Contents

Classification using data depth analysis

Abhay kumar & Akshay kumar

June 27, 2014

1 Introduction to data depth

Data Depth has emerged as an important tool for non-parametric analysis and inference of multivariate data [16]. Data depth technique facilitates systematic way of ordering the multivariate data. This is known as "center-outward ordering" [19] or "outlyingness" [11]. Generally, we analyse multivariate data on the presumption of following normal or near-normal distribution, but it is not always true. For analysing univariate data, various statistical moments, like location, scale, skewness, kurtosis etc are used. Straightforward extension of these statistical moments of univariate case is widely used for determining characteristics of multivariate data [11]. These moments are difficult to realize conceptually or graphically. In many cases various moments don't exist, making this approach inappropriate. Simple graphs or contours based on data depth analysis provide more intuitive visualisation of distributional properties as compared with statistical moments.

1.1 Desirable properties of Data Depth Functions

For using depth functions as an efficient tool for center outward ordering of points in $\mathbf{R}^d$, it should ideally satisfy the following properties. Let D(.;.): $\mathbf{R}^d \times \mathscr{F} \to \mathbf{R}^1$ be mapping of a depth function of points, whose distribution is given by $\mathscr{F}$.

P1: Affine Invariance
Under affine transformation of coordinate system, the depth of a point $x \in \mathbf{R}^d$ remains unchanged [19]. Affine transformation preserves collinearity and ratio of distances. It includes geometric contraction, expansion, dilation, reflection, rotation, translation etc. and all their possible combinations. Let the transformation be represented as $x \mapsto Ax + b$, where A is an invertible $d \times d$ matrix and b is a d-dimensional column vector, then

$$D(Ax + b; F_{Ax+b}) = D(x; F_x) \tag{1}$$

P2: Maximality at center
Any depth function should attains its maximum at "Center" of the distribution. Center is defined as a point of symmetry w.r.t some notion of symmetry like central symmetry, angular symmetry, halfspace symmetry etc [19].

$$D(\phi; F) = \sup_{x \in \mathbf{R}^d} D(x; F) \tag{2}$$

holds for any distribution $F \in \mathscr{F}$ having its center at ϕ.

P3: Monotonous w.r.t to the deepest point
As a point x moves away from the center along a ray passing through the center, the value of depth function should decrease monotonically [19].

$$D(x; F) \leq D(\phi + \alpha(x - \phi); F) \qquad \text{holds for } \alpha \in [0, 1] \tag{3}$$

P4: Vanishes at Infinity
The depth function should approach zero as euclidean distance of the point from the center approaches infinity [19].

$$D(x; F) = 0 \qquad \text{as } ||x|| \to \infty \tag{4}$$

1.2 Some Standard Depth Functions

Some of the most familiar depth function used for multivariate data analysis are as following:-

1.2.1 Halfspace Depth

Halfspace Depth (HD) [8, 13] of a point $x \in \mathbf{R}^d$ w.r.t distribution having probability measure P on $\mathbf{R}^d$ is defined as the minimum of probability mass of any closed halfspace containing x.

$$HD(x; P) = \inf_H \{P(H) : H \text{ is a closed halfspace containing } x \in \mathbf{R}^d\} \tag{5}$$

This is also known as "*Tukey Depth*" or "*location depth*" [9]. Halfspace depth of a point x with respect to an empirically distributed data set in $\mathbf{R}^d$ is defined as the minimum fraction of data points lying on either side of any possible hyperplane, drawn passing through x. Let us visualise halfspace depth at test point x for two-dimensional data set having four empirically distributed points [2] For the given distribution F, HD(x;F)=1/4 = 0.25

