Chapter 4
Nearest Neighbors for Classifying New Medicines (2 New and 25 Old Opioids)

General Purpose

Nearest neighbor methodology has a long history, and has, initially, been used for data imputation in demographic data files. This chapter is to assess whether it can also been used for classifying new medicines.

Specific Scientific Question

For most diseases a whole class of drugs rather than a single compound is available. Nearest neighbor methods can be used for identifying the place of a new drug within its class.

Example

Two newly developed opioid compounds are assessed for their similarities with the standard opioids in order to determine their potential places in therapeutic regimens. Underneath are the characteristics of 25 standard opioids and two newly developed opioid compounds.

This chapter was previously published in "Machine learning in medicine-cookbook 2" as Chap. 1, 2014.

© Springer International Publishing Switzerland 2015

T.J. Cleophas, A.H. Zwinderman, *Machine Learning in Medicine - a Complete Overview*, DOI 10.1007/978-3-319-15195-3_4

Drugname	analgesia score	antitussive score	constipation score	respiratory score	abuse score	eliminate time	duration time
buprenorphine	7,00	4,00	5,00	7,00	4,00	5,00	9,00
butorphanol	7,00	3,00	4,00	7,00	4,00	2,70	4,00
codeine	5,00	6,00	6,00	5,00	4,00	2,90	7,00
heroine	8,00	6,00	8,00	8,00	10,00	9,00	15,00
hydromorphone	8,00	6,00	6,00	8,00	8,00	2,60	5,00
levorphanol	8,00	6,00	6,00	8,00	8,00	11,00	20,00
mepriridine	7,00	2,00	4,00	8,00	6,00	3,20	14,00
methadone	9,00	6,00	6,00	8,00	6,00	25,00	5,00
morphine	8,00	6,00	8,00	8,00	8,00	3,10	5,00
nalbuphine	7,00	2,00	4,00	7,00	4,00	5,10	4,50
oxycodone	6,00	6,00	6,00	6,00	8,00	5,00	4,00
oxymorphine	8,00	5,00	6,00	8,00	8,00	5,20	3,50
pentazocine	7,00	2,00	4,00	7,00	5,00	2,90	3,00
propoxyphene	5,00	2,00	4,00	5,00	5,00	3,30	2,00
nalorphine	2,00	3,00	6,00	8,00	1,00	1,40	3,20
levallorphan	3,00	2,00	5,00	4,00	1,00	11,00	5,00
cyclazocine	2,00	3,00	6,00	3,00	2,00	1,60	2,80
naloxone	1,00	2,00	5,00	8,00	1,00	1,20	3,00
naltrexon	1,00	3,00	5,00	8,00	,00	9,70	14,00
alfentanil	7,00	6,00	7,00	4,00	6,00	1,60	,50
alphaprodine	6,00	5,00	6,00	3,00	5,00	2,20	2,00
fentanyl	6,00	5,00	7,00	5,00	4,00	3,70	,50
meptazinol	4,00	3,00	5,00	5,00	3,00	1,60	2,00
norpropoxyphene	8,00	6,00	8,00	5,00	7,00	6,00	4,00
sufentanil	7,00	6,00	8,00	6,00	8,00	2,60	5,00
newdrug1	5,00	5,00	4,00	3,00	6,00	5,00	12,00
newdrug2	8,00	6,00	3,00	4,00	5,00	7,00	16,00

