

K -Nearest Neighbour Algorithm

Dr. Lasanthi Watagoda

2/12/2018

Know your data

```
library(class)

data(iris)
str(iris)

## 'data.frame': 150 obs. of 5 variables:
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...

table(iris$Species)

##
## setosa versicolor virginica
## 50 50 50

head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1 5.1 3.5 1.4 0.2 setosa
## 2 4.9 3.0 1.4 0.2 setosa
## 3 4.7 3.2 1.3 0.2 setosa
## 4 4.6 3.1 1.5 0.2 setosa
## 5 5.0 3.6 1.4 0.2 setosa
## 6 5.4 3.9 1.7 0.4 setosa

iris

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1 5.1 3.5 1.4 0.2 setosa
## 2 4.9 3.0 1.4 0.2 setosa
## 3 4.7 3.2 1.3 0.2 setosa
## 4 4.6 3.1 1.5 0.2 setosa
## 5 5.0 3.6 1.4 0.2 setosa
## 6 5.4 3.9 1.7 0.4 setosa
## 7 4.6 3.4 1.4 0.3 setosa
## 8 5.0 3.4 1.5 0.2 setosa
## 9 4.4 2.9 1.4 0.2 setosa
## 10 4.9 3.1 1.5 0.1 setosa
## 11 5.4 3.7 1.5 0.2 setosa
## 12 4.8 3.4 1.6 0.2 setosa
## 13 4.8 3.0 1.4 0.1 setosa
## 14 4.3 3.0 1.1 0.1 setosa
## 15 5.8 4.0 1.2 0.2 setosa
## 16 5.7 4.4 1.5 0.4 setosa
```

## 17	5.4	3.9	1.3	0.4	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 19	5.7	3.8	1.7	0.3	setosa
## 20	5.1	3.8	1.5	0.3	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 23	4.6	3.6	1.0	0.2	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 39	4.4	3.0	1.3	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 41	5.0	3.5	1.3	0.3	setosa
## 42	4.5	2.3	1.3	0.3	setosa
## 43	4.4	3.2	1.3	0.2	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 45	5.1	3.8	1.9	0.4	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 49	5.3	3.7	1.5	0.2	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 51	7.0	3.2	4.7	1.4	versicolor
## 52	6.4	3.2	4.5	1.5	versicolor
## 53	6.9	3.1	4.9	1.5	versicolor
## 54	5.5	2.3	4.0	1.3	versicolor
## 55	6.5	2.8	4.6	1.5	versicolor
## 56	5.7	2.8	4.5	1.3	versicolor
## 57	6.3	3.3	4.7	1.6	versicolor
## 58	4.9	2.4	3.3	1.0	versicolor
## 59	6.6	2.9	4.6	1.3	versicolor
## 60	5.2	2.7	3.9	1.4	versicolor
## 61	5.0	2.0	3.5	1.0	versicolor
## 62	5.9	3.0	4.2	1.5	versicolor
## 63	6.0	2.2	4.0	1.0	versicolor
## 64	6.1	2.9	4.7	1.4	versicolor
## 65	5.6	2.9	3.6	1.3	versicolor
## 66	6.7	3.1	4.4	1.4	versicolor
## 67	5.6	3.0	4.5	1.5	versicolor
## 68	5.8	2.7	4.1	1.0	versicolor
## 69	6.2	2.2	4.5	1.5	versicolor
## 70	5.6	2.5	3.9	1.1	versicolor