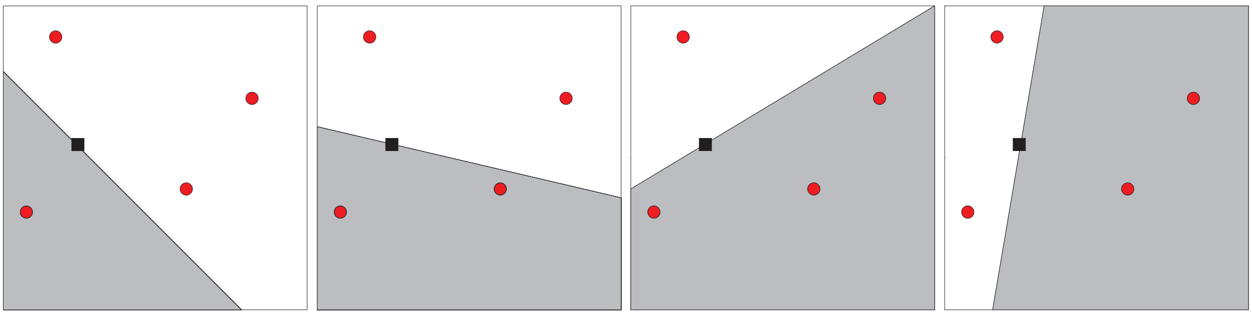

Figure 1: Illustration for finding Halfspace Depth [**?**]

1.2.2 Simplicial Depth

Simplicial Depth (SD) or Liu Depth [10] of a point $x \in \mathbf{R}^d$ relative to a probability measure P on $\mathbf{R}^d$ is defined as the probability that x belongs to a random simplex in $\mathbf{R}^d$.

$$SD(x; P) = P(x \in S[X_1, X_2, X_3, ..., X_{d+1}]) \tag{6}$$

(Note:A simplex in d-dimension is a d-dimensional polytope, which is convex-hull of its d+1 vertices.It is generalisation of triangle in 2D or tetrahedron in 3D to higher dimensions [18])

For an empirically distributed data set, the simplicial depth is defined as ratio of number of simplices containing x to total number of possible simplices.

$$SD(x; X_1, X_2.., X_n) = \frac{\text{Number of simplices containing } x}{\text{Total number of simplices}} \tag{7}$$

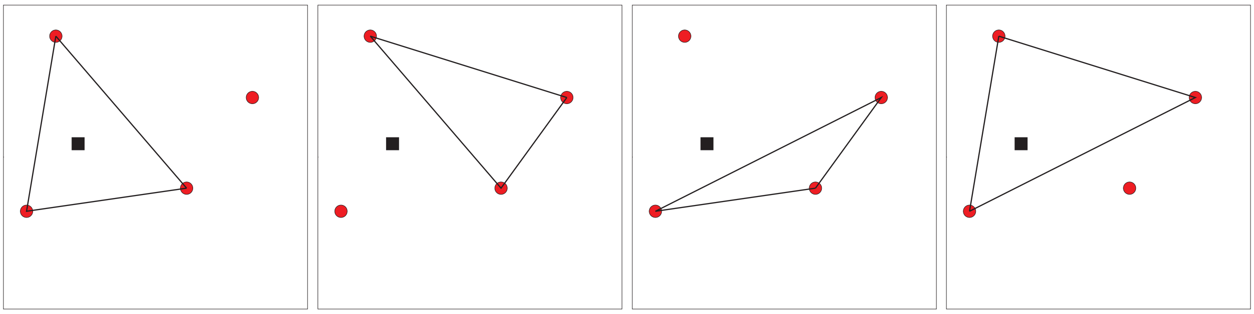

Figure 2: Illustration for calculating Simplicial Depth [2]; SD $= \frac{2}{\binom{4}{3}} = 0.5$

