**Homework 1 Clustering and RegressionSupakone kongprapan 6432172121**

**Instructions**

T1. 0.7

T2. Precision = 0.8

Recall = 0.66

F1 = 0.8

T3. Precision = 0.6

Recall = 0.75

F1 = 0.66

T4. Precision = 0.6

Recall = 0.75

F1 = 0.66

Because we use same data even the population characteristics change

OT1. First, I calculate

And I think term (TN-TP) is only term that can make ACC-F1 negative or positive.

So, Accuracy will be equal only when TP = TN. Greater than F1 only when TN > TP and less than F1 only when TP<TN.

**Hello Clustering**

T5.

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

Round 1 Start

Step 1 Initialization

K = 3

centriod : (3,3) (2,2) (-3,-3)

color : red green blue

Step 2 assign each point

(1,2) -> green

(3,3) -> red

(2,2) -> green

(8,8) -> red

(6,6) -> red

(7,7) -> red

(-3,-3) -> blue

(-2,-4) -> blue

(-7,-7) -> blue

A graph with different colored dots

Description automatically generated

Step 3 Update

new centriod :(6.0,6.0) (1.5,2.0) (-4.0,-4.666666666666667)

Step 4 Repeat

check if centriod change

centriod Change

repeat with new centriod :(6.0,6.0) (1.5,2.0) (-4.0,-4.666666666666667)

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

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

Round 2 Start

Step 1 Initialization

K = 3

centriod : (6.0,6.0) (1.5,2.0) (-4.0,-4.666666666666667)

color : red green blue

Step 2 assign each point

(1,2) -> green

(3,3) -> green

(2,2) -> green

(8,8) -> red

(6,6) -> red

(7,7) -> red

(-3,-3) -> blue

(-2,-4) -> blue

(-7,-7) -> blue

A graph with different colored dots

Description automatically generated

Step 3 Update

new centriod :(7.0,7.0) (2.0,2.3333333333333335) (-4.0,-4.666666666666667)

Step 4 Repeat

check if centriod change

centriod Change

repeat with new centriod :(7.0,7.0) (2.0,2.3333333333333335) (-4.0,-4.666666666666667)

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

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

Round 3 Start

Step 1 Initialization

K = 3

centriod : (7.0,7.0) (2.0,2.3333333333333335) (-4.0,-4.666666666666667)

color : red green blue

Step 2 assign each point

(1,2) -> green

(3,3) -> green

(2,2) -> green

(8,8) -> red

(6,6) -> red

(7,7) -> red

(-3,-3) -> blue

(-2,-4) -> blue

(-7,-7) -> blue

A graph with different colored dots

Description automatically generated

Step 3 Update

new centriod :(7.0,7.0) (2.0,2.3333333333333335) (-4.0,-4.666666666666667)

Step 4 Repeat

check if centriod change

centriod don't change

finish!!

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

T6.

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

Round 1 Start

Step 1 Initialization

K = 3

centriod : (-3,-3) (2,2) (-7,-7)

color : red green blue

Step 2 assign each point

(1,2) -> green

(3,3) -> green

(2,2) -> green

(8,8) -> green

(6,6) -> green

(7,7) -> green

(-3,-3) -> red

(-2,-4) -> red

(-7,-7) -> blue

A graph with different colored dots

Description automatically generated

Step 3 Update

new centriod :(-2.5,-3.5) (4.5,4.666666666666667) (-7.0,-7.0)

Step 4 Repeat

check if centriod change

centriod Change

repeat with new centriod :(-2.5,-3.5) (4.5,4.666666666666667) (-7.0,-7.0)

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

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

Round 2 Start

Step 1 Initialization

K = 3

centriod : (-2.5,-3.5) (4.5,4.666666666666667) (-7.0,-7.0)

color : red green blue

Step 2 assign each point

(1,2) -> green

(3,3) -> green

(2,2) -> green

(8,8) -> green

(6,6) -> green

(7,7) -> green

(-3,-3) -> red

(-2,-4) -> red

(-7,-7) -> blue

A graph with different colored dots

Description automatically generated

Step 3 Update

new centriod :(-2.5,-3.5) (4.5,4.666666666666667) (-7.0,-7.0)

Step 4 Repeat

check if centriod change

centriod don't change

finish!!

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

T7. For me I think startign point from T6 is better because I don’t like that (-7,-7) is cosider the same group as (-3,-3), (-2,-4)

For mesurement I think we can use Eucidien distance from main centriod as a mesurement for “goodness” in T5 it is 11.852337947930973 and in T6 it is 15.003583762738607

I think in K-mean algorithms, It is very important to try different set of centriod point to find “best” group for you

OT2. I think best k for me is 4, solely because in dataset It seem like it has 4 group.

**My heart will go on.**

T8.