Homework 2

Anubhav Shankar

2022-10-21

Hat Matrix: Maps the vector of response values to vector fitted values i.e. a vector of dependent values to the vector of fitted/predicted values. It describes the influence of each response variable on the individual fitted values. The diagonal elements of the projection matrix are the leverages.The sum of the diagonal elements should be 1.

```
Age <- c(6,5,4,3,2,2,1,1) # Create a vector of Age values
Price <- c(6,9,8,10,11,12,11,13) # Create a vector of Price values

# Quick verification - OPTIONAL
# is.vector(Age)
# is.vector(Price)

# Let's create a dataframe for ease of use
df <- data.frame(Age, Price)

str(df)
```

```
## 'data.frame':    8 obs. of  2 variables:
##  $ Age  : num  6 5 4 3 2 2 1 1
##  $ Price: num  6 9 8 10 11 12 11 13
```

Let's conduct a Linear Regression with the **Price** as the **dependent variable** and **Age** as the **independent variable**.

```
model <- lm(Price~Age, df)

summary(model) # Examine the result of the linear regression
```

```
##
## Call:
## lm(formula = Price ~ Age, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2500 -0.6875 -0.0625  0.7812  1.2500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.3750     0.6847  19.535 1.17e-06 ***
## Age          -1.1250     0.1976  -5.692  0.00127 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9682 on 6 degrees of freedom
## Multiple R-squared:  0.8437, Adjusted R-squared:  0.8177
## F-statistic:  32.4 on 1 and 6 DF,  p-value: 0.001269
```

Now, let’s examine the **Hat Matrix**.

```
sum(lm.influence(model)$hat) #Examine the sum of leverages i.e. the diagonals of the hat matrix. Ideally it should be 1
```

```
## [1] 2
```

Unfortunately, we get a value of 2. This indicates that the hat matrix is **not symmetrical**. We’ll run the regression model again but this time we’ll drop the intercept term and see how it goes.

```
mod <- lm(Price~ 0 + Age, df) # The addition of zero serves to drop the intercept term from the Linear Regression

summary(mod) # Examine the regression results
```

```
##
## Call:
## lm(formula = Price ~ 0 + Age, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.312 -1.180  4.953  7.867 10.781
##
## Coefficients:
##     Estimate Std. Error t value Pr(>|t|)
## Age   2.2188     0.7354   3.017   0.0195 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.205 on 7 degrees of freedom
## Multiple R-squared:  0.5653, Adjusted R-squared:  0.5032
## F-statistic: 9.103 on 1 and 7 DF,  p-value: 0.01946
```

Let us now examine the Hat matrix for this new model.

```
sum(lm.influence(mod)$hat) #Examine the sum of leverages i.e. the diagonals of the hat matrix. I
deally it should be 1
```

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

We see that the sum of the diagonals equals one (1). **This means that the matrix is symmetric**. However, we'll manually check to see if the diagonal values are in the range [0,1] and also verify that the Hat matrix is idempotent.

```
x <- df$Age # Save the independent variable in a separate matrix

H <- x %*% solve(t(x) %*% x) %*% t(x) # Generate the Hat Matrix

sum(diag(H))
```

