Loess Model: what if x-variable is not in order?

Yifan Li, yli2763@uwo.ca

2017-01-21

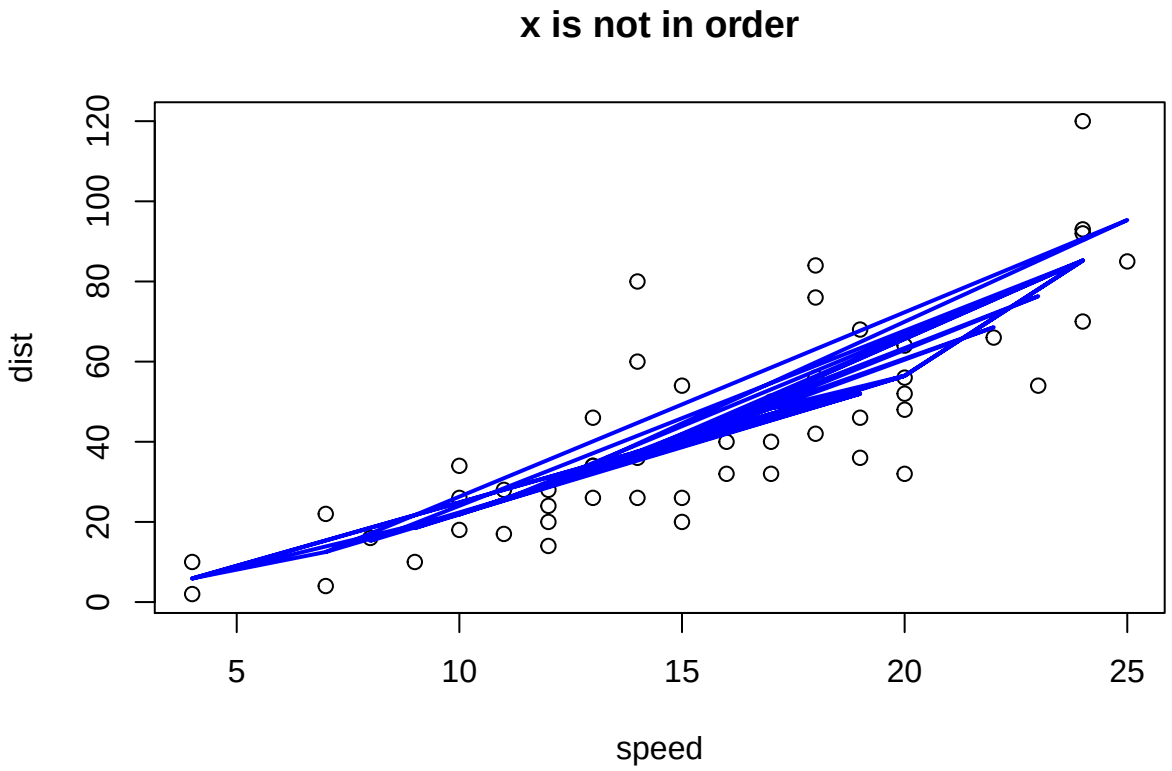
In the "carsNew" dataset, we will find that the x variable (column "speed") is not in order. We regard the distance (column "dist") as the dependent variable.

```
head(carsNew) #here the x variable is in a random order.
```

```
##   speed dist
## 1    18   56
## 2    19   68
## 3    12   28
## 4     9   10
## 5    10   18
## 6    12   24
```

We intend to apply loess model to fit the data. Usually we need the x variable to be in ascending order, but what if it is not?

```
#fit the data as it is
mod1 <- loess(dist~speed, carsNew)
with(carsNew, plot(speed, dist, main="x is not in order"))
lines(carsNew$speed, mod1$fitted, lwd=2, col="blue")
```



