'initialize the matrix

For col = 1 To 8

For x = 1 To 10

probability(x, col) = 0

Next

Next

List1.Clear

'count the total appear times of each attribute's discrete value

For col = 1 To 8

For x = 0 To 1483

Select Case CDbl(data3(x, col))

Case 1

probability(1, col) = probability(1, col) + 1

Case 2

probability(2, col) = probability(2, col) + 1

Case 3

probability(3, col) = probability(3, col) + 1

Case 4

probability(4, col) = probability(4, col) + 1

Case 5

probability(5, col) = probability(5, col) + 1

Case 6

probability(6, col) = probability(6, col) + 1

Case 7

probability(7, col) = probability(7, col) + 1

Case 8

probability(8, col) = probability(8, col) + 1

Case 9

probability(9, col) = probability(9, col) + 1

Case Else

probability(10, col) = probability(10, col) + 1

End Select

Next

Next

-----------------------------------------------------

Option Explicit

Dim in\_file As String

Dim out\_file As String

Dim atts As String

Dim class As String

Dim x As Integer

Dim y As Integer

Dim col As Integer

Dim data(1484, 10) As String 'data1 used in equal-width,changing value after discretization

Dim data2(1484, 10) As String 'data2 is a sorted matrix, used in equal-frequency

Dim data3(1484, 10) As String 'data3 used in equal-frequency

Dim frequency\_cutpoint(8, 9) As String '(attribute,splitting point)

Dim max(8) As Double 'save 8 attribute max value

Dim min(8) As Double 'save 8 attribute min value

Dim equal\_width(8) As Double 'save interval of each attribute

Dim interval As Integer 'interval is 10

Dim frequency As Integer 'frequency is 148

Dim freq\_probability(11, 10) As String

Dim freq\_probability\_1(11) As Double 'save attribute 1's probability of each discrete value (1-10)

Dim freq\_probability\_2(11) As Double 'save attribute 2's probability of each discrete value (1-10)

Dim freq\_probability\_3(11) As Double 'save attribute 3's probability of each discrete value (1-10)

Dim freq\_probability\_4(11) As Double 'save attribute 4's probability of each discrete value (1-10)

Dim freq\_probability\_5(11) As Double 'save attribute 5's probability of each discrete value (1-10)

Dim freq\_probability\_6(11) As Double 'save attribute 6's probability of each discrete value (1-10)

Dim freq\_probability\_7(11) As Double 'save attribute 7's probability of each discrete value (1-10)

Dim freq\_probability\_8(11) As Double 'save attribute 8's probability of each discrete value (1-10)

Dim freq\_probability\_9(11) As Double 'save attribute 9's probability of each discrete value (1-10)

Dim width\_probability(11, 10) As String

Dim width\_probability\_1(11) As Double 'save attribute 1's probability of each discrete value (1-10)

Dim width\_probability\_2(11) As Double 'save attribute 2's probability of each discrete value (1-10)

Dim width\_probability\_3(11) As Double 'save attribute 3's probability of each discrete value (1-10)

Dim width\_probability\_4(11) As Double 'save attribute 4's probability of each discrete value (1-10)

Dim width\_probability\_5(11) As Double 'save attribute 5's probability of each discrete value (1-10)

Dim width\_probability\_6(11) As Double 'save attribute 6's probability of each discrete value (1-10)

Dim width\_probability\_7(11) As Double 'save attribute 7's probability of each discrete value (1-10)

Dim width\_probability\_8(11) As Double 'save attribute 8's probability of each discrete value (1-10)

Dim width\_probability\_9(11) As Double 'save attribute 9's probability of each discrete value (1-10)

Dim h\_value\_att(10) As Double 'save each attribute's h\_value

Private Sub Partition\_click()

List1.Clear

'check whether the file name is empty

If infile.Text = "" Then

MsgBox "Please input the file names!", , "File Name"

infile.SetFocus

Else

in\_file = App.Path & "\" & infile.Text

'check whether the data file exists

If Dir(in\_file) = "" Then

MsgBox "Input file not found!", , "File Name"

infile.SetFocus

Else

'variale define

Dim each\_row\_output() As String

Open in\_file For Input As #1

'Open App.Path & "\test.txt" For Output As #2

Open App.Path & "\test-freq.txt" For Output As #3

y = 0 'from row 0

'read each line until end of file

Do While Not EOF(1)

Line Input #1, atts

each\_row\_output = Split(atts, " ")

col = 0

For x = 0 To UBound(each\_row\_output)

If each\_row\_output(x) <> "" Then 'if space,ignore

data(y, col) = each\_row\_output(x)

data2(y, col) = each\_row\_output(x)

data3(y, col) = each\_row\_output(x)

col = col + 1

End If

Next

y = y + 1 'goto next row

Loop

MsgBox "Input file has prepared successfully !"

