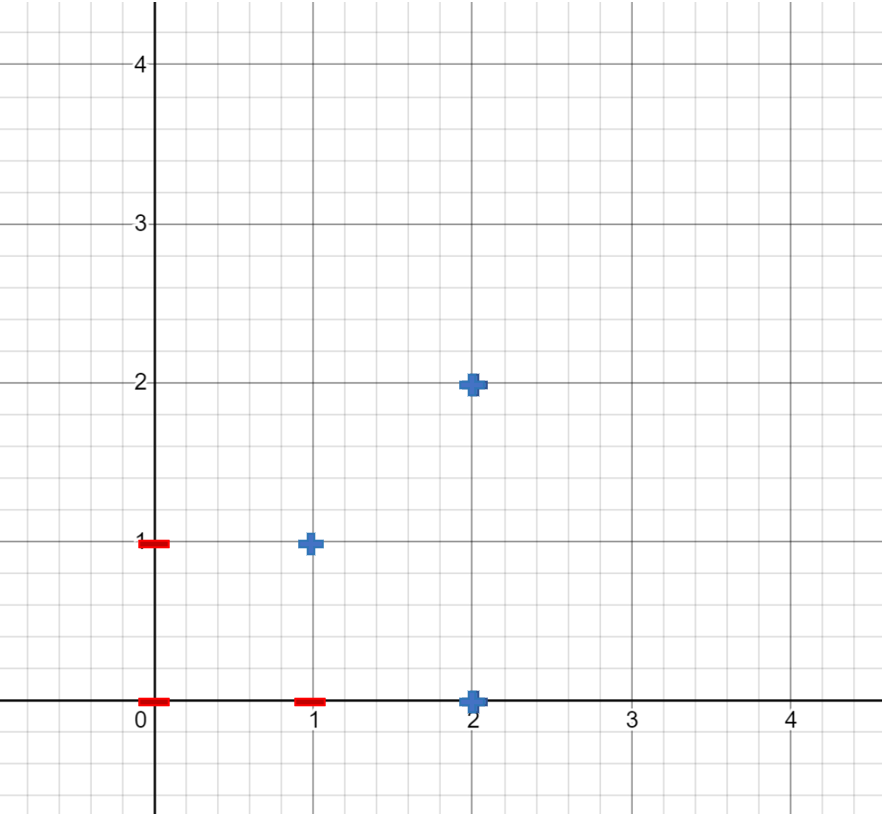
**Data Mining – Assignment 3**

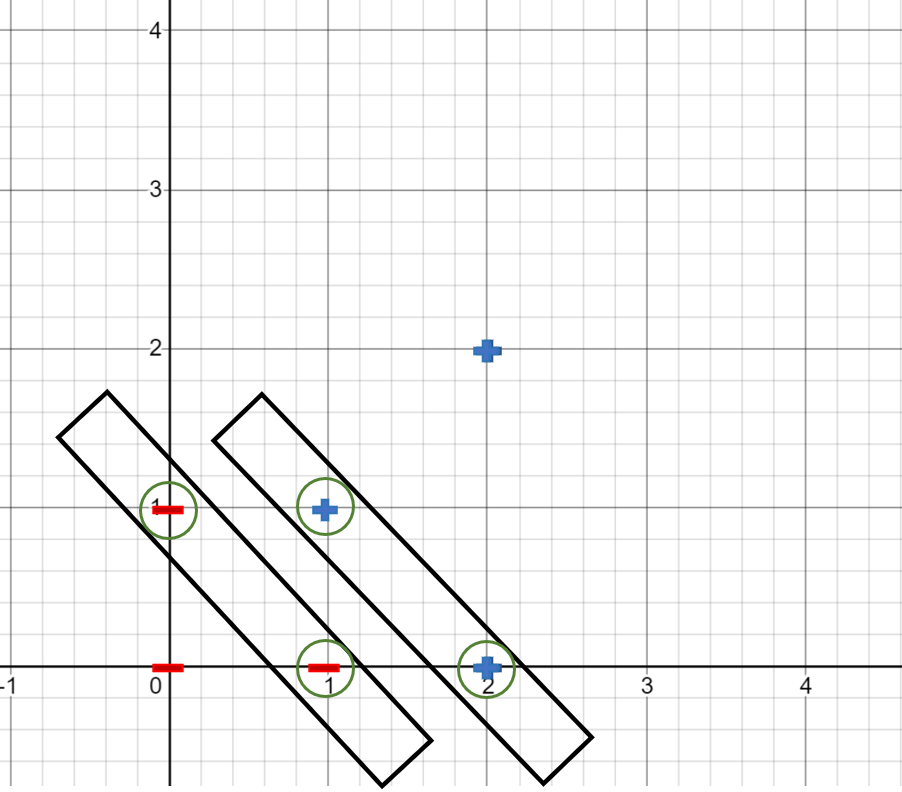
**薛劭杰 1930026143**

Q1.