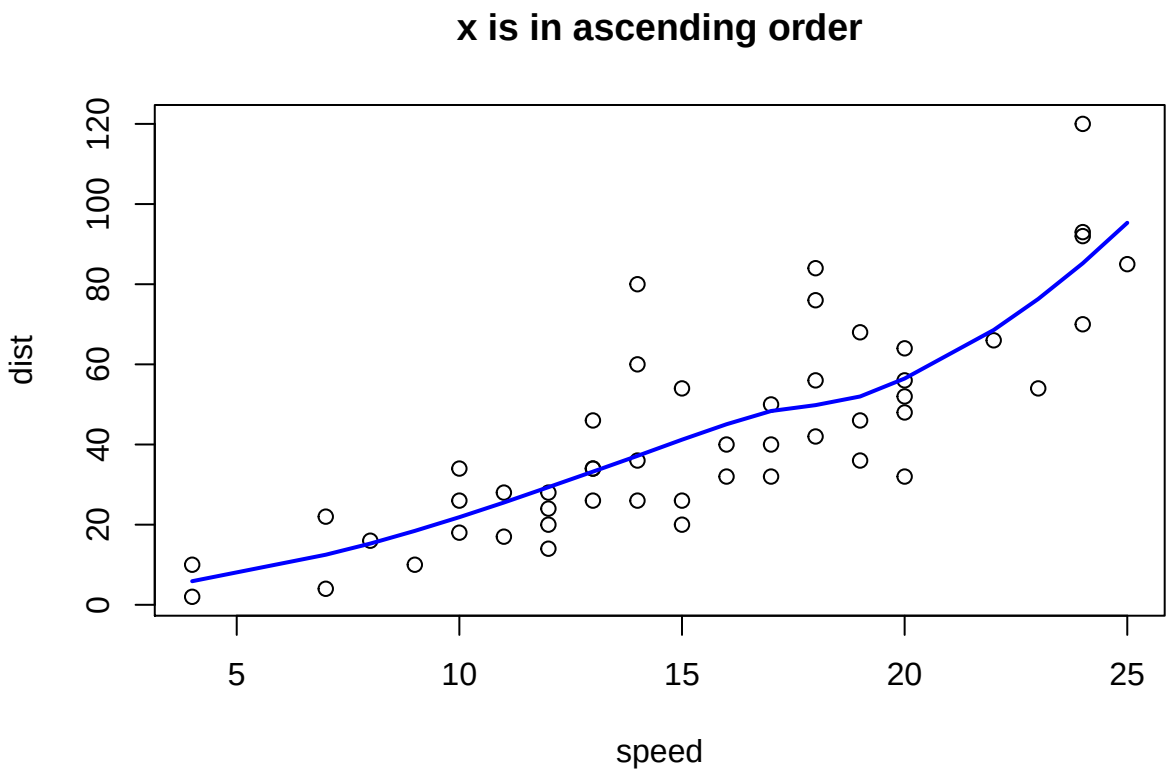
We can find that the plot is in a mess, because the function lines() will connect the fitted points according to the order in the current x vector (speed), which is not in ascending order. Let's look at the the fitted points in details:

```
##      speed fitted value
## [1,]    18     49.82481
## [2,]    19     51.98670
## [3,]    12     29.35039
## [4,]     9     18.44657
## [5,]    10     21.86532
## [6,]    12     29.35039
```

Therefore, in this case, it will connect points in this order: *speed* : $18 \to 19 \to 12 \to 9 \to 10 \to \cdots$. This will surely give us a series of messed straight lines.

In order to get the correct loess plot, we need to order the data frame with respect to the x-variable.

```
#create a new data frame
#with the speed in ascending order
IndexSort <- order(carsNew$speed)
carsNew2 <- carsNew[IndexSort,]
#build up the loess model again
mod2 <- loess(dist~speed, carsNew2)
with(carsNew2, plot(speed, dist, main="x is in ascending order"))
lines(carsNew2$speed, mod2$fitted, lwd=2, col="blue")
```



Since the x-variable is in increasing order now, the fitted points will be jointed in order (from left to right), then we can get the needed smooth plot for the loess model.

In fact, by comparing previous two models, we can find that there is no significant difference between them.

```
anova(mod1, mod2)
```

```
## Model 1: loess(formula = dist ~ speed, data = carsNew)
## Model 2: loess(formula = dist ~ speed, data = carsNew2)
##
## Analysis of Variance:   denominator df 44.62
##
##       ENP   RSS F-value Pr(>F)
## [1,] 4.78 10364
## [2,] 4.78 10364     Inf
```

Actually, according to the algorithm of loess model, the weight only depends on the value of data point rather than the order, so the two models are exactly the same. The order will matter when drawing the fitted plots.

In fact, x-variable can also be set in descending order, which will produce the same result.

```
#create a new data frame
#with the speed in descending order
IndexSort <- order(carsNew$speed, decreasing = TRUE)
carsNew3 <- carsNew[IndexSort,]
#build up the loess model again
mod3 <- loess(dist~speed, carsNew3)
with(carsNew3, plot(speed, dist, main="x is in descending order"))
lines(carsNew3$speed, mod3$fitted, lwd=2, col="blue")
```

The only difference is that the fitted points are connected from right to left. Here this will not give a mirror of the previous plot, because the coordinates have been fixed when we use the plot() to draw the data points, then the lines() function can only connect the fitted points on the given xlab.

Further comments about the effects from the order of x-variable on loess model:

As we learned in class, the local neighborhood weight for one data point (x_i, y_i) at the given point x is

$$w_i(x) = T\left(\frac{\Delta_i(x)}{\Delta(x,\alpha)}\right),$$

where $\Delta_i(x) = |x - x_i|$ only depend on the value of x_i, and

$$\Delta(x,\alpha) = \begin{cases} \Delta_{([\alpha n])}(x), & 0 < \alpha \leq 1 \\ \alpha\Delta_{(n)}(x), & \alpha > 1 \end{cases},$$

does not depend on i either, because $\Delta_{(k)}(x)$ is the $k-$th value of $\Delta_i(x),\ i = 1, 2, \cdots, n$, which consider all the data points no matter what the order of x_i is. Therefore, we can find that $w_i(x)$ does not really depend on the order of the data point (which is i). To be specific, if we change the location of one data point in the data set from the first one to the sixth one, the weight attached to that data point will not change. Finally, we can conclude that the order of x-variable does not affect the loess model itself.