MsgBox "please choose discretization method the right hand side!"

Close #1

End If

End If

End Sub

Private Sub EqualWidth\_click()

List1.Clear

interval = 10

'count width of each attribute

For col = 1 To 8

min(col) = 1

For x = 0 To 1483

If CDbl(data(x, col)) > max(col) Then

max(col) = CDbl(data(x, col))

ElseIf CDbl(data(x, col)) < min(col) Then

min(col) = CDbl(data(x, col))

End If

Next

equal\_width(col) = (max(col) - min(col)) / interval

Next

'print splitting point

For col = 1 To 8

List1.AddItem "Attribute : " & col & " " & " Width= " & equal\_width(col)

For x = 1 To (interval - 1)

List1.AddItem "splitting point = " & min(col) + x \* equal\_width(col)

Next

List1.AddItem ""

Next

'discretization to 1-10, each interval includes splitting point(lower bound)

For col = 1 To 8

For x = 0 To 1483

If CDbl(data(x, col)) >= (min(col) + (interval - 1) \* equal\_width(col)) Then 'the 10th interval

data(x, col) = 10

ElseIf CDbl(data(x, col)) <> -1 Then

For y = 1 To interval - 1 'the 1th to 9th

If data(x, col) < (min(col) + y \* equal\_width(col)) Then

data(x, col) = y

End If

Next

End If

Next

Next

'print

'For col = 1 To 8

'For x = 0 To 1483

'Print #2, data(x, col)

'Next

'Next

'initialize the matrix

For col = 1 To 9

For x = 1 To 10

width\_probability(x, col) = 0

Next

Next

'count the total appear times of each attribute's discrete value

For col = 1 To 8

For x = 0 To 1483

Select Case CDbl(data(x, col))

Case 1

width\_probability(1, col) = width\_probability(1, col) + 1

Case 2

width\_probability(2, col) = width\_probability(2, col) + 1

Case 3

width\_probability(3, col) = width\_probability(3, col) + 1

Case 4

width\_probability(4, col) = width\_probability(4, col) + 1

Case 5

width\_probability(5, col) = width\_probability(5, col) + 1

Case 6

width\_probability(6, col) = width\_probability(6, col) + 1

Case 7

width\_probability(7, col) = width\_probability(7, col) + 1

Case 8

width\_probability(8, col) = width\_probability(8, col) + 1

Case 9

width\_probability(9, col) = width\_probability(9, col) + 1

Case Else

width\_probability(10, col) = width\_probability(10, col) + 1

End Select

Next

Next

For x = 0 To 1483

Select Case data(x, 9)

Case "CYT"

width\_probability(1, 9) = width\_probability(1, 9) + 1

Case "NUC"

width\_probability(2, 9) = width\_probability(2, 9) + 1

Case "MIT"

width\_probability(3, 9) = width\_probability(3, 9) + 1

Case "ME3"

width\_probability(4, 9) = width\_probability(4, 9) + 1

Case "ME2"

width\_probability(5, 9) = width\_probability(5, 9) + 1

Case "ME1"

width\_probability(6, 9) = width\_probability(6, 9) + 1

Case "EXC"

width\_probability(7, 9) = width\_probability(7, 9) + 1

Case "VAC"

width\_probability(8, 9) = width\_probability(8, 9) + 1

Case "POX"

width\_probability(9, 9) = width\_probability(9, 9) + 1

Case Else

width\_probability(10, 9) = width\_probability(10, 9) + 1

End Select

Next

'transform 2-d array to 1-d array

For x = 1 To 10

width\_probability\_1(x) = width\_probability(x, 1)

width\_probability\_2(x) = width\_probability(x, 2)

width\_probability\_3(x) = width\_probability(x, 3)

width\_probability\_4(x) = width\_probability(x, 4)

width\_probability\_5(x) = width\_probability(x, 5)

width\_probability\_6(x) = width\_probability(x, 6)

width\_probability\_7(x) = width\_probability(x, 7)

width\_probability\_8(x) = width\_probability(x, 8)

width\_probability\_9(x) = width\_probability(x, 9)

