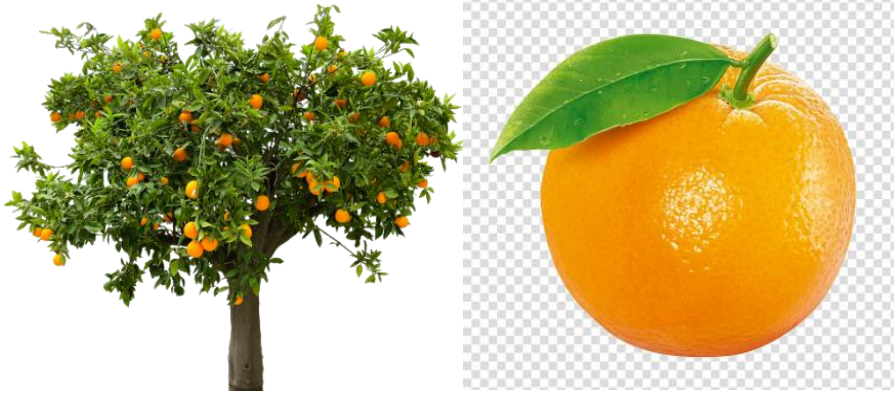
NAME-Anmol Sinha
E_MAIL- 1805553@kiit.ac.in

Linear Regression

- In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

- In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.

- After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.



{ R-Project Code }

```
> # Orange Dataset
> data("Orange")
> head(Orange)
```

```
   Tree  age circumference
1     1  118            30
2     1  484            58
3     1  664            87
4     1 1004           115
5     1 1231           120
6     1 1372           142
```

> Orange

```
   Tree  age circumference
1     1  118            30
2     1  484            58
3     1  664            87
4     1 1004           115
5     1 1231           120
6     1 1372           142
7     1 1582           145
8     2  118            33
9     2  484            69
10    2  664           111
11    2 1004           156
12    2 1231           172
13    2 1372           203
14    2 1582           203
15    3  118            30
16    3  484            51
17    3  664            75
18    3 1004           108
19    3 1231           115
20    3 1372           139
21    3 1582           140
22    4  118            32
23    4  484            62
24    4  664           112
25    4 1004           167
```

```
26       4 1231                179
27       4 1372                209
28       4 1582                214
29       5  118                 30
30       5  484                 49
31       5  664                 81
32       5 1004                125
33       5 1231                142
34       5 1372                174
35       5 1582                177
> cor(Orange$circumference, Orange$age)
[1] 0.9135189
> cor(Orange$age, Orange$circumference)
[1] 0.9135189
> plot(Orange$age, Orange$circumference)
> plot(Orange$circumference, Orange$age)
> model <- lm(age ~ circumference, data= Orange)
> summary(model)

Call:
lm(formula = age ~ circumference, data = Orange)

Residuals:
     Min       1Q   Median       3Q      Max
-317.88 -140.90   -17.20    96.54   471.16

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      16.6036    78.1406   0.212    0.833
circumference     7.8160     0.6059  12.900 1.93e-14 *
---
Signif. codes:  0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 203.1 on 33 degrees of freedom
```

```
Multiple R-squared:    0.8345,  Adjusted R-squared:    0.8295
F-statistic: 166.4 on 1 and 33 DF,    p-value: 1.931e-14

> # Prediction
> predict(model, data.frame("circumference"=100))
        1
798.2035
> predict(model, data.frame("age"=100))
Error in eval(predvars, data, env) : object 'circumference' not
found
> model <- lm(circumference ~ age, data= Orange)
> summary(model)

Call:
lm(formula = circumference ~ age, data = Orange)

Residuals:
     Min       1Q   Median       3Q      Max
-46.310 -14.946   -0.076   19.697   45.111

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.399650   8.622660   2.018   0.0518 .
age          0.106770   0.008277  12.900 1.93e-14 *
---
Signif. codes:  0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.74 on 33 degrees of freedom
Multiple R-squared:    0.8345,  Adjusted R-squared:    0.8295
F-statistic: 166.4 on 1 and 33 DF,    p-value: 1.931e-14

> predict(model, data.frame("age"=100))
        1
28.07668
```

```
> plot(Orange$circumference, Orange$age, xlab='circumference',
ylab='Age')
> abline(model,col="red",lty=2,lwd=3)
> model <- lm(age ~ circumference, data= Orange)
> plot(Orange$circumference, Orange$age, xlab='circumference',
ylab='Age')
> abline(model,col="red", lty=2,lwd=3)
```