Bagadi, R. (2017). Dynamic Thresholding For Linear Binary Classifiers. {Version 2} ISSN 1751-3030. *PHILICA.COM Article number 1146.*
http://www.philica.com/display_article.php?article_id=1146

Dynamic Thresholding For Linear Binary Classifiers. {Version 2} ISSN 1751-3030

Author:

Ramesh Chandra Bagadi
Data Scientist
INSOFE (International School Of Engineering),
Hyderabad, India.
rameshcbagadi@uwalumni.com
+91 9440032711

Research Manuscript

Abstract

In this research investigation, the author has detailed a novel method of finding the Thresholding for Linear Binary Classifiers.

Theory

Method 1:

If y_l (for $l = 1 \; to \; n$) are the points, that have to be divided using a linear binary classifier, we can select the Threshold value y_t using the equation

$$\sum_{i=1}^{m} (y_i - y_t) y_i = \sum_{j=1}^{n-m} |(y_j - y_t)| y_j \qquad \text{Equation 1}$$

with $i \neq j$, $y_i > y_t$, $y_j < y_t$ and $y_i, y_j \in y_l$. But since we do not know y_t, we first order all the y_l in increasing order and choose y_t to be in between the y_l values, i.e., $y_l < y_t < y_{l+1}$ (for $l = 1 \; to \; n-1$). That is for n number of points, we need to choose $(n-1)$ number of domains of y_t.Values of y_t within one of these domains gives us the best y_t, the best being the one which satisfies the above stated equation 1 best.

Method 2:

Case 1:

If x_l, y_l (for $l = 1 \; to \; n$) are the points, that have to be divided using a linear binary classifier, we can select the Threshold value y_t using the equation

$$\sum_{i=1}^{m}(y_i - y_t)(x_t - x_i) = \sum_{j=1}^{n-m}(y_t - y_j)(x_j - x_t) \qquad \text{Equation 2}$$

with $i \neq j$, $y_i > y_t$, $y_j < y_t$, $x_i < x_t$, $x_j > x_t$ and $y_i, y_j \in y_l$. But since we do not know x_t, y_t, we first order all the y_l in increasing order and choose y_t to be in between the y_l values, i.e., $y_l < y_t < y_{l+1}$ (for $l = 1 \; to \; n-1$). We similarly, order all x_l in increasing order and choose x_t to be in between the x_l values, i.e., $x_l < x_t < x_{l+1}$ (for $l = 1 \; to \; n-1$). That is for n number of points, we need to choose $(n-1)$ number of domains (each) of x_t and y_t. Now, we have to choose the values of x_t, y_t within one of their (respective) domains such that they satisfy the above stated equation 2 best.

Case 2:

If y_i (for $i = 1 \; to \; n$) are the points, that have to be divided using a linear binary classifier, we can select the Threshold value y_t using the equation

$$\sum_{i=1}^{m}(y_t - y_i)(x_t - x_i) = \sum_{j=1}^{n-m}(y_j - y_t)(x_j - x_t) \qquad \text{Equation 3}$$

with $i \neq j$, $y_i < y_t$, $y_j > y_t$, $x_i < x_t$, $x_j > x_t$ and $y_i, y_j \in y_l$. But since we do not know x_t, y_t, we first order all the y_l in increasing order and choose y_t to be in between the y_l values, i.e.,

$y_l < y_t < y_{l+1}$ (for $l = 1\; to\; n-1$). We similarly, order all x_l in increasing order and choose x_t to be in between the x_l values, i.e., $x_l < x_t < x_{l+1}$ (for $l = 1\; to\; n-1$). That is for n number of points, we need to choose $(n-1)$ number of domains (each) of x_t and y_t. Now, we have to choose the values of x_t, y_t within one of their (respective) domains such that they satisfy the above stated equation 2 best.

References

1.http://www.philica.com/advancedsearch.php?author=12897
2.http://www.vixra.org/author/ramesh_chandra_bagadi