1) a) purity

Good\_docs(w1)= max(4,1,1)=4

Good\_docs(w2)= max(2,4,0)=4

Good\_docs(w3)= max(1,0,4)=4

Purity= 1/17 \*(4+4+4)=12/17=0.71

b)

contingency table

|  |  |  |
| --- | --- | --- |
|  | Same cluster | Different cluster |
| Same class | Tp 19 | Fn 22 |
| Different class | Fp 21 | tn 74 |

TP+FP=

TP=

FP=21

Same class= TP+FN=41

FN=22

TP + FP + TN + FN =

TN=74

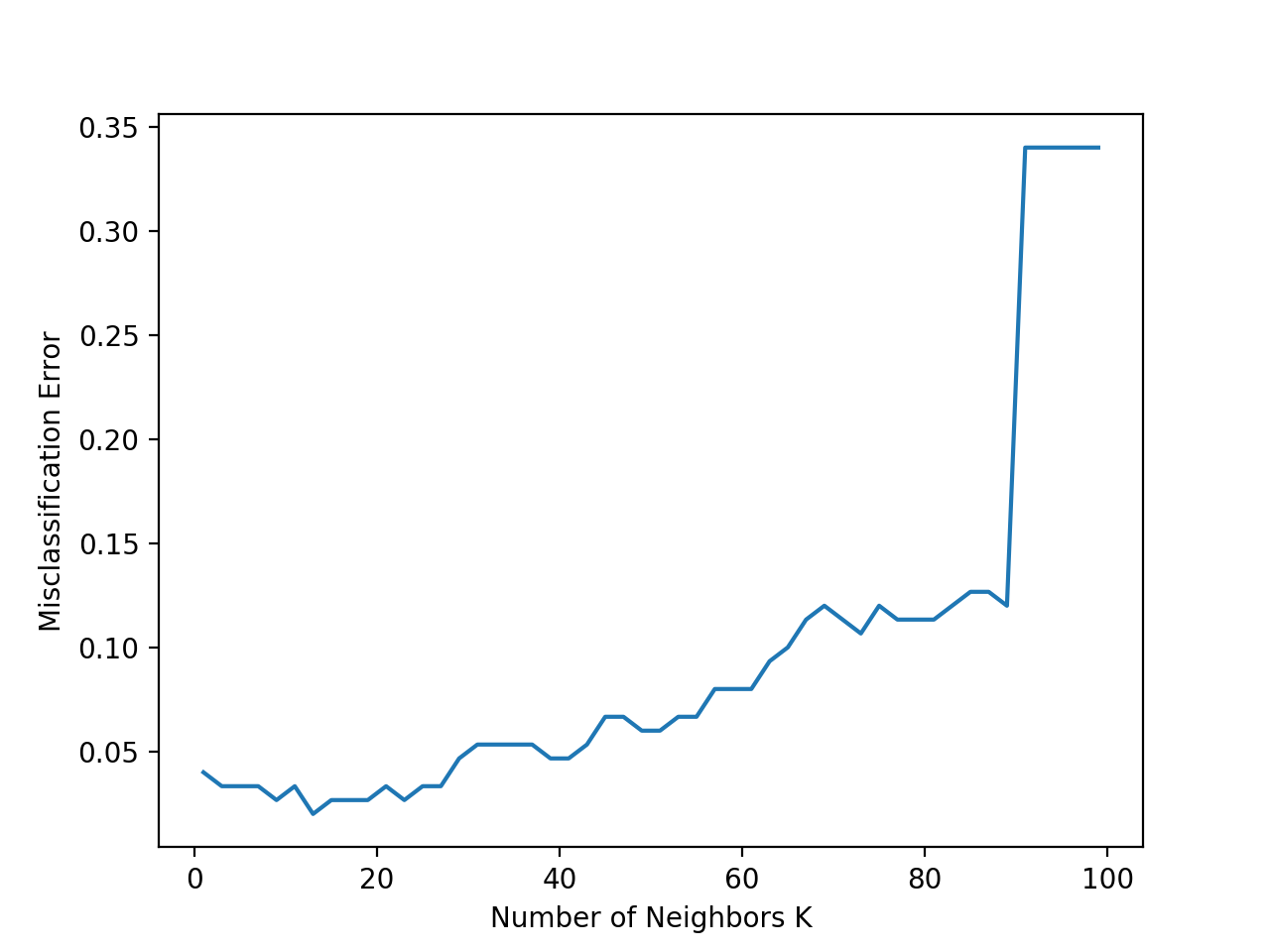
Rand Index=(19+74)/136=0.68

P = tp/(tp + fp) = 19/40=0.475

R = tp/(tp + fn) = 19/41=0.463

Fβ=1 = 2\*P\*R/(P+R) = 0.44/0.938=0.469

2)



3)

I will graduate in next Dec. I want to work full time from next May. I want to get a full-time job as data scientist or data analyst. I am trying so hard to study python, R and other language. I also need to study big data like spark or apache software. Find a job in U.S. as international student is hard. Luckily, I start to make connection early. Now I do not have much thought about my future, I need to find a job asap and find a way to provide my family.