1.2.3 Mahalanobis Depth

Mahalanobis distance-It is a descriptive statistic which provides the relative distance between points $x, y \in \mathbf{R}^d$ w.r.t (with respect to) a $d \times d$ positive-definite matrix.(notes:for any non-zero z column vector $Z^T MZ$ is positive for a positiv-definite matrix).The mahalanobis distance of a point x from a data sample with mean μ and covariance matrix S is defined as [17] :-

$$D_m(x) = \sqrt{(x-\mu)^\top S^{-1}(x-\mu)} \tag{8}$$

Let us consider a random sample $\{X_1, X_2, \ldots, X_n\}$ from multivariate distribution, where second moment exists.For the given sample of size n, sample version of Mahalanobis Depth of X_i, (MD_i) is defined as [5, 19] :-

$$MD_i = \frac{1}{1 + \underbrace{(X_i - \bar{X})^\top S^{-1}(X_i - \bar{X})}_{D_m^2(X_i)}} \tag{9}$$

where, $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$

1.2.4 L_1 depth

L_1 depth (L_1) [9, 14] of a point x with respect to a random sample, $S = \{X_1, X_2, \ldots, X_n\}$ in $\mathbf{R}^d$ is defined as one minus the average of the unit vectors, directed from x towards all observation in S.

$$L_1(x, S) = 1 - \left\|\frac{1}{n}\sum_{i=1}^n u_i(x)\right\| \tag{10}$$

where, $u_i(x) = \frac{x - X_i}{\|x - X_1\|}$. Since, $u_i(x)$ are unit vectors for $x \neq X_i$ [14],

$$0 \leq L_1(x, S) \leq 1 \tag{11}$$

For x lying far away from the center, all unit vectors gets summed up i.e. $\lim_{||x|| \to \infty} ||\bar{u}(x)|| = 1$ and hence L_1 depth approaches zero [14].

$$\lim_{||x|| \to \infty} L_1(x, S) = 0 \tag{12}$$

For points near center, value of L_1 depth is higher as unit vector cancel each other.

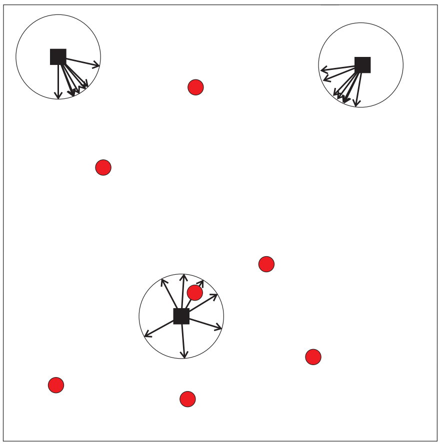

Figure 3: Illustration for L_1 depth[1]

1.2.5 Oja depth

(Note: Please include this in types of depth function as it is already defined in R and we have done misclassification using it also.)
Oja depth of a point $x \in \mathbf{R}^d$ w.r.t to a sample $S = \{X_1, X_2, \ldots, X_n\}$, whose distribution is given by F, is defined as [11]

$$O(x; F) = \frac{1}{1 + E_F[volume(S[x, X_1, \ldots, X_d])]} \tag{13}$$

where, $S[x, X_1, \ldots, X_d]$ represents the closed simplex with vertices x and random d points from the sample S.
For an empirically distributed sample in $\mathbf{R}^2$, Oja depth of a test point x is defined as [?]

$$O(x, F) = \left[1 + \frac{1}{\binom{n}{2}} \sum_{0<i<j<n} \text{Area of traingle with vertices } x, X_i, X_j \right]^{-1} \tag{14}$$

2 Classification using Data depth

Suppose there are n points, $\{X_1, X_2, \ldots X_n\}$. These observations have been categorised into l 'Populations' or 'Class' namely, $P_0, P_1, \ldots, P_{l-1}$ containing $n_0, n_1, \ldots, n_{l-1}$ points respectively (note: $n_1 + n_2 + \ldots + n_{l-1} = n$). Given a test point or a test cluster, we will classify it to one of the classes using depth-based classifier. A classifier is a function $m : X \to Y$ (where, $x \in \mathbf{R}^d$ and $Y \in \{0, 1, \ldots, l-1\}$) that associates a test point x with its corresponding "class" Y.

