Review of Stratified Psychiatry via Convexity-Based Clustering with Applications Towards Moderator Analysis

Thaddeus Tarpey, Eva Petkova, and Liangyu Zhu

Hanchao Zhang

October 21, 2021

1 Convexity-Based Clustering: From Discriminant Analysis to Partitioning

We focus on the population with two groups. In supervised learning, each observation comes with a class label indicating to which sub-population the observation belongs. Suppose the density in each sub-population is f_1 and f_2, using these information in discriminant analysis, the optimal rule for classification in terms of minimizing the probability of misclassification in Bayes' rule is to classfy an observation $\boldsymbol{x}$ to population 1 if

$$\frac{\pi_1 f_1(\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})} \geq \frac{\pi_2 f_2(\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})} \tag{1}$$

and it classifies to population 2 otherwise. The mixture density function in the denominator of equation 1 is

$$f(x) = \pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x}) \tag{2}$$

The magnitude of subtracting the left hand side by the right hand side for an observation $\boldsymbol{x}$ in equation 1 is a measure of the strength in how well the observation can be classified to one or the other sub-population.

$$\phi(\lambda(\boldsymbol{x})) = \left(\frac{\pi_1 f_1(\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})} - \frac{\pi_2 f_2(\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})} \right)^2 \tag{3}$$

$$= \left(\frac{\pi_1 f_1(\boldsymbol{x})}{f(\boldsymbol{x})} - \frac{\pi_2 f_2(\boldsymbol{x})}{f(\boldsymbol{x})} \right)^2 \tag{4}$$

where $\lambda(\boldsymbol{x}) = \frac{\pi_1 f_1(\boldsymbol{x})}{f(\boldsymbol{x})} - \frac{\pi_2 f_2(\boldsymbol{x})}{f(\boldsymbol{x})}$, and $\phi(\cdot) = (\cdot)^2$. Or we can also define $\lambda(\boldsymbol{x}) = \frac{\pi_2 f_2(\boldsymbol{x})}{f(\boldsymbol{x})}$, and $\phi(\lambda) = (1 - 2\lambda)^2$. The objective function that we want to maximize becomes:

$$
\begin{aligned}
\mathscr{L} &= \sum_{j=1}^{k} P(B_j)\phi(E[\lambda(X)|X \in B_j]) && (5)\\
&= \sum_{j=1}^{k} P(B_j)(1 - 2E[\lambda(X)|X \in B_j])^2 && (6)\\
&= \sum_{i=1}^{k} P(B_j)\left(1 - \frac{2}{P(B_j)} \int_{B_j} \frac{\pi_2 f_2(x)}{\cancel{f(x)}} \cancel{f(x)} dx\right)^2 && (7)\\
&= \sum_{i=1}^{k} P(B_j)\left(1 - \frac{2\pi_2 P_2(B_j)}{P(B_j)}\right)^2 && (8)\\
&= \sum_{i=1}^{k} P(B_j)\left(\frac{\pi_1 P_1(B_j) + \pi_2 P_2(B_j) - 2\pi_2 P_2(B_j)}{P(B_j)}\right)^2 && (9)\\
&= \sum_{i=1}^{k} \frac{\big(\pi_1 P_1(B_j) - \pi_2 P_2(B_j)\big)^2}{P(B_j)} && (10)
\end{aligned}
$$

2 Semi-Supervised Discriminant Clustering Algorithm

1. Start with initial partition $B_1, \ldots, B_k$
2. Calculate the support points as $w_j = E[\lambda(\boldsymbol{x})|\boldsymbol{x} \in B_j] = \frac{\pi_2 P_2(B_j)}{P(B_j)}$
3. Determine a minimum support plane partition

$$D_j = \{\lambda \in \mathbb{R} : ||\lambda - w_j|| < ||\lambda - w_h||, h \neq j\} \tag{11}$$

4. Update the partition by $B_j = \lambda^{-1}(D_j)$, where by if

$$\lambda(\boldsymbol{x}) = \frac{\pi_2 f_2(\boldsymbol{x})}{f(\boldsymbol{x})} \in D_j, \text{ then } \boldsymbol{x} \to B_j \tag{12}$$

5. Repeat 2-4 steps until a convergence criterion is met