KNN Case Study

Abhirup Sen

03/06/2021

```
library(ISLR)
dim(Caravan)
```

```
## [1] 5822   86
```

```
attach(Caravan)
summary(Purchase)
```

```
##   No  Yes
## 5474  348
```

#standerdize and scale

```
standerdized.X = scale(Caravan[,-86])
var(Caravan[,1])
```

```
## [1] 165.0378
```

```
var(Caravan[,2])
```

```
## [1] 0.1647078
```

```
var(standerdized.X[,1])
```

```
## [1] 1
```

```
var(standerdized.X[,2])
```

```
## [1] 1
```

```
library(class)
test = 1:1000
train.X = standerdized.X[-test,]
test.X= standerdized.X[test,]
train.Y = Purchase[-test]
test.Y = Purchase[test]
set.seed(1)

knn.pred = knn(train.X,test.X, train.Y, k = 1)
mean(test.Y != knn.pred)
```

```
## [1] 0.118
```

```
mean(test.Y!="No")
```

```
## [1] 0.059
```

```
table(knn.pred,test.Y)
```

```
##         test.Y
## knn.pred  No Yes
##      No  873  50
##      Yes  68   9
```

```
knn.pred = knn(train.X,test.X,train.Y,k=3)
table(knn.pred,test.Y)
```

```
##         test.Y
## knn.pred  No Yes
##      No  920  54
##      Yes  21   5
```

```
knn.pred = knn(train.X,test.X,train.Y,k=5)
table(knn.pred,test.Y)
```

```
##         test.Y
## knn.pred  No Yes
##      No  930  55
##      Yes  11   4
```

#comparing with Logistic

```
glm.fits = glm(Purchase~., data = Caravan, family = binomial, subset = -test)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
glm.probs = predict(glm.fits, Caravan[test,], type = "response")
glm.pred = rep("No",1000)
glm.pred[glm.probs>0.5]="Yes"
table(glm.pred,test.Y)
```

```
##         test.Y
## glm.pred  No Yes
##      No  934  59
##      Yes   7   0
```

```
glm.pred=rep("No",1000)
glm.pred[glm.probs>0.25]="Yes"
table(glm.pred,test.Y)
```

```
##         test.Y
## glm.pred  No Yes
##      No  919  48
##      Yes  22  11
```