2.1 Maximum-depth classifier

Maximum-depth classifier classifies a test point x into the class, with respect to which it has highest depth. Let P_j denote j^{th} population. Here, we can use any depth function, like halfspace depth, simplicial depth, Oja depth, mahalanobis depth, L_1 depth etc. to classify the test point. [10, 6].

$$\hat{m}_D(x; P_0, P_1) = \mathbb{I}\Big[D(x, P_1) > D(x, P_0)\Big] \tag{15}$$

Generalisation of Maximum-depth classifier for a test point in L populations Above defined maximum-depth classifier for two-class problem can be generalised for the multi-class problem.

$$\hat{m}_D(x; P_0, \ldots, P_j, \ldots, P_{l-1}) = \underset{j}{\operatorname{argmax}} \Big[D(X, P_j)\Big] \tag{16}$$

It will identify the class out of l classes with respect to which the test point x has highest depth, and classify the test point to that class.

2.2 Proposed Maximum-depth based classifier for a test cluster in two population

Suppose we have two data clusters namely Population1(P_1) and Population0 (P_0) and a test cluster ($X = \{x_1, x_2, .., x_m\}$), having m points. Then the test cluster is said to be belonging to (P_1) if the number of points in X, having more depth relative to (P_1) than that to (P_0) is greater than the number of points in X, having more depth relative to (P_0) than that to (P_1) This is formulated as:-

$$\hat{h}_D(X; P_0, P_1) = \mathbb{I}\left[\underbrace{\sum_{i=1}^{m} \mathbb{I}\left[D(X_i, P_1) > D(X_i, P_0)\right]}_{n(X,P_1)} > \underbrace{\sum_{i=1}^{m} \mathbb{I}\left[D(X_i, P_0) > D(X_i, P_1)\right]}_{n(X,P_0)}\right] \tag{17}$$

where, $n(X, P_j)$ represents number of points of test cluster X having highest depth w.r.t P_j ($\forall j = \{0, 1\}$).
$\hat{h}_D(X; P_0, P_1) = 1 \Rightarrow$ Test cluster belongs to P_1
$\hat{h}_D(X; P_0, P_1) = 0 \Rightarrow$ Test cluster belongs to P_0

Proposed generalisation of maximum-depth classifier for a test cluster in L populations:

Maximum-depth classifier for a test cluster for two-class problem can be easily generalised for multi-class problem.

$$\hat{h}_D(X; P_1, P_2, ...P_j..., P_l) = \underset{j}{\operatorname{argmax}} \Big[n(X, P_j)\Big] \tag{18}$$

where, $n(X, P_j)$ denotes number of data points of the test cluster X having highest depth w.r.t P_j ($\forall j = \{0, 1, \ldots, l-1\}$). If maximum number of points of test cluster belongs to P_j then the test cluster is classified to P_j.

2.3 *K*NN depth-based Classifier

Depth based neighbourhood: For defining depth based neighbourhood, we exploit maximality at center(P2) of depth functions or symmetrization with respect to x. Firstly we construct an x-outward ordering of points. Symmetrization construction involves adding to the original sample $(X_1, X_2.X_3, .., X_n)$, their reflections w.r.t. to x i.e.$(2x - X_1, 2x - X_2, 2x - X_3, .., 2x - X_n)$. As a result of symmetrization, x becomes the deepest point of the final sample.Let $\mathbf{R}_x^{\beta(n)}$, where $\beta = k/n$ denote the smallest depth based neighbourhood that contains atleast β proportion of total sample points [12].

