Project 1

Evgeni Vershinin, Kristina Ødegård

2024-02-08

**Problem 1
a)
Quantative: Time, income earned, horsepower of a car. Qvalitative: Marital status, origin, Gender.
b) KNN, LDA, QDA can be used for multi-class classifications.
c)
d)The nereast neighbour for k=1 is a blue dot, so our classification is blue. For K=3, two of the nearest neighbors are red and 1 blue. This gives 2/3 red, so it is red. For K=5 we have 3/5 red, so it is red.

```
library(MASS)
data(Boston)
data = Boston

model = lm(medv ~ rm + age,data=data)
summary(model)
```

```
##
## Call:
## lm(formula = medv ~ rm + age, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.555  -2.882  -0.274   2.293  40.799
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.27740    2.85676  -8.848  < 2e-16 ***
## rm            8.40158    0.41208  20.388  < 2e-16 ***
## age          -0.07278    0.01029  -7.075 5.02e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.316 on 503 degrees of freedom
## Multiple R-squared:  0.5303, Adjusted R-squared:  0.5284
## F-statistic: 283.9 on 2 and 503 DF,  p-value: < 2.2e-16
```

```
cor_matrix = cor(data.frame(data$medv, data$rm, data$age))
print(cor_matrix)
```

```
##            data.medv    data.rm   data.age
## data.medv  1.0000000  0.6953599 -0.3769546
## data.rm    0.6953599  1.0000000 -0.2402649
## data.age  -0.3769546 -0.2402649  1.0000000
```

```r
model2 = lm(medv ~ rm + age + nox, data=data)
summary(model2)
```

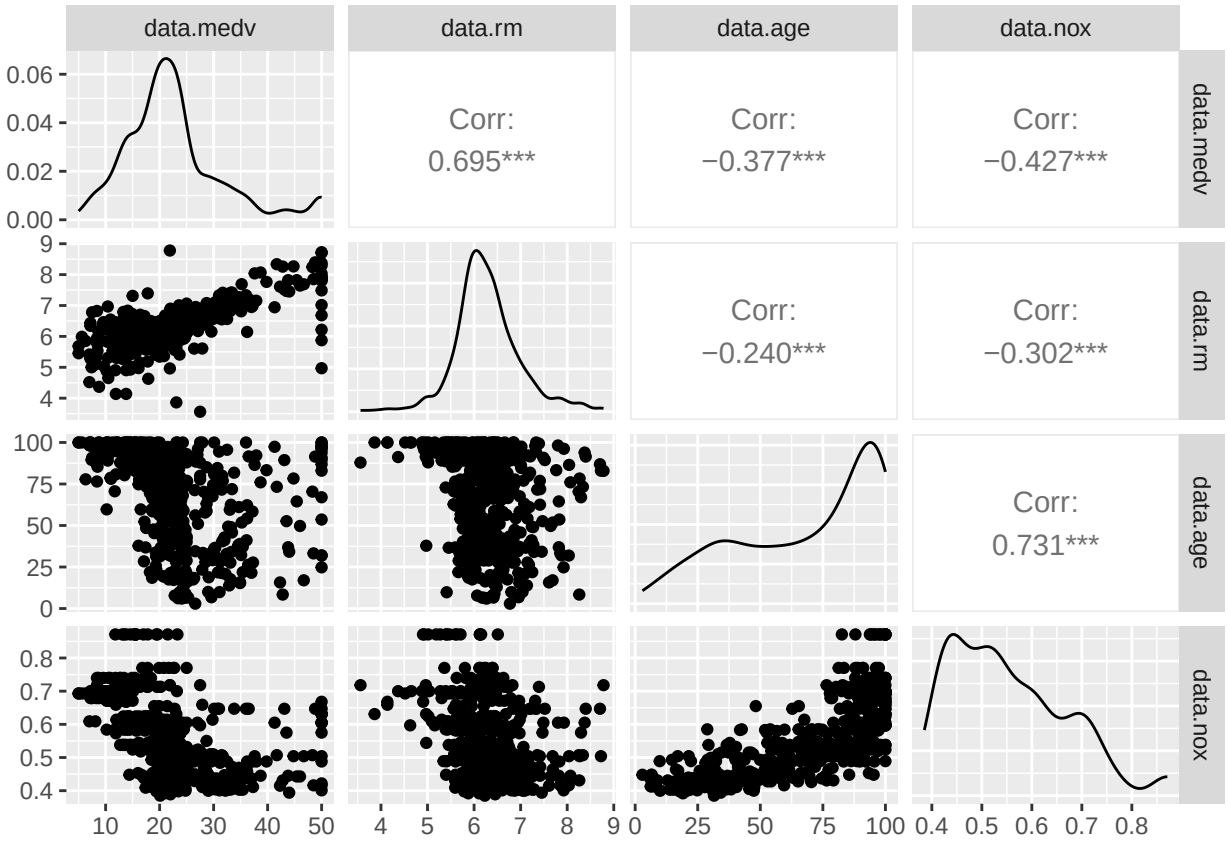
```
##
## Call:
## lm(formula = medv ~ rm + age + nox, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.343  -3.168  -0.539   2.221  40.260
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.08308    3.33919  -5.715 1.88e-08 ***
## rm            8.12542    0.41525  19.568  < 2e-16 ***
## age          -0.03686    0.01449  -2.544 0.011269 *
## nox         -12.47877    3.58434  -3.481 0.000542 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.247 on 502 degrees of freedom
## Multiple R-squared:  0.5413, Adjusted R-squared:  0.5386
## F-statistic: 197.5 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
ggpairs(data.frame(data$medv, data$rm, data$age, data$nox))
```



**1 e) IV Looking at the correlation between Age and NOX, it is 0.731, which is quite high which suggest it has multicollinearity, which means they give the some of the same information for the model.