KNN IN MAPREDUCE?

Anchalia, P. P., & Roy, K. The k-Nearest Neighbor Algorithm Using MapReduce Paradigm.

Possible implementation of KNN with K=1 in Mapreduce

Computer 1

X	Y	Class
0.5	3	Y
1	2	N
-2	0,5	N

$p = (p_x=1.3, p_y=2.1)$
Which class?

```
Map(k=null, v=(x,y,class)):
  return(k=null, (distance((x,y), (p_x, p_y)), class)
```

Computer 2

X	Y	Class
0.9	1	N
2	2	Y
-2	1	N

map

(, (1.2, Y))
(, (0.32, N))
(, (3.67, N))

map

(, (1.17, Y))
(, (0.71, N))
(, (3.48, N))

Sort and shuffle. Not required

```
Reduce(k=null, v=((d_1,class_1), (d_2, class_2), …):
  compute minimum d_i
  return (di, class_i)
```

reduce

(**N,** 0.32)

Possible implementation of KNN with K=1 in Mapreduce

Computer 1

X	Y	Class
0.5	3	Y
1	2	N
-2	0,5	N

$p = (p_x=1.3, p_y=2.1)$
Which class?

```
Map(k=null, v=(x,y,class)):
  return(k=null, (distance((x,y), (p_x, p_y)), class)
```

Computer 2

X	Y	Class
0.9	1	N
2	2	Y
-2	1	N

map

(, (1.2, Y))
(, (0.32, N))
(, (3.67, N))

map

(, (1.17, Y))
(, (0.71, N))
(, (3.48, N))

Notice that we could use a combiner here

Sort and shuffle. Not required

```
Reduce(k=null, v=((d_1,class_1), (d_2, class_2), …):
  compute minimum d_i
  return (di, class_i)
```

reduce

(**N,** 0.32)

Second possible implementation of KNN with K=1 in Mapreduce

Computer 1

X	Y	Class
0.5	3	Y
1	2	N
-2	0,5	N

$p = (p_x=1.3, p_y=2.1)$
Which class?

```
Map(k=class, v=(x,y)):
  return(k=class, (distance((x,y), (p_x, p_y)))
```

Computer 2

X	Y	Class
0.9	1	N
2	2	Y
-2	1	N