Next

'calculate probability, 1484 is the number of observation

For x = 1 To 10

width\_probability\_1(x) = width\_probability\_1(x) / 1484

width\_probability\_2(x) = width\_probability\_2(x) / 1484

width\_probability\_3(x) = width\_probability\_3(x) / 1484

width\_probability\_4(x) = width\_probability\_4(x) / 1484

width\_probability\_5(x) = width\_probability\_5(x) / 1484

width\_probability\_6(x) = width\_probability\_6(x) / 1484

width\_probability\_7(x) = width\_probability\_7(x) / 1484

width\_probability\_8(x) = width\_probability\_8(x) / 1484

width\_probability\_9(x) = width\_probability\_9(x) / 1484

Next

h\_value\_att(1) = h\_value(width\_probability\_1)

h\_value\_att(2) = h\_value(width\_probability\_2)

h\_value\_att(3) = h\_value(width\_probability\_3)

h\_value\_att(4) = h\_value(width\_probability\_4)

h\_value\_att(5) = h\_value(width\_probability\_5)

h\_value\_att(6) = h\_value(width\_probability\_6)

h\_value\_att(7) = h\_value(width\_probability\_7)

h\_value\_att(8) = h\_value(width\_probability\_8)

h\_value\_att(9) = h\_value(width\_probability\_9)

End Sub

Private Sub EqualFrequency\_click()

'use bubble-sort to sort attribute and renew data2

Dim Tag As Double

List1.Clear

For col = 1 To 8

For x = 0 To 1483

For y = x To 1483

If CDbl(data2(x, col)) > CDbl(data2(y, col)) Then

'swift

Tag = data2(x, col)

data2(x, col) = data2(y, col)

data2(y, col) = Tag

'change class

class = data2(x, 0)

data2(x, 0) = data2(y, 0)

data2(y, 0) = class

End If

Next y

Next x

Next col

interval = 10

frequency = CInt(1484 / interval) '148

For col = 1 To 8

For x = 1 To 9

frequency\_cutpoint(col, x) = (CDbl(data2(x \* frequency - 1, col)) + CDbl(data2(x \* frequency, col))) / 2

Next

Next

For col = 1 To 8

List1.AddItem "Attribute : " & col

For x = 1 To interval - 1

List1.AddItem "splitting point = " & frequency\_cutpoint(col, x)

Next

List1.AddItem ""

Next

For col = 1 To 8

For x = 0 To 1483

If CDbl(data3(x, col)) >= frequency\_cutpoint(col, interval - 1) Then 'the 10th interval

data3(x, col) = 10

ElseIf CDbl(data3(x, col)) <> -1 Then

For y = 1 To interval - 1 'the 1th to 9th

If data3(x, col) < frequency\_cutpoint(col, y) Then

data3(x, col) = y

End If

Next

End If

Next

Next

'print

'For col = 1 To 1

'For x = 0 To 1483

'Print #3, data3(x, col)

'Next

'Next

'initialize the matrix

For col = 1 To 9

For x = 1 To 10

freq\_probability(x, col) = 0

Next

Next

'count the total appear times of each attribute's discrete value

For col = 1 To 8

For x = 0 To 1483

Select Case CDbl(data3(x, col))

Case 1

freq\_probability(1, col) = freq\_probability(1, col) + 1

Case 2

freq\_probability(2, col) = freq\_probability(2, col) + 1

Case 3

freq\_probability(3, col) = freq\_probability(3, col) + 1

Case 4

freq\_probability(4, col) = freq\_probability(4, col) + 1

Case 5

freq\_probability(5, col) = freq\_probability(5, col) + 1

Case 6

freq\_probability(6, col) = freq\_probability(6, col) + 1

Case 7

freq\_probability(7, col) = freq\_probability(7, col) + 1

Case 8

freq\_probability(8, col) = freq\_probability(8, col) + 1

Case 9

freq\_probability(9, col) = freq\_probability(9, col) + 1

Case Else

freq\_probability(10, col) = freq\_probability(10, col) + 1

End Select

Next

Next

For x = 0 To 1483

Select Case data3(x, 9)