## 71	5.9	3.2	4.8	1.8 versicolor
## 72	6.1	2.8	4.0	1.3 versicolor
## 73	6.3	2.5	4.9	1.5 versicolor
## 74	6.1	2.8	4.7	1.2 versicolor
## 75	6.4	2.9	4.3	1.3 versicolor
## 76	6.6	3.0	4.4	1.4 versicolor
## 77	6.8	2.8	4.8	1.4 versicolor
## 78	6.7	3.0	5.0	1.7 versicolor
## 79	6.0	2.9	4.5	1.5 versicolor
## 80	5.7	2.6	3.5	1.0 versicolor
## 81	5.5	2.4	3.8	1.1 versicolor
## 82	5.5	2.4	3.7	1.0 versicolor
## 83	5.8	2.7	3.9	1.2 versicolor
## 84	6.0	2.7	5.1	1.6 versicolor
## 85	5.4	3.0	4.5	1.5 versicolor
## 86	6.0	3.4	4.5	1.6 versicolor
## 87	6.7	3.1	4.7	1.5 versicolor
## 88	6.3	2.3	4.4	1.3 versicolor
## 89	5.6	3.0	4.1	1.3 versicolor
## 90	5.5	2.5	4.0	1.3 versicolor
## 91	5.5	2.6	4.4	1.2 versicolor
## 92	6.1	3.0	4.6	1.4 versicolor
## 93	5.8	2.6	4.0	1.2 versicolor
## 94	5.0	2.3	3.3	1.0 versicolor
## 95	5.6	2.7	4.2	1.3 versicolor
## 96	5.7	3.0	4.2	1.2 versicolor
## 97	5.7	2.9	4.2	1.3 versicolor
## 98	6.2	2.9	4.3	1.3 versicolor
## 99	5.1	2.5	3.0	1.1 versicolor
## 100	5.7	2.8	4.1	1.3 versicolor
## 101	6.3	3.3	6.0	2.5 virginica
## 102	5.8	2.7	5.1	1.9 virginica
## 103	7.1	3.0	5.9	2.1 virginica
## 104	6.3	2.9	5.6	1.8 virginica
## 105	6.5	3.0	5.8	2.2 virginica
## 106	7.6	3.0	6.6	2.1 virginica
## 107	4.9	2.5	4.5	1.7 virginica
## 108	7.3	2.9	6.3	1.8 virginica
## 109	6.7	2.5	5.8	1.8 virginica
## 110	7.2	3.6	6.1	2.5 virginica
## 111	6.5	3.2	5.1	2.0 virginica
## 112	6.4	2.7	5.3	1.9 virginica
## 113	6.8	3.0	5.5	2.1 virginica
## 114	5.7	2.5	5.0	2.0 virginica
## 115	5.8	2.8	5.1	2.4 virginica
## 116	6.4	3.2	5.3	2.3 virginica
## 117	6.5	3.0	5.5	1.8 virginica
## 118	7.7	3.8	6.7	2.2 virginica
## 119	7.7	2.6	6.9	2.3 virginica
## 120	6.0	2.2	5.0	1.5 virginica
## 121	6.9	3.2	5.7	2.3 virginica
## 122	5.6	2.8	4.9	2.0 virginica
## 123	7.7	2.8	6.7	2.0 virginica
## 124	6.3	2.7	4.9	1.8 virginica

```
## 125      6.7      3.3      5.7      2.1 virginica
## 126      7.2      3.2      6.0      1.8 virginica
## 127      6.2      2.8      4.8      1.8 virginica
## 128      6.1      3.0      4.9      1.8 virginica
## 129      6.4      2.8      5.6      2.1 virginica
## 130      7.2      3.0      5.8      1.6 virginica
## 131      7.4      2.8      6.1      1.9 virginica
## 132      7.9      3.8      6.4      2.0 virginica
## 133      6.4      2.8      5.6      2.2 virginica
## 134      6.3      2.8      5.1      1.5 virginica
## 135      6.1      2.6      5.6      1.4 virginica
## 136      7.7      3.0      6.1      2.3 virginica
## 137      6.3      3.4      5.6      2.4 virginica
## 138      6.4      3.1      5.5      1.8 virginica
## 139      6.0      3.0      4.8      1.8 virginica
## 140      6.9      3.1      5.4      2.1 virginica
## 141      6.7      3.1      5.6      2.4 virginica
## 142      6.9      3.1      5.1      2.3 virginica
## 143      5.8      2.7      5.1      1.9 virginica
## 144      6.8      3.2      5.9      2.3 virginica
## 145      6.7      3.3      5.7      2.5 virginica
## 146      6.7      3.0      5.2      2.3 virginica
## 147      6.3      2.5      5.0      1.9 virginica
## 148      6.5      3.0      5.2      2.0 virginica
## 149      6.2      3.4      5.4      2.3 virginica
## 150      5.9      3.0      5.1      1.8 virginica
```

```
dim(iris)
```

```
## [1] 150    5
```