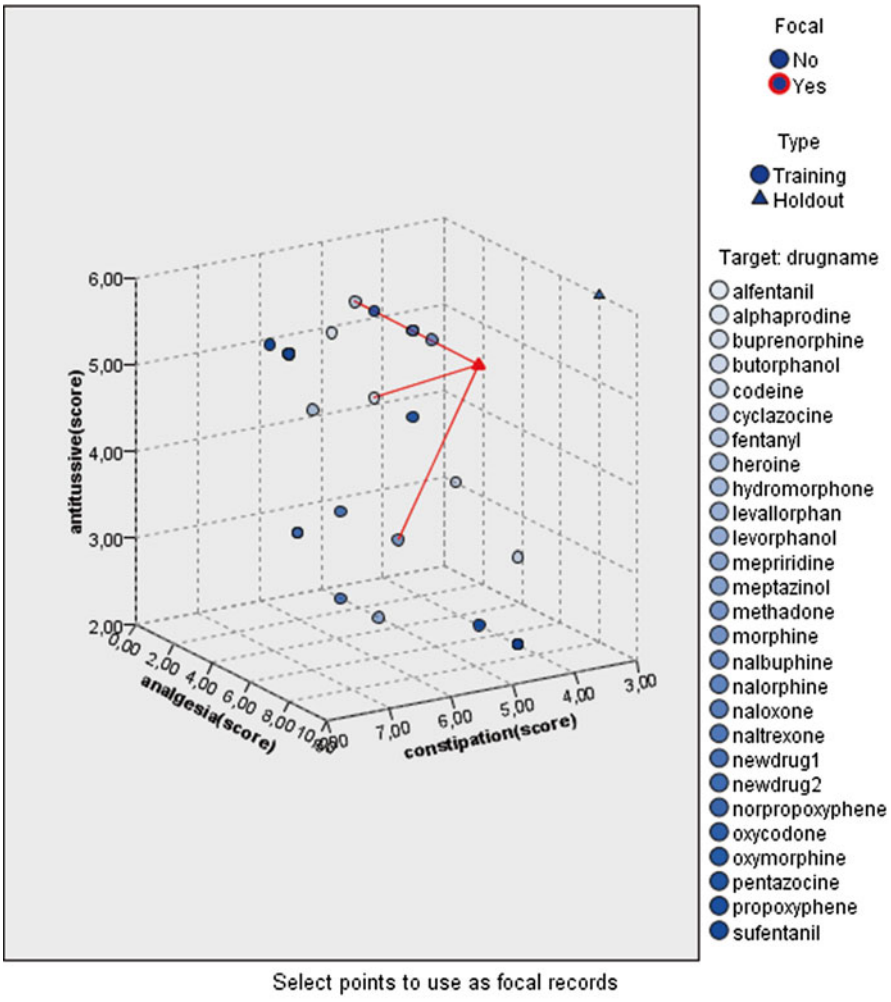
Var = variable
Var 1 analgesia score (0–10)
Var 2 antitussive score (0–10)
Var 3 constipation score (0–10)
Var 4 respiratory depression score (1–10)
Var 5 abuse liability score (1–10)
Var 6 elimination time ($t_{1/2}$ in hours)
Var 7 duration time analgesia (hours)

The data file is entitled “nearestneighbor” and is in extras.springer.com.

SPSS statistical software is used for data analysis. Start by opening the data file. The drug names included, eight variables are in the file. A ninth variable entitled “partition” must be added with the value 1 for the opioids 1–25 and 0 for the two new compounds (cases 26 and 27).