Case "CYT"

freq\_probability(1, 9) = freq\_probability(1, 9) + 1

Case "NUC"

freq\_probability(2, 9) = freq\_probability(2, 9) + 1

Case "MIT"

freq\_probability(3, 9) = freq\_probability(3, 9) + 1

Case "ME3"

freq\_probability(4, 9) = freq\_probability(4, 9) + 1

Case "ME2"

freq\_probability(5, 9) = freq\_probability(5, 9) + 1

Case "ME1"

freq\_probability(6, 9) = freq\_probability(6, 9) + 1

Case "EXC"

freq\_probability(7, 9) = freq\_probability(7, 9) + 1

Case "VAC"

freq\_probability(8, 9) = freq\_probability(8, 9) + 1

Case "POX"

freq\_probability(9, 9) = freq\_probability(9, 9) + 1

Case Else

freq\_probability(10, 9) = freq\_probability(10, 9) + 1

End Select

Next

'transform 2-d array to 1-d array

For x = 1 To 10

freq\_probability\_1(x) = freq\_probability(x, 1)

freq\_probability\_2(x) = freq\_probability(x, 2)

freq\_probability\_3(x) = freq\_probability(x, 3)

freq\_probability\_4(x) = freq\_probability(x, 4)

freq\_probability\_5(x) = freq\_probability(x, 5)

freq\_probability\_6(x) = freq\_probability(x, 6)

freq\_probability\_7(x) = freq\_probability(x, 7)

freq\_probability\_8(x) = freq\_probability(x, 8)

freq\_probability\_9(x) = freq\_probability(x, 9)

Next

'calculate probability, 1484 is the number of observation

For x = 1 To 10

freq\_probability\_1(x) = freq\_probability\_1(x) / 1484

freq\_probability\_2(x) = freq\_probability\_2(x) / 1484

freq\_probability\_3(x) = freq\_probability\_3(x) / 1484

freq\_probability\_4(x) = freq\_probability\_4(x) / 1484

freq\_probability\_5(x) = freq\_probability\_5(x) / 1484

freq\_probability\_6(x) = freq\_probability\_6(x) / 1484

freq\_probability\_7(x) = freq\_probability\_7(x) / 1484

freq\_probability\_8(x) = freq\_probability\_8(x) / 1484

freq\_probability\_9(x) = freq\_probability\_9(x) / 1484

Next

h\_value\_att(1) = h\_value(freq\_probability\_1)

h\_value\_att(2) = h\_value(freq\_probability\_2)

h\_value\_att(3) = h\_value(freq\_probability\_3)

h\_value\_att(4) = h\_value(freq\_probability\_4)

h\_value\_att(5) = h\_value(freq\_probability\_5)

h\_value\_att(6) = h\_value(freq\_probability\_6)

h\_value\_att(7) = h\_value(freq\_probability\_7)

h\_value\_att(8) = h\_value(freq\_probability\_8)

h\_value\_att(9) = h\_value(freq\_probability\_9)

End Sub

Private Sub EntropyBase\_click()

List1.Clear

List1.AddItem "nothing"

End Sub

Private Sub forward\_click()

List1.Clear

For x = 1 To 9

List1.AddItem h\_value\_att(x)

Next

End Sub

Function h\_value(freq\_prob() As Double) As Double

Dim temp As Double

temp = 0

Dim i As Integer

Dim log\_value As Double

For i = 1 To interval

log\_value = log2(freq\_prob(i))

temp = temp - freq\_prob(i) \* log\_value

Next

h\_value = temp

End Function

Function log2(x As Double) As Double

If (x = 0) Then

log2 = 0

Else

log2 = Log(x) / Log(2)

End If

End Function

Dim select(9) as integer

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| attr1 | Attr2 | Attr3 | Attr4 | Attr5 | Attr6 | Attr7 | Attr8 |

Function goodness(select) As Double

Max=0

For i = 1 to 8

If goodness(select(i))>max

max=goodness(select(i))

select(i)=1

Next

End Function