Linear Data Analysis
Classification - Linear Seperability

Cain Susko

Queen's University
School of Computing

March 8, 2022

a Seperating two Clusters

we will expolre how to seperate clusters by a hyper-plane. Binary clustering uses 1 hyper-plane. multiple clusterings require multiple hyperplanes.

Calssification from Clustering Consider the output of clustering:

$$k \text{ centriods using } L_2 \text{ vector norm } ||\cdot||_2^2$$

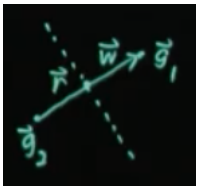
how do we classify a new vector $\vec{u}$? We should first simplify to a binary clustering for classification

- centriod $\vec{g}_1$, 'positive'
- centriod $\vec{g}_2$, 'negative'

the geometry is as such that there is a hyper-plane seperating $\vec{g}_1, \vec{g}_2$
Consider a vector $\vec{v}$ on the boundary: $||\vec{v} - \vec{g}_1||^2 = ||\vec{v} - \vec{g}_2||^2$
which defines the vector that is equidistant from $\vec{g}_1, \vec{g}_2$, thus showing us the hyper-plane seperating the 2 centriods, which is perpendicular to the difference vector of $\vec{g}_1, \vec{g}_2$.

b A Hyper-plane From Cluster Centriods

given the clusters $\vec{g}_1, \vec{g}_2$ and the hyper plane $\mathbb{H}$ with a normal vector $\vec{w}$ and a biased scalar b.



where $\vec{w}$ is the difference vector between $\vec{g}_1, \vec{g}_2$ directed from $\vec{g}_2$ to $\vec{g}_1$. and $\vec{r}$ is the reference point of the hyperplane and is the midpoint between $\vec{g}_1, \vec{g}_2$. Thus:

$$\vec{w} = \vec{g}_1 - \vec{g}_2$$
$$\vec{r} = \frac{\vec{g}_1 + \vec{g}_2}{2}$$

Consider $\vec{v}$ is in $\mathbb{H}$. we know that $\vec{v}$ must satisfy the following equation.

$$\begin{aligned}
\vec{w} \cdot (\vec{v} - \vec{r}) &= 0 \\
\vec{w} \cdot \vec{v} - \vec{w} \cdot \vec{r} &= 0 \\
\vec{w} \cdot \vec{v} + b &= 0 \\
b &= -\vec{w} \cdot \vec{r} \\
b &= -[\vec{g}_1 - \vec{g}_2][\frac{\vec{g}_1 + \vec{g}_2}{2}]
\end{aligned}$$

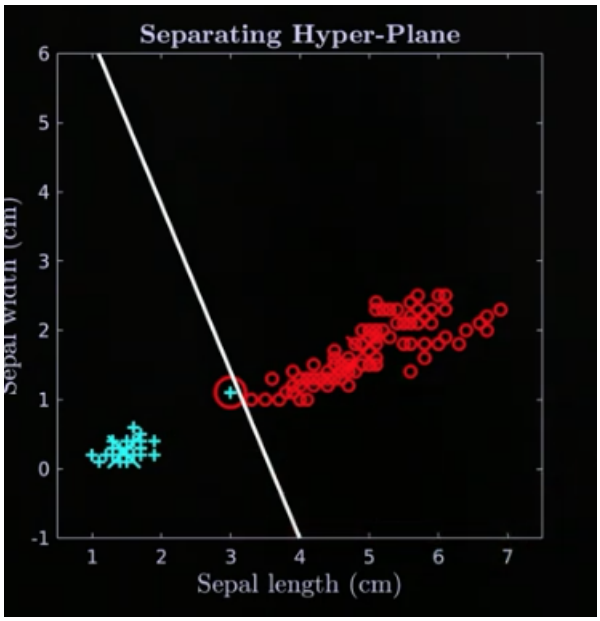
Consider a new vector $\vec{u}$. Is $\vec{u}$ in S_1 or S_2?
We shall say that $\vec{u}$ is in S_1 if and only if:

$$\vec{w}^\top \vec{u} + b \geq 0$$

else, it is in S_2 (as the clustering is binary)

Clustering Iris Data

Now, we shall do this computation on fisher's iris data. after doing k-means this is the clustering with it's corresponding hyper-plane:



we can see by the cyan outlier that the kmeans algorithm made some decisions that a human expert may not support.

c Hyper-planes for Multiple Clusters

let us try to work out the partitioning of a set into 3 clusters. each with a corresponding centriod.

$$S_1, S_2, S_3$$
$$\vec{g}_1, \vec{g}_2, \vec{g}_3$$

we shall now decide how each hyper-plane will seperate each cluster.

$$\mathbb{H}_{12}, \mathbb{H}_{13}, \mathbb{H}_{23}$$

where each plane seperates cluster 1 and 2, 1 and 3, and 2 and 3 respectively. Note that in the above diagram, the reason the hyper-planes are not extended to infinity (this part represented by the dotted line) is because if we want to know if $\vec{u}$ (see diagram) is on the positive side of $\vec{g}_3$ then we check just that; and having the planes extend is visually distracting.

Logical Method A logical method for finding out if $\vec{u}$ is in a specific cluster is:
$\vec{u} \in S_1$ iff:

$$\vec{w}_{12}^{\top}\vec{u} + b_{12} \geq 0 \quad \wedge \quad \vec{w}_{13}^{\top}\vec{u} + b_{13} \geq 0$$

New Operator consider the vector alpha: $\alpha \in \mathbb{R}^m$ we can say that alpha is greater than 0 if and only if all of alpha's entries are also less than 0 using the new operator:

$$\vec{\alpha} \geq \vec{0} \text{ iff } \forall_i(\alpha_i \geq 0)$$

where $\forall$ is the new operator, meaning 'for all'

Gather Values we can now gather the values from the hyper-plane calculations as:

$$\vec{b} = \begin{bmatrix} b_{12} \\ b_{13} \end{bmatrix}$$
$$W_1 = \begin{bmatrix} \vec{w}_{12}^{\top} \\ \vec{w}_{13}^{\top} \end{bmatrix}$$

thus, we can say that $\vec{u} \in S_1$ if and only if:

$$W_1\vec{u} + \vec{b}_1 \geq \vec{0}$$

thus we can see that it is easy to deal with k clusters when trying to find a hyper-plane.

d The Davies-Bouldin Index for Clusters

the DB index is used to score the clustering of data such that

$$\vec{x}_i \in S_1 \vee S_2$$

the DB index measures the closeness and spread-outness of a clustering. Note: S_1 has m_1 members, S_2 has m_2 members. Each set will have a centriod $\vec{g}_1, \vec{g}_2$

cluster distnace We will first measure the distance between $\vec{g}_1, \vec{g}_2$ by doing the calculation

$$\frac{1}{||\vec{g}_1 - \vec{g}_2||}$$

cluster dispursion We then also want to score the cluster by the mean distance within a partition (cluster). This can be calculated as:

$$d_1 = \frac{1}{m_1} \sum_{i=1}^{m_2} ||\vec{x}_i - \vec{g}_1||$$

Thus, the equation for d_2 would be:

$$d_2 = \frac{1}{m_2} \sum_{j=1}^{m_2} ||\vec{x}_j - \vec{g}_2||$$

thus, the measure of the dispursion is:

$$d_1 + d_2$$

and finally, the DB index can be calculated as:

$$DB =^{def} \frac{d_1 + d_2}{||\vec{g}_1 - \vec{g}_2||}$$

thus, the smaller the DB index, the farther apart the clusters are and the less dispursed they are within the cluster.