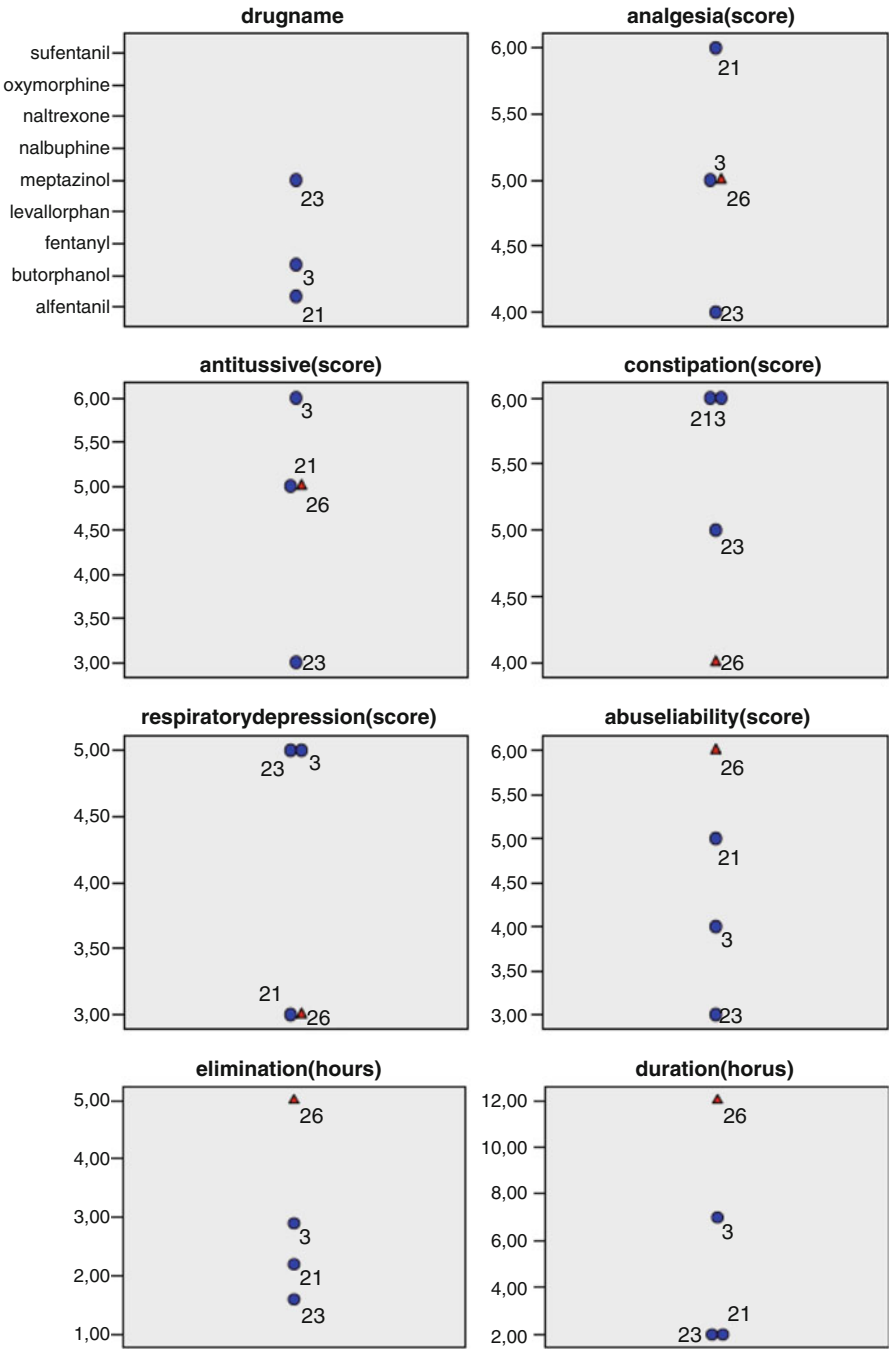
Then command:

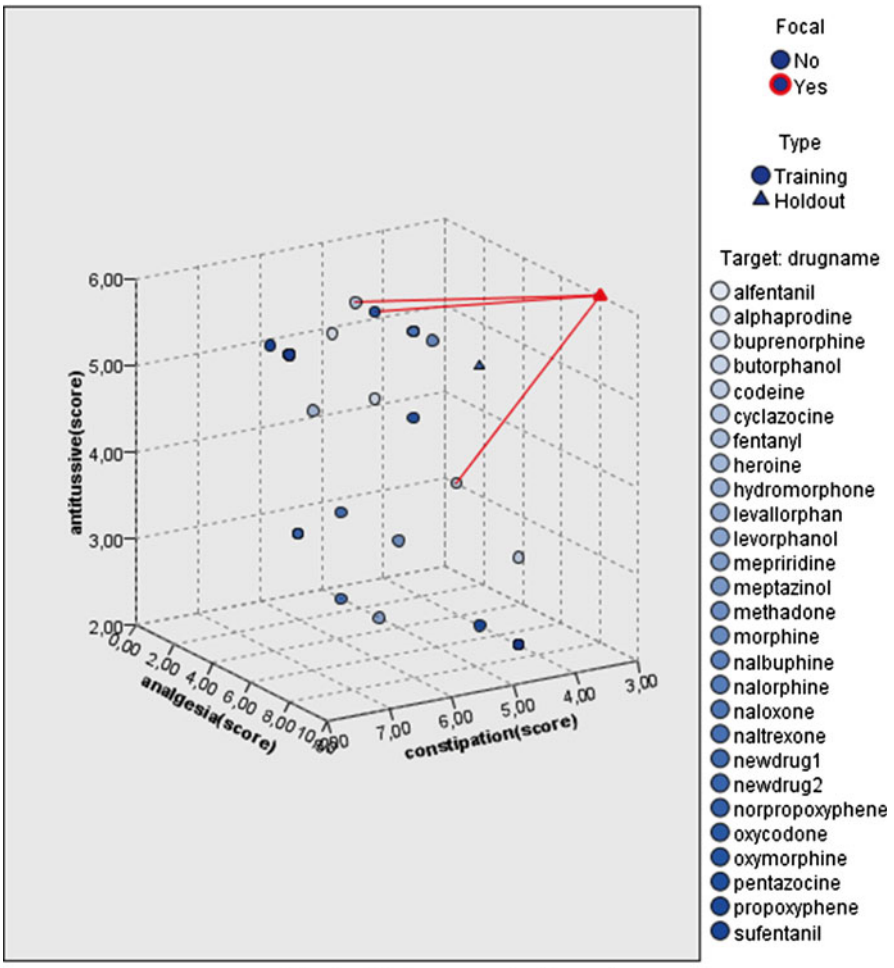
Analyze….Classify….Nearest Neighbor Analysis….enter the variable "drugsname" in Target….enter the variables "analgesia" to "duration of analgesia" in Features…. click Partitions….click Use variable to assign cases….enter the variable "Partition"….click OK.



