Midterm

#Yiming Xu

#10438422

#1

Take (0,2), (0,0), (2,2) as examples:

Distance from (0,0) to (0,2), d1 = 16

Distance from (0,2) to (2,2), d2 = 16

Distance from (0,0) to (2,2), d3 = 64

d1+ d2 = 32 < 64 = d3 , so it is not a proper distance function.

#2

1. p(long\_medium) = 0.05
2. p(medium\_pr) = p(s)+p(l) = 0.07 + 0.05 = 0.12
3. p(short\_sleeves) = p(s)+p(m)+p(l) = 0.11+0.27+0.18 = 0.56

p(long\_sleeves) = 1-p(short\_sleeves) = 0.44

1. p(m) = 0.08/(0.04+0.08+0.03) = 0.533
2. p(short) = 0.08/(0.08+0.1) = 0.444

p(long) = 0.1/0.18 = 0.556

#3

a)

1. k=2 and method =“unweighted vote” is used



****

****

**Since the distance of X to 2 and x to 1 are two nearest points, X’s income is either 90k or 100k, 50% each**

2. k=3 and method = distance weighted voted is used

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Id | Age | Asset size |  |  | Income |
| X | 30 | 60 | 0.2 | 0.25 |  |
| 1 | 25 | 50 | 0 | 0 | 100k |
| 2 | 33 | 60 | 0.2 | 0.4 | 90k |
| 3 | 35 | 80 | 0.6 | 0.5 | 150k |

Age mean = (30+25+33+35)/4 =30.75, mins = 25, max = 45, range = 20

(30-25)/20 = 0.25

(25-25)/20 = 0

(33-25)/20=0.4

(35-25)/20 =0.5

Asset mean =( 60+50+60+80)/4 =62.5, mins = 50, max=100, range = 50

(60-50)/50=0.2

(50-50)/50=0

(60-50)/50=0.2

(80-50)/50=0.6



****

****







Since the distance of X to 2 and X to 1 and X to 3 are the three nearest points, X’s income = 90K.

b)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Id | Age | Asset size |  |  | Income |
| X | 30 | 60 | 0.2 | 0.25 |  |
| 1 | 25 | 50 | 0 | 0 | Medium |
| 2 | 33 | 60 | 0.2 | 0.4 | Low |
| 3 | 35 | 80 | 0.6 | 0.5 | High |
| 4 | 30 | 60 | 0.2 | 0.25 | Medium |
| 5 | 30 | 60 | 0.2 | 0.25 | High |
| 6 | 30 | 60 | 0.2 | 0.25 | High |



(30-25)/20=0.25

(25-25)/20=0

(33-25)/20=0.4

(35-25)/20=0.5



(60-50)/50=0.2

(50-50)/50=0

(60-50)/50=0.2

(80-50)/50=0.6

























Since the distance of X to 4 and X to 5 and X to 6 are the three nearest points, X’s income is high.