***k*-Nearest Neighbours (*k*NN) classifier**, $\hat{m}_D^{\beta(n)}(x)$ classifies a test point x into class '1' iff there are more points from class '1' than from class '0' in the smallest depth based neighbourhood of x that contains total k data points [12].

$$\hat{m}_D^{\beta(n)}(x; P_0, P_1) = \mathbb{I}\left[\underbrace{\sum_{i=1}^{n} \mathbb{I}\left[Y_i = 1\right] W_i^{\beta(n)}(x)}_{\hat{\eta}_1} > \underbrace{\sum_{i=1}^{n} \mathbb{I}\left[Y_i = 0\right] W_i^{\beta(n)}(x)}_{\hat{\eta}_0}\right] \tag{19}$$

with $W_i^{\beta(n)}(x) = \frac{1}{K_x^{\beta(n)}} \mathbb{I}\left[X_i \in R_x^{\beta(n)}\right]$ where $K_x^{\beta(n)} = \sum_{j=1}^{n} \mathbb{I}\left[X_j \in R_x^{\beta(n)}\right]$ denotes total number of points in $R_x^{\beta(n)}$. Y_i denotes the class to which the data point X_i belongs to. Here, $\hat{\eta}_1$ and $\hat{\eta}_0$ represent fraction of points in $\mathbf{R}_x^{\beta(n)}$ that belongs to class '1' and class '0' respectively.
(NOTE: If $\hat{\eta}_1 > \hat{\eta}_0$ then $\hat{\eta}_1 > \frac{1}{2}$ ($\because \hat{\eta}_1 + \hat{\eta}_0 = 1$, always))

$$\hat{m}_D^{\beta(n)}(x) = \begin{cases} 1, & \text{if } \hat{\eta}_1 > \hat{\eta}_0 \\ 0, & \text{if } \hat{\eta}_1 < \hat{\eta}_0 \end{cases} \tag{20}$$

$\hat{m}_D^{\beta(n)}(x) = 1 \Rightarrow$ Test point belongs to P_1
$\hat{m}_D^{\beta(n)}(x) = 0 \Rightarrow$ Test point belongs to P_0

2.4 Proposed *K*NN depth-based Classifier for a cluster in two class problem

KNN depth-based Classifier can also be used to classify a test cluster ($X = \{x_1, x_2, \ldots, x_j, \ldots, x_m\}$), having m points. Here, we will use KNN depth-based Classifier [Eqn:19] to classify each point of the test cluster. $\hat{m}_D^{\beta(n)}(x_j)$ is KNN depth-based Classifier which classifies j^{th} point of the test cluster. It will be 1 if that point belongs to class '1'. Hence, $\sum_{j=1}^{m} \hat{m}_D^{\beta(n)}(x_j)$ signifies total number of points of the test cluster belonging to class '1'. Similarly $1 - \hat{m}_D^{\beta(n)}(x_j)$ will be 1 if x_j belongs to class '0' and hence, $\sum_{j=1}^{m}[1 - \hat{m}_D^{\beta(n)}(x_j)]$ signifies the total number of points of the test cluster belonging to class '0'. If the total number of points of test cluster belonging class '1' is more than those belonging to class '0', we will classify the test cluster to class '1' and vice versa.

$$\hat{m}_D^{\beta(n)}(X;P_0,P_1) = \mathbb{I}\Bigg[\sum_{j=1}^{m}\overbrace{\mathbb{I}\Bigg(\sum_{i=1}^{n}\mathbb{I}\left[Y_i=1\right]W_i^{\beta(n)}(x) > \sum_{i=1}^{n}\mathbb{I}\left[Y_i=0\right]W_i^{\beta(n)}(x)\Bigg)}^{\hat{m}_D^{\beta(n)}(x_j)} \tag{21}$$

$$> \sum_{i=1}^{m}\underbrace{\mathbb{I}\Bigg(\sum_{i=1}^{n}\mathbb{I}\left[Y_i=0\right]W_i^{\beta(n)}(x) > \sum_{i=1}^{n}\mathbb{I}\left[Y_i=1\right]W_i^{\beta(n)}(x)\Bigg)}_{1-\hat{m}_D^{\beta(n)}(x_j)}\Bigg]$$