Mixup to remove the order.

```
gp <- runif(nrow(iris)) # mix up the data set using this order
gp
```

```
## [1] 0.3621005579 0.9127957784 0.2899519110 0.9765533579 0.6417517532
## [6] 0.9212999842 0.2549693468 0.0267289870 0.4446314252 0.3426514913
## [11] 0.7279655009 0.6650242966 0.6862884082 0.6640912110 0.5783065960
## [16] 0.9775648266 0.9759020756 0.2117969026 0.3871603620 0.0988107231
## [21] 0.6061542630 0.8766701899 0.7969567254 0.1342353954 0.3628753391
## [26] 0.8590396561 0.9393610077 0.3581233732 0.7466434201 0.6935195660
## [31] 0.2669847845 0.7437292670 0.7199546404 0.0724671097 0.6984240951
## [36] 0.3546229936 0.0899193753 0.6057727204 0.9327982236 0.4787994148
## [41] 0.3022015407 0.1121878501 0.3426661454 0.5806459698 0.4761845781
## [46] 0.0657336209 0.1040223623 0.9963526239 0.0077424524 0.6698556775
## [51] 0.9315738820 0.3465957765 0.9872793984 0.4515121835 0.1795202831
## [56] 0.5494884737 0.2848433740 0.8348733354 0.7386295302 0.1924885123
## [61] 0.2670651977 0.1045327610 0.5492642124 0.7335859525 0.0963471327
## [66] 0.9986777715 0.7858086580 0.1446443875 0.7937206733 0.9500710673
## [71] 0.3962900061 0.0368400998 0.3405494350 0.4488672877 0.4707916095
## [76] 0.0041355030 0.8114406976 0.4710592509 0.5500778246 0.0440397165
## [81] 0.8746685707 0.5696219641 0.4581797277 0.0369662393 0.9334289334
## [86] 0.4885814737 0.6967754578 0.6987855816 0.1113616906 0.4292661520
```

```
## [91] 0.8376392527 0.7643393860 0.7728236164 0.5529018529 0.7163106659
## [96] 0.1000489180 0.6900929490 0.1996207007 0.9454076185 0.5289808712
## [101] 0.7646626388 0.1129235497 0.8563886099 0.4629099539 0.1933420654
## [106] 0.7148069842 0.6110793436 0.5554480294 0.6554778717 0.6108181491
## [111] 0.4488256823 0.9926903483 0.5490696935 0.8761014559 0.9202197159
## [116] 0.0925799049 0.7246764861 0.0602161386 0.5342295796 0.2527704656
## [121] 0.4297461566 0.6041131429 0.6509262621 0.2257374299 0.0008185299
## [126] 0.6110424770 0.8188287204 0.2114439476 0.1498408194 0.5196917937
## [131] 0.4516063768 0.0160616294 0.8366670858 0.2306084416 0.9959781975
## [136] 0.8429338289 0.3348966404 0.0665075311 0.6552303587 0.8581187704
## [141] 0.7202524268 0.3582940579 0.6426849570 0.8614209050 0.9719420087
## [146] 0.7751728841 0.9643785697 0.5740743191 0.1162542414 0.8876883790
```

```
iris <- iris[order(gp),]
#iris
head(iris)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 125           6.7           3.3           5.7           2.1 virginica
##  76           6.6           3.0           4.4           1.4 versicolor
##  49           5.3           3.7           1.5           0.2      setosa
## 132           7.9           3.8           6.4           2.0 virginica
##   8           5.0           3.4           1.5           0.2      setosa
##  72           6.1           2.8           4.0           1.3 versicolor
```

Normalize your data

```
summary(iris[,c(1,2,3,4)]) # should we Normalize?
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
## Min.      :4.300   Min.      :2.000   Min.      :1.000   Min.      :0.100
## 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
## Median :5.800   Median :3.000   Median :4.350   Median :1.300
## Mean    :5.843   Mean    :3.057   Mean    :3.758   Mean    :1.199
## 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
## Max.    :7.900   Max.    :4.400   Max.    :6.900   Max.    :2.500
```

```
#Define your own Normalize function
```

```
normalize <- function(x){
  return((x - min(x))/(max(x)-min(x)))
}
```

```
#Use lapply to normalize multiple columns at once.
```

```
iris_n <- as.data.frame(lapply(iris[,c(1,2,3,4)],normalize))
str(iris_n)
```

```
## 'data.frame':   150 obs. of  4 variables:
## $ Sepal.Length: num  0.667 0.639 0.278 1 0.194 ...
## $ Sepal.Width : num  0.542 0.417 0.708 0.75 0.583 ...
## $ Petal.Length: num  0.7966 0.5763 0.0847 0.9153 0.0847 ...
## $ Petal.Width : num  0.8333 0.5417 0.0417 0.7917 0.0417 ...
```

```
summary(iris_n)
```

```
##   Sepal.Length   Sepal.Width   Petal.Length   Petal.Width
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##   1st Qu.:0.2222   1st Qu.:0.3333   1st Qu.:0.1017   1st Qu.:0.08333
##   Median :0.4167   Median :0.4167   Median :0.5678   Median :0.50000
##   Mean    :0.4287   Mean    :0.4406   Mean    :0.4675   Mean    :0.45806
##   3rd Qu.:0.5833   3rd Qu.:0.5417   3rd Qu.:0.6949   3rd Qu.:0.70833
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.00000
```

Tranning X, Y and Test X, Y

```
iris_train <- iris_n[1:129,] # Training X
iris_test  <- iris_n[130:150,] # Test X

iris_train_target <- iris[1:129,5] # Training Y
iris_test_target  <- iris[130:150,5] # Test Y
```

```
#require(class)
```

```
k <- 13 # sqrt(n) make it odd
```

```
m1 <- knn(train = iris_train, test = iris_test, cl = iris_train_target, k = k)
```

Confusion Metrix

```
CM <- table(iris_test_target, m1) # x and y
CM
```

```
##           m1
## iris_test_target setosa versicolor virginica
##      setosa      9          0          0
##      versicolor  0          6          0
##      virginica   0          1          5
```