Distance Weighted Discrimination

Adam Rahman
February 8, 2019

Given two sets of points in a matrix $X \in \mathbb{R}^n$ with associated class variables $[-1,1]$ in $Y = \text{diag}(y)$, distance weighted discrimination ([1]) seeks to classify the points into two distinct subsets by finding a hyperplane between the two sets of points. Mathematically, the distance weighted discrimination problem seeks a hyperplane defined by a normal vector, $\omega$, and position, $\beta$, such that each element in the residual vector $r = YX^T \omega + \beta y$ is positive and large. Since the class labels are either 1 or -1, having the residuals be positive is equivalent to having the points on the proper side of the hyperplane.

Of course, it may be impossible to have a perfect separation of points using a linear hyperplane, so an error term $\xi$ is introduced. Thus, the perturbed residuals are defined to be

$$r = YX^T \omega + \beta y + \xi$$

Distance Weighted Discrimination solves the following optimization problem to find the optimal hyperplane[1].

$$\begin{align*}
\text{minimize} & \sum_{i=1}^{n} \frac{1}{r_i} + C \xi^T \\
\text{subject to} & r = YX^T \omega + \beta y + \xi \\
& \omega^T \omega \leq 1 \\
& r \geq 0 \\
& \xi \geq 0
\end{align*}$$

where $C > 0$ is a penalty parameter to be chosen.

The function dwd takes as input two $n \times p$ matrices $X_1$ and $X_2$ containing the points to be separated, as well as a penalty term $C \geq 0$ penalizing the movement of a point on the wrong side of the hyperplane to the proper side, and returns the optimal solution using sq1p to the distance weighted discrimination problem.

R> out <- dwd(X1,X2,C)

Numerical Example

Consider two point configurations - An and Ap - which we would like to classify using distance weighted discrimination. Each point configuration is a matrix containing 50 points in three dimensional space.

R> data(Andwd)
R> data(Apdwd)
R> d <- ncol(Andwd)
R> head(Andwd)

```
V1    V2   V3
[1,] 0.214 -1.577 -1.525
```

R> data(Apdwd)
R> head(Apdwd)

```
V1    V2   V3
[1,] 0.214 -1.577 -1.525
```
Distane weighed discrimination is used to seperate these two configurations by specifying an appropriate penalization term. Here, we will take a value of 0.5.

```r
R> out <- dwd(Apdwd,Andwd,0.5)
```

The information defining the seperating hyperplane ($\omega$ and $\beta$) is stored in the $X$ output vector.

```r
X <- out$X
```

```r
omega <- X[[1]][2:(d+1)]
beta <- X[[1]][d+3]
```

```r
omega
  [,1]
[1,] 0.6567689
[2,] 0.4857645
[3,] 0.5767907
```

```r
beta
[1] -0.7520769
```

References