2.5 Proposed KNN depth-based Classifier for a test cluster in multi-class problem

Let us consider a multi-class problem for a given sample $(X_1, X_2, \ldots, X_i, \ldots, X_n)$ where we have to classify a test cluster X $(= \{x_1, x_2, \ldots, x_j \ldots, x_m\})$ into any one of the l populations, namely $(P_0, P_1, \ldots, P_k, \ldots, P_{l-1})$ using KNN depth-based Classifier.

$$\hat{m}_D^{\beta(n)}(X;P_0,P_1,...P_k...,P_{l-1}) = \underset{k}{\operatorname{argmax}}\Big[\sum_{j=1}^{m}\mathbb{I}\Big(\sum_{i=1}^{n}\mathbb{I}[Y_i=k]W_i^{\beta(n)}(x)\Big)\Big] \tag{22}$$

2.6 Properties of proposed classifiers for cluster

1.Affine Invariance: Above Depth-based classifiers are defined as a function of various depth functions, like Halfspace Depth, Simplicial Depth, Oja depth etc. Since these depth functions are affine invariant [P1], related classifier would also be affine invariant.
Proof:
Let $X \mapsto Y$ and $Y = AX + b$ represent the affine transformation. Let $\hat{h}_D(x; P_{1x}, P_{0x})$ denote the maximum-depth classifier for original sample and $\hat{h}_D(y; P_{1y}, P_{0y})$ denote the maximum-depth classifier for the affine-transformed sample.

$$\begin{aligned}
\hat{h}_D(y;P_{1y},P_{0y}) &= \mathbb{I}\left[D(y,P_{1y}) > D(y,P_{0y})\right] \\
&= \mathbb{I}\left[D(x,P_{1x}) > D(x,P_{0x})\right] \quad (\because \text{(D(.;.) is affine invariant)}) \\
&= \hat{h}_D(x;P_{1x},P_{0x})
\end{aligned} \tag{23}$$

Hence, all maximum-depth classifiers (note: whether for a test point or a test cluster) are affine invariant.

2.Robustness Analysis: Most statstical depth functions will give zero depth to any outlier point x, making it impossible to be classified. Symmetrisation construction w.r.t x, involved in defining depth-based neighbourhood makes x the center of the resulting data sample. Since kNN depth-based classifier is based on depth-based neighbourhood, it is robust to outliers [12].

2.7 Pros and cons of depth function

Random vector X in $\mathbf{R}^d$ is said to be
(a)**Centrally symmetric** : about θ if $X - \theta \stackrel{d}{=} \theta - X$ (where $\stackrel{d}{=}$ denotes "equal in distribution")[19].
(b)**Angularly Symmetric** : about θ if $(X-\theta)/||X-\theta|| \stackrel{d}{=} (\theta - X)/||X-\theta||$ [19].
(c)**Halfspace symmetry** : about θ if $P(X \in H \geq 1/2)$ for every closed halfspace containing θ [19].

Halfspace depth:satisfies all four properties, but its computation time is relatively higher [15]. It is also robust to outliers [4].
Simplicial depth may fail to satisfy the 'monotonicity' property [P3] for centrally symmetric discrete distributions. It may also fails to satisfy the 'Maximality' property [P2] for Halfspace symmetric discrete distributions [19]. Its computation is also slow.[7]
Mahalanobis depth is not 'robust' since it depends on means and covariance, which are non robust measures. Existence of second moment is must for defining mahalanobis depth. It may fail to satisfy 'maximality' property for angularly symmetric distributions [19].
L_1 **depth**: is computationally much faster and easier. But it is not robust for classification purpose as L_1 depth of a test point far away from the convex hull containing sample points is positive. It hardly attains zero for any outliers.
Oja depth: It satisfies all four properties. Generally, it is robust. But it may not be robust for certain sample, having a small breakdown point [3]. (Note: Higher the breakdown point of an estimator, more robust it is.)