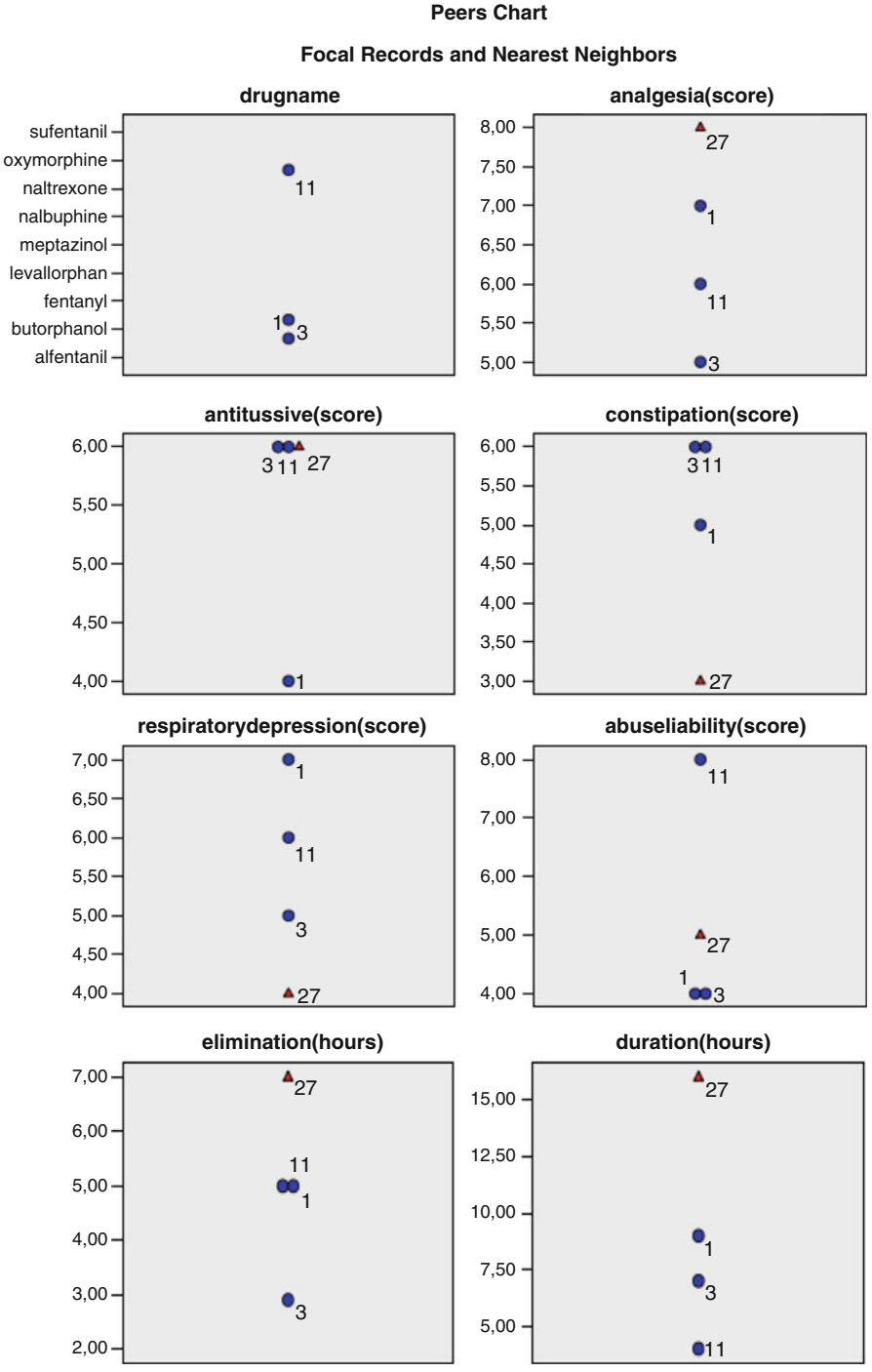
The above figure shows as an example the place of the two new compounds (the small triangles) as compared with those of the standard opioids. Lines connect them to their 3 nearest neighbors. In SPSS' original output sheets the graph can by double-clicking be placed in the "model viewer", and, then, (after again clicking on it) be

