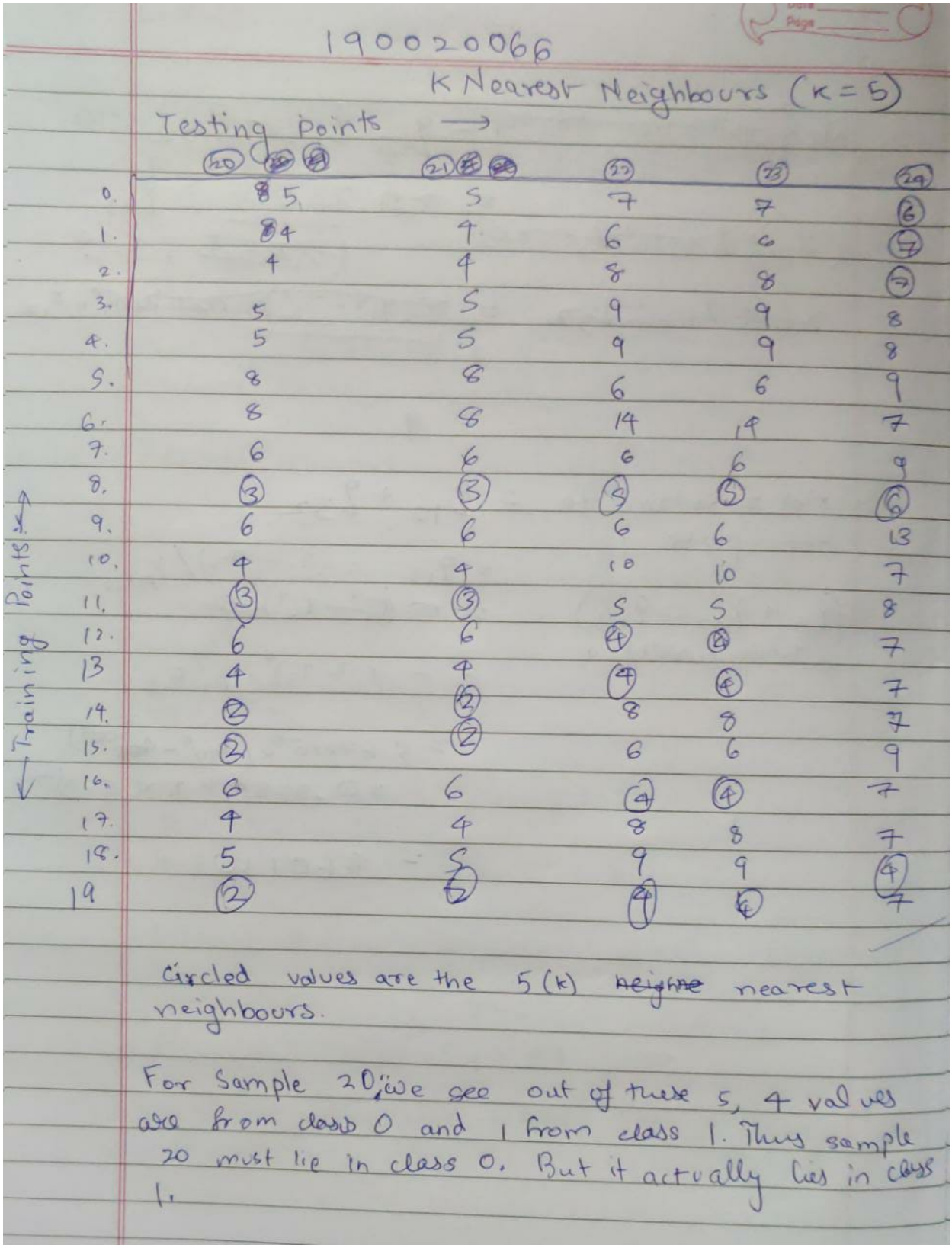
dataset: Somerville Happiness Survey Data Set
model: K-Nearest Neighbours

D = decision attribute (D) with values 0 (unhappy) and 1 (happy)
Attributes X1 to X6 have values 1 to 5.
X1 = the availability of information about the city services
X2 = the cost of housing
X3 = the overall quality of public schools
X4 = your trust in the local police
X5 = the maintenance of streets and sidewalks
X6 = the availability of social community events

Sample Data: Training: 0 to 19, Testing: 20 to 24

	D	X1	X2	X3	X4	X5	X6
0	0	3	3	3	4	2	4
1	0	3	2	3	5	4	3
2	1	5	3	3	3	3	5
3	0	5	4	3	3	3	5
4	0	5	4	3	3	3	5
5	1	5	5	3	5	5	5
6	0	3	1	2	2	1	3
7	1	5	4	4	4	4	5
8	0	4	1	4	4	4	4
9	0	4	4	4	2	5	5
10	0	3	2	3	3	2	3
11	0	4	4	3	4	4	4
12	1	5	2	4	5	5	5
13	0	4	2	4	5	4	3
14	0	4	1	3	3	4	3
15	1	3	2	4	3	4	4
16	0	5	3	4	5	4	5
17	1	5	1	4	3	4	5
18	0	5	1	2	4	4	5
19	0	4	2	4	4	4	4
20	1	4	2	3	3	4	4
21	1	4	2	3	3	4	4
22	0	4	3	5	5	5	4
23	0	4	3	5	5	5	4
24	1	5	1	2	5	2	4



190020066

K Nearest Neighbours (K = 5)

Training Points ↓	Testing Points → (20)	(21)	(22)	(23)	(24)
0.	8 5	5	7	7	(6)
1.	8 4	4	6	6	(7)
2.	4	4	8	8	(7)
3.	5	5	9	9	8
4.	5	5	9	9	8
5.	8	8	6	6	9
6.	8	8	14	14	7
7.	6	6	6	6	9
8.	(3)	(3)	(5)	(5)	(6)
9.	6	6	6	6	13
10.	4	4	10	10	7
11.	(3)	(3)	5	5	8
12.	6	6	(4)	(4)	7
13	4	4	(4)	(4)	7
14.	(2)	(2)	8	8	7
15.	(2)	(2)	6	6	9
16.	6	6	(4)	(4)	7
17.	4	4	8	8	7
18.	5	(5)	9	9	(4)
19	(2)	(2)	(4)	(4)	7

Circled values are the 5 (k) ~~neighbe~~ nearest neighbours.

For sample 20, we see out of these 5, 4 values are from class 0 and 1 from class 1. Thus sample 20 must lie in class 0. But it actually lies in class 1.

190020066

similarly sample 21 must lie in class 0, sample 22 in class 0, sample

Similarly:

Sample	Our output	Actual output class (1)
21	0	1
22	0	0
23	0	0
24	0	1

Thus we obtain an accuracy of 40%.

~~Upon comparing to the Github codes, we~~ The github (SK learn) classifier also gives us an accuracy of 40% for k=5.

For distance, I have used

$$|X| = \sum_{i=1}^{n} |x_i|$$ (n = no of features)

Train dataset : 0 to 19 (20 total)
Test dataset : 20 to 24. (5 total)

I haven't scaled the values because all features lie in 0 to 5.