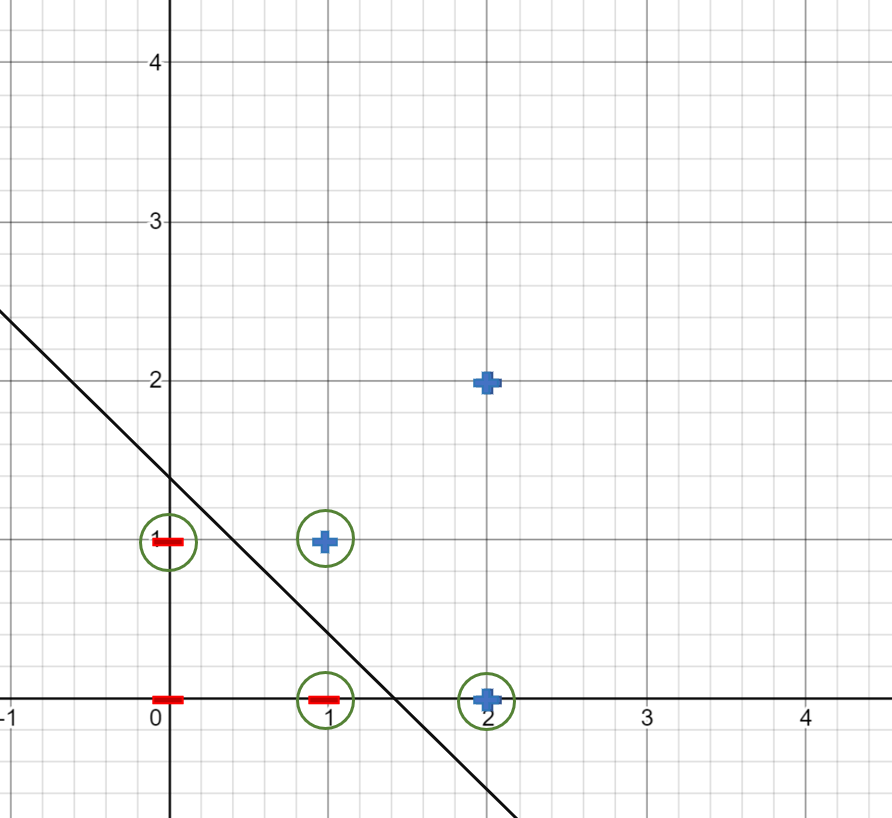
(a)



(2)



(3)



(d)

In the vectors, we select the point1: (1, 0), point3: (1, 1) and point5: (2, 0), and we can the three functions: , there are three unknowns and three the equations, so we can get all results: . So, the optional separating hyperplane of this system is: , the bias is -3.

The margin for each point:

For .

For .

For .

For .

For .

For .

For the minimum margin .

Q2.

Jaccard coefficient: . We set to 1 and to 0

Adopt Jaccard similarity to measure the closeness between samples:

With set to 5:

The largest five Jaccard similarity is .

Then for their classes:

-

There are 4 Mammals and 1 Non-mammals, so classify the test to Mammals.

(b)

We have to compute for each attribute A. First of all, calculate the entropy for the category.

Then focus on

Then focus on

Then focus on

Then focus on

So, split using the attribute .

Give Birth

***Yes***

***No***

Mammals Non-Mammals

It can get the result and it not need to split again.

The Give birth of the test one is , so classify the test to Mammals.

Set the set is split on an attribution A into subset yes and no

Gini index:

Focus on :

Focus on :

Focus on :

Focus on :

Focus on :

(c)

Firstly, we compute the prior probability for each class:

Set , -

,

To derive for , we need to compute the following:

,

,

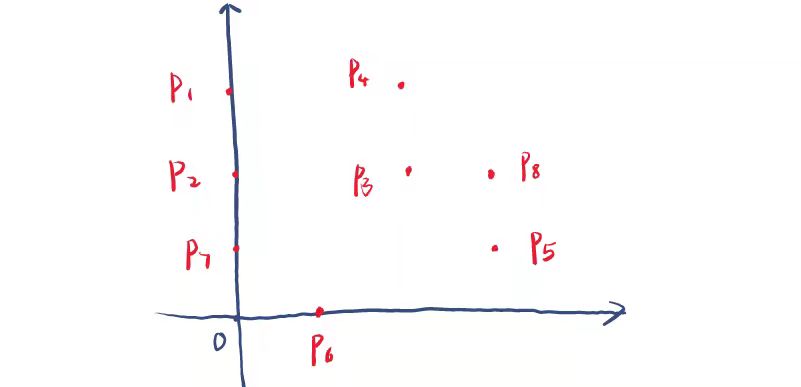
,

,

Given the previous probabilities, we obtain:

, classify the test to () Mammals.

Q3.



The eight points in the coordinate system like above picture:

Firstly, we can calculate the Euclidean distance for each point:

For point , ,, , , , .

For point , ,, , , .

For point , ,, , .

For point , ,, .

For point , ,.

For point , .

For point .

and . For each point, select the other points which distance is smaller

As

are core points

are border points

is a noise point

Now run DBSCAN algorithm:

For each visited points, support start with ( is a core point and a cluster is form ), then mark as the visited and retrieve -,

* Next add (visited), as , no append.
* Next add (visited), as , no append.
* Finish for

Now are visited

For each unvisited point, suppose continue with ( is a core point and a cluster is form ).

mark as the visited and retrieve -,

* Next add (visited), as , no append.
* Next add (visited), as , append to .
* Next add (visited), as , no append.
* Finish for

Now are visited.

For each unvisited point, suppose continue with .

mark as the visited and retrieve -, .

* is not a core point, mark the as noise point

All points are visited: