

ECM2002 – Machine Learning Algorithms Lab L1+L2

Register Number: 19BLC1186

Name: Tarun Sidhu

Lab Exercise No: 2

Date: 15/2/2021

Dataset: Covid-19 Dataset

Task: Carry out linear regression analysis on the given dataset and summarise observations

Conclusion:

Code:

```
library(ISLR)
attach(covid_19_data)
plot(Confirmed~Recovered)
fit1=lm(Confirmed~Recovered)
abline(fit1)
summary(fit1)
fit2=lm(Confirmed~ Recovered + I(Recovered^2))
summary(fit2)
points(Recovered, fitted(fit2), col ="Red", pch=20)
fit3=lm(Confirmed~ Recovered + I(Recovered^2) + I(Recovered^3))
summary(fit3)
points(Deaths, fitted(fit3), col ="blue", pch=20)
fit4=lm(Confirmed~ poly(Recovered,4))
summary(fit4)
```

```
points(Recovered, fitted(fit4), col="green", pch=20)
```

Output:

```
View(covid_19_data)
```

```
> library(ISLR)
```

```
> attach(covid_19_data)
```

```
> plot(Confirmed~Recovered)
```

```
> fit1=lm(Confirmed~Recovered)
```

```
> abline(fit1)
```

```
> summary(fit1)
```

Call:

```
lm(formula = Confirmed ~ Recovered)
```

Residuals:

Min	1Q	Median	3Q	Max
-3541331	-29857	-26719	-12076	3018813

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.022e+04	2.783e+02	108.6	<2e-16 ***
Recovered	5.487e-01	2.042e-03	268.7	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123700 on 205949 degrees of freedom

Multiple R-squared: 0.2595, Adjusted R-squared: 0.2595
F-statistic: 7.219e+04 on 1 and 205949 DF, p-value: < 2.2e-16

```
> fit2=lm(Confirmed~ Recovered + I(Recovered^2))  
> summary(fit2)
```

Call:

```
lm(formula = Confirmed ~ Recovered + I(Recovered^2))
```

Residuals:

Min	1Q	Median	3Q	Max
-1338450	-19409	-18642	-12489	3029618

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.942e+04	2.510e+02	77.38	<2e-16 ***
Recovered	1.091e+00	2.919e-03	373.58	<2e-16 ***
I(Recovered^2)	-2.254e-07	9.525e-10	-236.66	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 109700 on 205948 degrees of freedom

Multiple R-squared: 0.4179, Adjusted R-squared: 0.4178

F-statistic: 7.391e+04 on 2 and 205948 DF, p-value: < 2.2e-16

```
> points(Recovered, fitted(fit2), col ="Red", pch=20)
```

```
> fit3=lm(Confirmed~ Recovered + I(Recovered^2) + I(Recovered^3))
> summary(fit3)
```

Call:

```
lm(formula = Confirmed ~ Recovered + I(Recovered^2) + I(Recovered^3))
```

Residuals:

Min	1Q	Median	3Q	Max
-1062424	-16627	-16348	-13120	3032661

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.638e+04	2.518e+02	65.04	<2e-16 ***
Recovered	1.304e+00	4.194e-03	310.87	<2e-16 ***
I(Recovered^2)	-4.769e-07	3.711e-09	-128.51	<2e-16 ***
I(Recovered^3)	4.186e-14	5.976e-16	70.05	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 108400 on 205947 degrees of freedom

Multiple R-squared: 0.4314, Adjusted R-squared: 0.4314

F-statistic: 5.208e+04 on 3 and 205947 DF, p-value: < 2.2e-16

```
> points(Deaths, fitted(fit3), col ="blue", pch=20)
```

```
> fit4=lm(Confirmed~ poly(Recovered,4))
```

```
> summary(fit4)
```

Call:

```
lm(formula = Confirmed ~ poly(Recovered, 4))
```

Residuals:

Min	1Q	Median	3Q	Max
-1196983	-18103	-17694	-12773	3030938

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.540e+04	2.379e+02	190.87	<2e-16 ***
poly(Recovered, 4)1	3.323e+07	1.080e+05	307.81	<2e-16 ***
poly(Recovered, 4)2	-2.595e+07	1.080e+05	-240.40	<2e-16 ***
poly(Recovered, 4)3	7.592e+06	1.080e+05	70.33	<2e-16 ***
poly(Recovered, 4)4	4.349e+06	1.080e+05	40.29	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 108000 on 205946 degrees of freedom

Multiple R-squared: 0.4358, Adjusted R-squared: 0.4358

F-statistic: 3.978e+04 on 4 and 205946 DF, p-value: < 2.2e-16

```
> points(Recovered, fitted(fit4), col="green", pch=20)
```

Plot:

