases might be left for an algorithm to sort out.

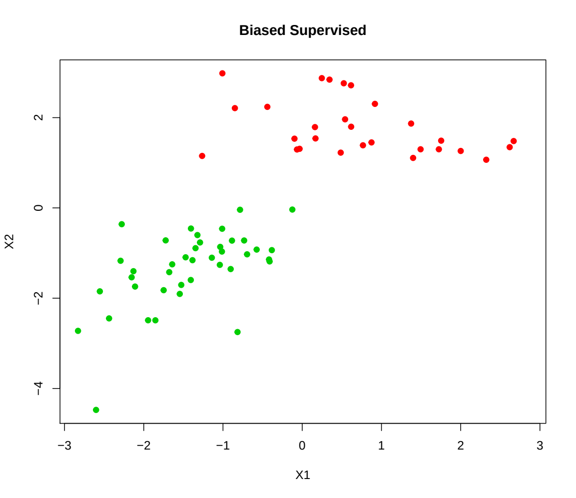
Bias Supervised

- Suppose that the following cases have been hand-labeled



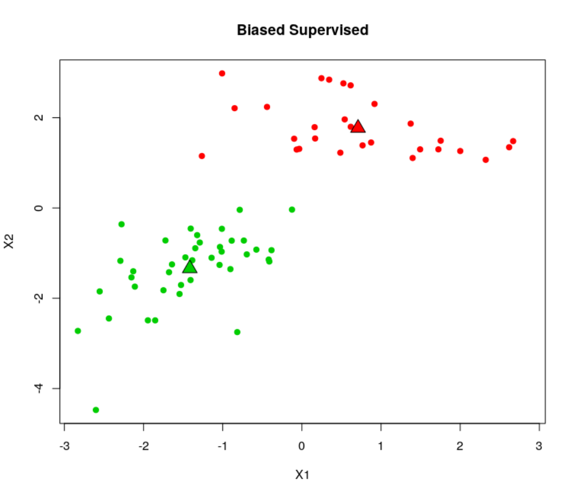
Bias Supervised

- We then estimate the means...



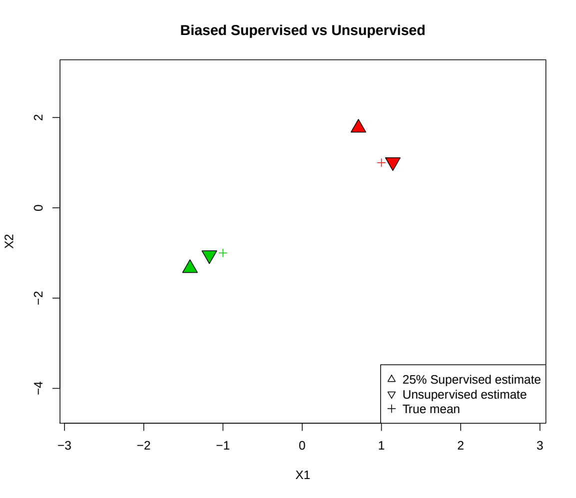
Bias Supervised

- We then estimate the means...

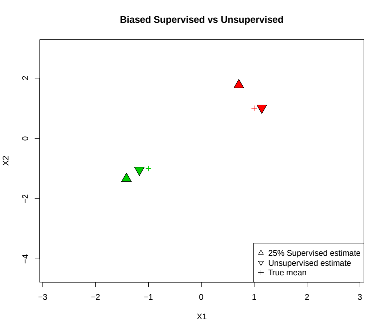


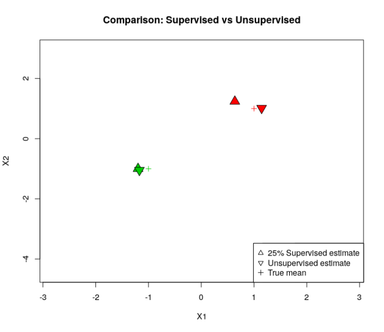
Bias Supervised vs Unsupervised

- And again compare to unsupervised...



Side by side


Biased Supervised vs Unsupervised
X1
X2
△ 25% Supervised estimate
▽ Unsupervised estimate
+ True mean
−3
−2
−1
0
1
2
3
2
0
−2
−4


Comparison: Supervised vs Unsupervised
X1
X2
△ 25% Supervised estimate
▽ Unsupervised estimate
+ True mean
-3
-2
-1
0
1
2
3
2
0
-2
-4

Comments

- Of course, if the group structure is less clear, unsupervised methods can fail horribly on this type of estimation example.
- In which case supervised methods will outperform.
- Furthermore, it seems like there should be some way to take advantage of labeled responses to improve on (blind) unsupervised methods.

Semi-supervised Machine Learning

- There is a third option....or rather, a third set of options[1].
- Semi-supervised methods use both labeled and unlabeled observations in tandem during the model fitting process.
- While such an approach is usable in several classification models, we will focus on mixture models (surprise, surprise).

[1] Vrbik, I., & McNicholas, P. D. (2015). Fractionally-supervised classification. Journal of Classification, 32(3), 359-381.

Semi-supervised Modelling: Specifics

- Suppose X_u contains all the predictors with unlabeled response (unobserved y_u), and X_l contains all predictors with labeled response (observed y_l).

Paradigm	Avail. Data	Estimate $\hat{f}$ with	Provides
Supervised	y_l, X_l, X_u	y_l, X_l	$\hat{y}_l$, $\hat{y}_u$
Unsupervised	y_l, X_l, X_u	X_l, X_u	$\hat{y}_l$, $\hat{y}_u$
Semi-supervised	y_l, X_l, X_u	y_l, X_l, X_u	$\hat{y}_l$, $\hat{y}_u$

- For mixture models:
  - Supervised = Discriminant Analysis (Gaussian unconstrained = QDA)
  - Unsupervised = Model-based clustering (Gaussian unconstrained = Mclust 'VVV' model)
  - Semi-supervised = sometimes referred to as Model-based classification

Semi-supervised Modelling: Specifics

- One reason to focus on mixture models (beyond my affinity for them), is that implementing semi-supervised modelling is relatively trivial.
- The basic idea is to take your labeled cases (y_l) and both pre-determine group membership, as well as not allow that group membership to change during the model-fitting process.
- This is easy to implement in the EM algorithm...let's quickly review

Recall: EM Algorithm for Clustering

1. Start the algorithm with random values for $\hat{z}_{ig}$. (there are alternative starting options)

2. Assuming those $\hat{z}$ are correct, estimate parameters $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$ (via MLEs — hence, this is the maximization of EM)

3. Assuming those parameters are correct, find the expected value of group memberships

$$\hat{z}_{ig} = \frac{\pi_g \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g)}{\sum_{g=1}^{G} \pi_g \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g)}$$

   (this is the expectation of EM)

4. Repeat 2. and 3. until 'changes' are minimal. (The log-likelihood of the model is monitored for convergence)

EM Algorithm for Semi-supervised Classification

1. Start the algorithm with random values for $\hat{z}_{ig}$ where $i \in X_u$. For $j \in X_l$, set $\hat{z}_{jg} = 1$ if $y_l = g$, otherwise 0

2. Assuming those $\hat{z}_{ig}$ are correct, estimate parameters $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$

3. Assuming those parameters are correct, find the expected value of group memberships for $i \in X_u$

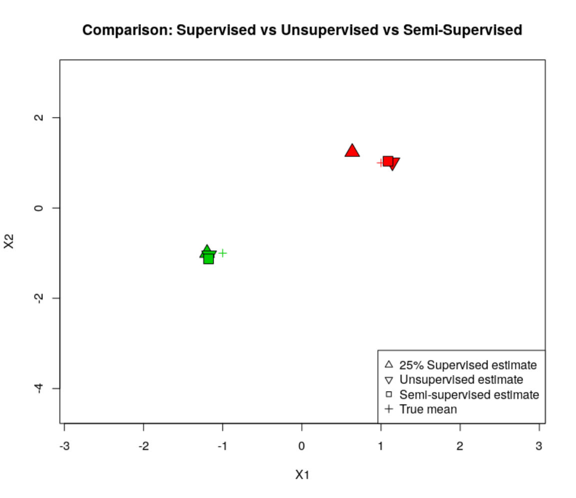
$$\hat{z}_{ig} = \frac{\pi_g \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g)}{\sum_{g=1}^{G} \pi_g \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g)}$$

   (aka, update unlabeled estimates, leave labeled estimates as initialized)

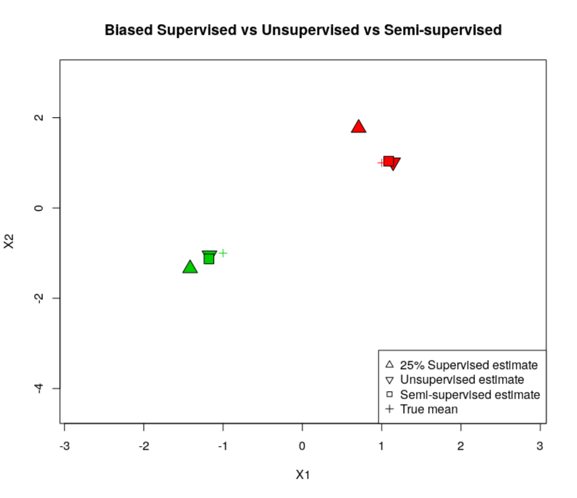
4. Repeat 2. and 3. until 'changes' are minimal. (The log-likelihood of the model is monitored for convergence)

25% Supervised vs Unsupervised vs Semi-Supervised

- We can reanalyze those same simulations from before using this paradigm.



Bias Supervised vs Unsupervised vs Semi-Supervised



Comments

- More recent versions of MCLUST can run supervised classification through MclustDA() and semi-supervised through MclustSSC().
- tEIGEN can be used semi-supervised by inputting NA's for the unlabeled observations in the "known" vector.


UBC
THE UNIVERSITY OF BRITISH COLUMBIA