Depth functions	Pros	Cons
Halfspace Depth	It satisfies all four properties	Its computation time is relatively higher
Simplicial Depth	It may fail to satisfy the 'monotonicity' property [P3] for centrally symmetric discrete distributions. It may also fails to satisfy the 'Maximality' property [P2] for Halfspace symmetric discrete distributions	Its computation is also slow.
Mahalanobis Depth	Its computation is fast as well.	It is not 'robust' since it depends on means and covariance, which are non robust measures. Existence of second moment is must for defining mahalanobis depth. It may fail to satisfy 'maximality' property for angularly symmetric distributions
L_1 Depth	It is computationally much faster and easier	It is not robust for classification purpose as L_1 depth of a test point far away from the convex hull containing sample points is positive. It hardly attains zero for any outliers.
Oja Depth	It satisfies all four properties. Generally, it is robust.	But it may not be robust for certain sample, having a small breakdown point. (Note: Higher the breakdown point of an estimator, more robust it is.The breakdown point of an estimator is the fraction of data that must be moved to infinity before the estimator is also moved to infinity.)

2.8 Improvements in depth based classifier

Maximum-depth classifier has low computation time as compared to KNN depth based classifier. On a contrary, KNN depth based classifier is highly robust as compared to maximum-depth based classifier. For optimising computation time and robustness, we have introduced a weighted depth based classifier. We will exploit the 'low computation time' of maximum-depth classifier and 'robustness' of KNN depth based classifier. We will use maximum-depth classifier for points having higher depth value (say upto top 80%) and KNN depth based classifier for the remaining points (outliers).

$$\hat{W}_D(X; P_0, P_1) = \hat{h}_D(X; P_0, P_1)\mathbb{I}\Big[D(X) \geq 0.2 * MD\Big] + \hat{m}_D^{\beta(n)}(X; P_0, P_1)\mathbb{I}\Big[D(X) \geq 0.2 * MD\Big] \quad (24)$$

Where MD denotes the maximum depth value in a given sample. For points having depth greater than 20% value of MD, we will use maximum-depth classifier. And for points having depth upto 20% of MD, we will use robust KNN depth based classifier. Here 20% is arbitrarily chosen benchmark or threshold depth value. We should reduce this threshold value to optimise computation time, but compromising robustness. Similarly increasing this threshold value improves robustness, but reduces computation time. Hence, there is a trade-off between computation time and robustness depending on threshold value.

3 Problem statement

(Please elaborate the problem statement) Let us consider four source, S_1, S_2, S_3, S_4 located at vertices of a rectangle($l_1 \times l_2$, assuming $l_1 > l_2$ without any loss of generality) and four test nodes, N_1, N_2, N_3, N_4 in an isotropic medium. Given the test vectors V_1, V_2, V_3, V_4, we have to classify it to which node do these vectors belong to.

Let us define a Population matrix, P =

$$\begin{pmatrix} P_{11} & P_{21} & P_{31} & P_{41} \\ P_{12} & P_{22} & P_{32} & P_{42} \\ P_{13} & P_{23} & P_{33} & P_{43} \\ P_{14} & P_{24} & P_{34} & P_{44} \end{pmatrix} \tag{25}$$

where, P_{kl} = Population of observations obtained at node N_k due to source, $S_l \; \forall \; k, l = \{1, 2, 3, 4\}$. Observation at each node will constitute of four vectors, obtained from each source. Suppose the given vectors, $V_1 = \begin{pmatrix} v_{11} \\ v_{12} \\ \vdots \\ v_{1m_1} \end{pmatrix}, V_2 = \begin{pmatrix} v_{21} \\ v_{22} \\ \vdots \\ v_{2m_2} \end{pmatrix}, V_1 = \begin{pmatrix} v_{31} \\ v_{32} \\ \vdots \\ v_{3m_3} \end{pmatrix}, V_1 = \begin{pmatrix} v_{41} \\ v_{42} \\ \vdots \\ v_{4m_4} \end{pmatrix}$, have m_1, m_2, m_3 and m_4 points respectively. We have to classify these vectors to the node it belong to.

