Improvement to nls() : Solution to Tests(Easy)

Arkajyoti Bhattacharjee

Theory :

The logistic sigmoid growth curve is given by -

$$y = f(t) = \frac{L}{1 + e^{-k(t-x_0)}} ...(\Delta)$$

Here, the growth $y = f(t)$ is a function of time t.
x_0 : the x value of the sigmoid's midpoint.
L : the curve's maximum value.
k : the logistic growth rate or steepness of the curve.

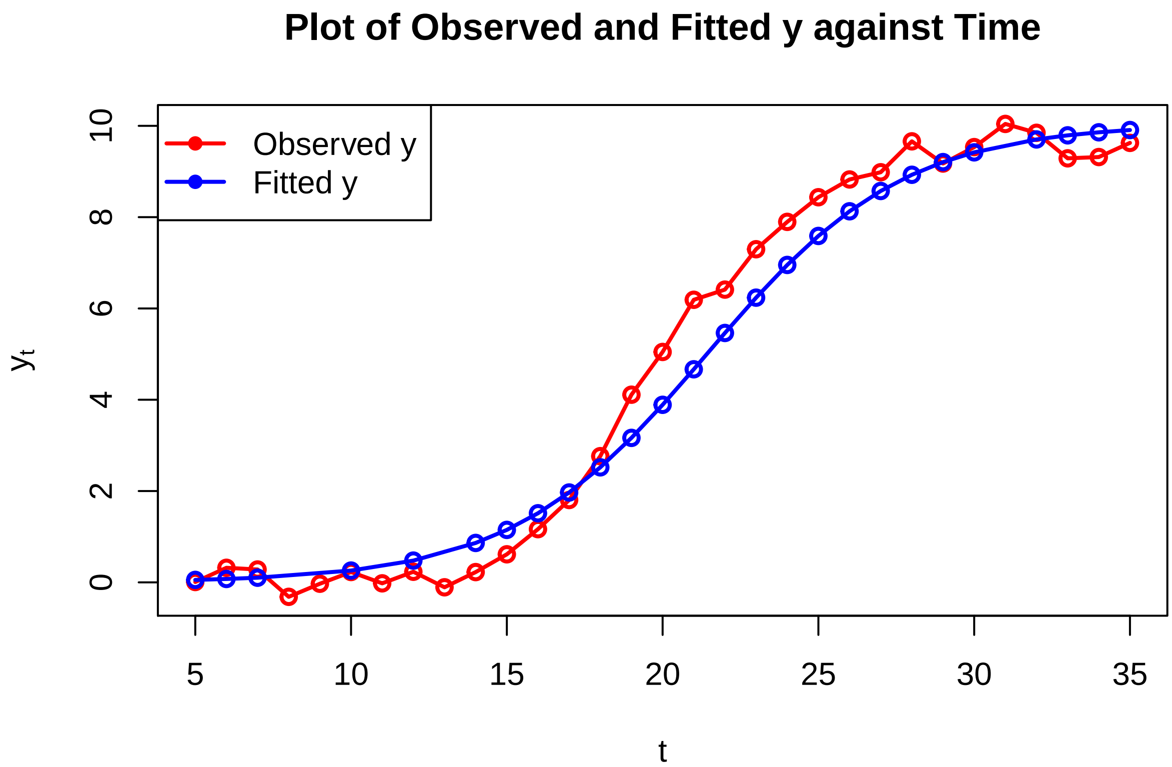
Now, given the test data, we need to estimate a logistic sigmoid growth curve of the form as in Eq. Δ.

R Code :

```r
### "Improvements to nls()" - GSOC'21
## Easy

# test data
time <- c( 5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,
 30, 31, 32, 33, 34, 35)
y <- c( 0.0074203,  0.3188325,  0.2815891, -0.3171173, -0.0305409,  0.2266773,
  -0.0216102,  0.2319695, -0.1082007,  0.2246899,  0.6144181,  1.1655192,
  1.8038330,  2.7644418,  4.1104270,  5.0470456,  6.1896092,  6.4128618,
  7.2974793,  7.8965245,  8.4364991,  8.8252770,  8.9836204, 9.6607736,
  9.1746182,  9.5348823, 10.0421165,  9.8477874,  9.2886090,  9.3169916,
  9.6270209 )

# Plot y vs. time
plot(time,y,xlab="t",ylab=expression(y[t]),
main="Plot of Observed and Fitted y against Time",lwd=2,col="red",type="o")

# clean the data
time <- time[y>=0]
```

```r
y <- y[y>=0]

## Approximate estimation of x0,k, L based on above plot

#approximate value of x0
x0 <- mean(time)

# approximate value of L
L<-max(y)

# approximate value of k
# I remove y for which ln(L/y-1) is NA
time<-time[-which(L/y<=1)]
y<-y[-which(L/y<=1)]

k <- log(L/y-1)/(x0-time)
k<-mean(k)

# plot of the fitted values
lines(time, L/(1+exp(-k*(time-x0))),type="o",lwd=2,col="blue")
legend("topleft",legend=c("Observed y","Fitted y"),
col=c("red","blue"),lwd=2,,pch=16)

# print the parameter values
cat("The estimated value of x0 is : ",x0,"\n")
cat("The estimated value of L is : ",L,"\n")
cat("The estimated value of k is : ",k,"\n")
```

Output :




```
## The estimated value of x0 is :  21.44444

## The estimated value of L is :  10.04212

## The estimated value of k is :  0.3175419
```

Explanation :

I first plot the graph to know how the given $y's$ look as a function of time. Since, the logistic function is strictly positive, I first discard the negative y values from the data and the corresponding time points. Then, to estimate x_0 we take the mean of the time points as the mid-value of y should correspond to the mid-value of $time$.Now, L being the maximum value of the growth curve, I take L to be the maximum of the given $y's$. Finally, given y, $time$ and estimated x_0 , L, I compute $k's$ by plugging these values in Eq. Δ and then take the mean of the resulting vector of $k's$. In this process, I had to compute $k's$ as $k = \frac{log(L/y-1)}{x0-time}$. So, I deleted those $y's$ and corresponding $time$ points for which $\frac{L}{y} \leq 1$.

Using the above procedure, my estimated logistic growth function is -

$$\hat{y} = \hat{f}(t) = \frac{10.04212}{1 + e^{-0.3175419(t-21.44444)}}$$