

Lab 7

[Vaishak Balachandra]

Q.N. 1) Suppose we have height, weight and T-shirt size of some players and we need to predict the T-shirt size of a new players given only height and weight information we have. Data including height, weight and T-shirt size information is shown below (Also provided in the Brightspace page)

Height (in cms)	Weight (in kgs)	T Shirt Size
158	58	M
158	59	M
158	63	M
160	59	M
160	60	M
163	60	M
163	61	M
160	64	L
163	64	L
165	61	L
165	62	L
165	65	L
168	62	L
168	63	L
168	66	L
170	63	L
170	64	L
170	68	L

Perform the KNN to classify the players in order to determine their T-shirt size. Use K=5

```
> # Q1
> Q1 <- read.csv("Q1.csv")
> head(Q1)
  Height..in.cms. Weight..in.kgs. T.Shirt.Size
1           158           58             M
2           158           59             M
3           158           63             M
4           160           59             M
5           160           60             M
6           163           60             M
> names(Q1)
[1] "Height..in.cms." "Weight..in.kgs." "T.Shirt.Size"
> names(Q1) = c("Height", "Weight", "Size")
> dim(Q1)
[1] 18  3
> attach(Q1)
>
>
> set.seed(111)
> # install.packages("e1071")
> library(e1071)
> index = createDataPartition(Q1$Size, p = 0.7, list = F)
```

```

>
> train = Q1[index, -3] # except the label, choosing the rest rows in index
> test = Q1[-index,-3]
>
> train_label = Q1[index,3]
> test_label = Q1[-index, 3]
>
> # install.packages("class")
> library(class)
> pred = knn(train = train, test = test, cl = train_label, k = 5)
> pred
[1] M M L L L
Levels: L M
>
> table(test_label, pred)
      pred
test_label L M
      L 3 0
      M 0 2
>
> # install.packages("gmodels")
> library(gmodels)
> CrossTable(x = test_label, y = pred, prop.chisq = F)

```

Cell Contents

	N
N / Row Total	
N / Col Total	
N / Table Total	

Total Observations in Table: 5

test_label	pred		Row Total
	L	M	
L	3	0	3
	1.000	0.000	0.600
	1.000	0.000	
	0.600	0.000	
M	0	2	2
	0.000	1.000	0.400
	0.000	1.000	
	0.000	0.400	
Column Total	3	2	5
	0.600	0.400	

```

> confusionMatrix(reference=factor(test_label), data = pred)
Confusion Matrix and Statistics

```

```

      Reference
Prediction L M
      L 3 0
      M 0 2

```

```

      Accuracy : 1
      95% CI : (0.4782, 1)
No Information Rate : 0.6
P-Value [Acc > NIR] : 0.07776

```

Kappa : 1

McNemar's Test P-Value : NA

Sensitivity : 1.0

Specificity : 1.0

Pos Pred Value : 1.0

Neg Pred Value : 1.0

Prevalence : 0.6

Detection Rate : 0.6

Detection Prevalence : 0.6

Balanced Accuracy : 1.0

'Positive' Class : L

> # Accuracy: 100%

> # Sensitivity = 100%

> # Specificity = 100%