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

```
H # Echo the Hat Matrix
```

```
##        [,1]       [,2]       [,3]    [,4]       [,5]       [,6]       [,7]
## [1,] 0.3750 0.31250000 0.25000000 0.18750 0.12500000 0.12500000 0.06250000
## [2,] 0.3125 0.26041667 0.20833333 0.15625 0.10416667 0.10416667 0.05208333
## [3,] 0.2500 0.20833333 0.16666667 0.12500 0.08333333 0.08333333 0.04166667
## [4,] 0.1875 0.15625000 0.12500000 0.09375 0.06250000 0.06250000 0.03125000
## [5,] 0.1250 0.10416667 0.08333333 0.06250 0.04166667 0.04166667 0.02083333
## [6,] 0.1250 0.10416667 0.08333333 0.06250 0.04166667 0.04166667 0.02083333
## [7,] 0.0625 0.05208333 0.04166667 0.03125 0.02083333 0.02083333 0.01041667
## [8,] 0.0625 0.05208333 0.04166667 0.03125 0.02083333 0.02083333 0.01041667
##            [,8]
## [1,] 0.06250000
## [2,] 0.05208333
## [3,] 0.04166667
## [4,] 0.03125000
## [5,] 0.02083333
## [6,] 0.02083333
## [7,] 0.01041667
## [8,] 0.01041667
```

```
H %*% H # Idempotent check
```

```
##        [,1]       [,2]       [,3]    [,4]       [,5]       [,6]       [,7]
## [1,] 0.3750 0.31250000 0.25000000 0.18750 0.12500000 0.12500000 0.06250000
## [2,] 0.3125 0.26041667 0.20833333 0.15625 0.10416667 0.10416667 0.05208333
## [3,] 0.2500 0.20833333 0.16666667 0.12500 0.08333333 0.08333333 0.04166667
## [4,] 0.1875 0.15625000 0.12500000 0.09375 0.06250000 0.06250000 0.03125000
## [5,] 0.1250 0.10416667 0.08333333 0.06250 0.04166667 0.04166667 0.02083333
## [6,] 0.1250 0.10416667 0.08333333 0.06250 0.04166667 0.04166667 0.02083333
## [7,] 0.0625 0.05208333 0.04166667 0.03125 0.02083333 0.02083333 0.01041667
## [8,] 0.0625 0.05208333 0.04166667 0.03125 0.02083333 0.02083333 0.01041667
##            [,8]
## [1,] 0.06250000
## [2,] 0.05208333
## [3,] 0.04166667
## [4,] 0.03125000
## [5,] 0.02083333
## [6,] 0.02083333
## [7,] 0.01041667
## [8,] 0.01041667
```

```
t(H)
```

```
##        [,1]       [,2]       [,3]    [,4]       [,5]       [,6]       [,7]
## [1,] 0.3750 0.31250000 0.25000000 0.18750 0.12500000 0.12500000 0.06250000
## [2,] 0.3125 0.26041667 0.20833333 0.15625 0.10416667 0.10416667 0.05208333
## [3,] 0.2500 0.20833333 0.16666667 0.12500 0.08333333 0.08333333 0.04166667
## [4,] 0.1875 0.15625000 0.12500000 0.09375 0.06250000 0.06250000 0.03125000
## [5,] 0.1250 0.10416667 0.08333333 0.06250 0.04166667 0.04166667 0.02083333
## [6,] 0.1250 0.10416667 0.08333333 0.06250 0.04166667 0.04166667 0.02083333
## [7,] 0.0625 0.05208333 0.04166667 0.03125 0.02083333 0.02083333 0.01041667
## [8,] 0.0625 0.05208333 0.04166667 0.03125 0.02083333 0.02083333 0.01041667
##            [,8]
## [1,] 0.06250000
## [2,] 0.05208333
## [3,] 0.04166667
## [4,] 0.03125000
## [5,] 0.02083333
## [6,] 0.02083333
## [7,] 0.01041667
## [8,] 0.01041667
```

From the above results we can conclude that the Hat Matrix is symmetrical and idempotent. Also, all the diagonal values are in the range [0,1].

```
set.seed(123) # To ensure reproducability

A<-matrix(sample(1:5,16, replace=TRUE),4,4)
B<-matrix(sample(1:5,16, replace=TRUE),4,4)

sum(diag(A %*% B))
```

```
## [1] 122
```

```
sum(diag(B %*% A))
```

```
## [1] 122
```

As the sum of the respective matrix diagonal's equal 1, we can conclude that the property Trace(AB) = Trace(BA).

References:

1.) https://stats.stackexchange.com/questions/125886/linear-model-trace-of-the-hat-matrix-in-r#125887 (https://stats.stackexchange.com/questions/125886/linear-model-trace-of-the-hat-matrix-in-r#125887)
2.) https://math.stackexchange.com/questions/209743/what-is-a-idempotent-matrix (https://math.stackexchange.com/questions/209743/what-is-a-idempotent-matrix)