(explaining our concept): We will classify each point in these test vectors into different sixteen populations,P_{kl} according to maximum-depth classifier. If total number of points belonging to P_{11}, P_{12}, P_{13} and P_{14} is maximum, then we will conclude that the given vectors corresponds to node N_1.Similarly If total number of points belonging to P_{21}, P_{22}, P_{23} and P_{24} is maximum, then we will conclude that the given vectors corresponds to node N_2 , and so on.Let x_{ij} represent j^{th} point of i^{th} test vector.

$$\hat{h}_D(V_1, V_2, V_3, V_4; P) = \underset{k}{\operatorname{argmax}} \left[\underbrace{\sum_{i=1}^{4} \underbrace{\sum_{j=1}^{m_i} \underbrace{\sum_{l=1}^{4} \left[\underbrace{\mathbb{I}\Big(\{k,l\} = \underset{\{a,b\}}{\operatorname{argmax}} D(x_{ij}, P_{ab})\Big)}_{=1, \text{ if } x_{ij} \text{ has highest depth w.r.t } P_{kl}} \right]}_{\text{Summation over Population related to } k^{th} \text{ test node}}}_{\text{Summation over } m_i \text{ points of } i^{th} \text{ vector}}}_{\text{Summation over all four test vectors } V_1, V_2, V_3, V_4} \right] \tag{26}$$

References

[1] L1 depth @http://depth.johnhugg.com/DepthExplorerICORSslides.pdf.

[2] Stanislav nagy, data depth slides.

[3] Greg Aloupis, Stefan Langerman, Michael Soss, and Godfried Toussaint. Algorithms for bivariate medians and a fermat–torricelli problem for lines. *Computational geometry*, 26(1):69–79, 2003.

[4] Gerda Claeskens, Mia Hubert, Leen Slaets, and Kaveh Vakili. Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109(505):411–423, 2014.

[5] Maman A Djauhari and Rian F Umbara. A redefinition of mahalanobis depth function. *Malaysian Journal of Fundamental and Applied Sciences*, 3(1), 2008.

[6] Anil K. Ghosh and Probal Chaudhuri. On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, 11(1):1–27, 01 2005.

[7] Daniel Hlubinka and Ondrej Vencalek. Depth-based classification for distributions with non-convex support. *Journal of Probability and Statistics*, 2013, 2013.

[8] Joseph L Hodges. A bivariate sign test. *The Annals of Mathematical Statistics*, pages 523–527, 1955.

[9] John Hugg, Eynat Rafalin, Kathryn Seyboth, and Diane L Souvaine. An experimental study of old and new depth measures. In *ALENEX*, pages 51–64. SIAM, 2006.

[10] Regina Y Liu et al. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.

[11] Regina Y Liu, Jesse M Parelius, Kesar Singh, et al. Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). *The Annals of Statistics*, 27(3):783–858, 1999.

[12] Davy Paindaveine and Germain Van Bever. Nonparametrically consistent depth-based classifiers. *arXiv preprint arXiv:1204.2996*, 2012.

[13] John W Tukey. Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, volume 2, pages 523–531, 1975.

[14] Yehuda Vardi and Cun-Hui Zhang. The multivariate l1-median and associated data depth. *Proceedings of the National Academy of Sciences*, 97(4):1423–1426, 2000.

[15] O Vencálek. Generalization of halfspace depth.

[16] Ondřej Vencálek. Concept of data depth and its applications. *Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica*, 50(2):111–119, 2011.

[17] Wikipedia. Mahalanobis distance — wikipedia, the free encyclopedia, 2014. [Online; accessed 17-June-2014].

[18] Wikipedia. Simplex — wikipedia, the free encyclopedia, 2014. [Online; accessed 17-June-2014].

[19] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 04 2000.