interactively rotated in order to improve the view of the distances. SPSS uses 3 nearest neighbors by default, but you can change this number if you like. The names of the compounds are given in alphabetical order. Only three of seven variables are given in the initial figure, but if you click on one of the small triangles in this figure, an auxiliary view comes up right from the main view. Here are all the details of the analysis. The upper left graph of it shows that the opioids 21, 3, and 23 have the best average nearest neighbor records for case 26 (new drug 1). The seven figures alongside and underneath this figure give the distances between these three and case 26 for each of the seven features (otherwise called predictor variables).


drugname
sufentanil
oxymorphine
naltrexone
nalbuphine
meptazinol
levallorphan
fentanyl
butorphanol
alfentanil
23
3
21
analgesia(score)
6,00
5,50
5,00
4,50
4,00
21
3
26
23
antitussive(score)
6,00
5,50
5,00
4,50
4,00
3,50
3,00
3
21
26
23
constipation(score)
6,00
5,50
5,00
4,50
4,00
213
23
26
respiratorydepression(score)
5,00
4,50
4,00
3,50
3,00
23
3
21
26
abuseliability(score)
6,00
5,50
5,00
4,50
4,00
3,50
3,00
26
21
3
23
elimination(hours)
5,00
4,00
3,00
2,00
1,00
26
3
21
23
duration(horus)
12,00
10,00
8,00
6,00
4,00
2,00
26
3
21
23



If you click on the other triangle (representing case 27 (newdrug 2) in the initial figure), the connecting lines with the nearest neighbors of this drug comes up. This is shown in the above figure, which is the main view for drug 2. Using the same manoeuvre as above produces again the auxiliary view showing that the opioids 3, 1, and 11 have the best average nearest neighbor records for case 27 (new drug 2). The seven figures alongside and underneath this figure give again the distances between these three and case 27 for each of the seven features (otherwise called predictor variables). The auxiliary view is shown underneath.


Peers Chart
Focal Records and Nearest Neighbors
drugname
sufentanil
oxymorphine
naltrexone
nalbuphine
meptazinol
levallorphan
fentanyl
butorphanol
alfentanil
11
1
3
analgesia(score)
8,00
7,50
7,00
6,50
6,00
5,50
5,00
27
1
11
3
antitussive(score)
6,00
5,50
5,00
4,50
4,00
3 11 27
1
constipation(score)
6,00
5,50
5,00
4,50
4,00
3,50
3,00
3 11
1
27
respiratorydepression(score)
7,00
6,50
6,00
5,50
5,00
4,50
4,00
1
11
3
27
abuseliability(score)
8,00
7,00
6,00
5,00
4,00
11
27
1
3
elimination(hours)
7,00
6,00
5,00
4,00
3,00
2,00
27
11
1
3
duration(hours)
15,00
12,50
10,00
7,50
5,00
27
1
3
11

Conclusion

Nearest neighbor methodology enables to readily identify the places of new drugs within their classes of drugs. For example, newly developed opioid compounds can be compared with standard opioids in order to determine their potential places in therapeutic regimens.

Note

Nearest neighbor cluster methodology has a long history and has initially been used for missing data imputation in demographic data files (see Statistics applied to clinical studies 5th edition, 2012, Chap. 22, Missing data, pp 253–266, Springer Heidelberg Germany, from the same authors).