ASM Practice

Smoothing and regression splines

Maria Gkotsopoulou & Ricard Monge Calvo & Amalia Vradi

14/12/2019

1

Estimate the regression function m(instant) of cnt as a function of instant using a cubic regression splines estimated with the R function smooth.splines and choosing the smoothing parameter by Generalized Cross Validation.

a)

The chosen Smoothness penalization hyperparameter λ by GCV is 1.0050377×10^{-7}.

b)

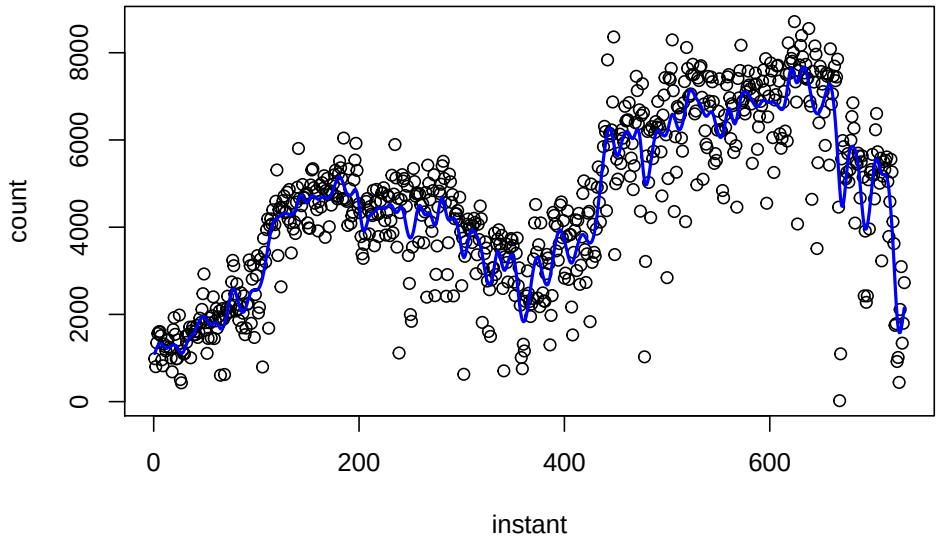
The corresponding equivalent number of degrees of freedom of the spline regression's linear estimator is 93.

c)

140 knots were used.

d)

We show a scatterplot of the data points with the fitted spline regression:



e)

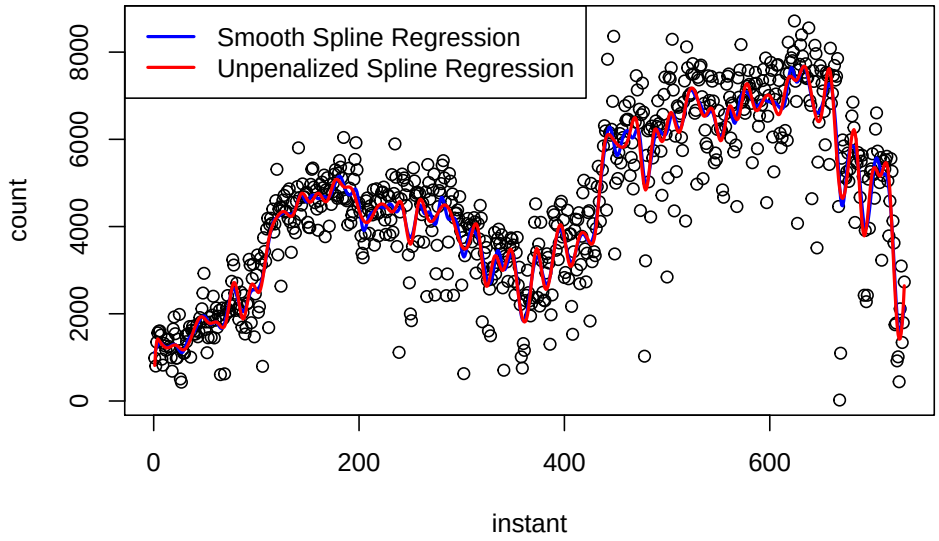
We estimate now m(instant) by unpenalized regression splines combining the R functions *bs* and *lm*, using the knots where *n.knots* is the previous value of df minus 4.

```
x <- bikes$instant # x
y <- bikes$cnt # y
n <- length(x)
n.knots <- sm.sp.1$df -4
```

```
my.knots <- quantile(x,((1:n.knots)-.5)/n.knots)
b.kn <- range(x)+c(-1,1)*.1*(diff(range(x)))
X.bs <- bs(x, knots=my.knots, Boundary.knots = b.kn)
sreg <- lm(y~X.bs)
```

f)

Plot the scatter plot with the different spline regressions



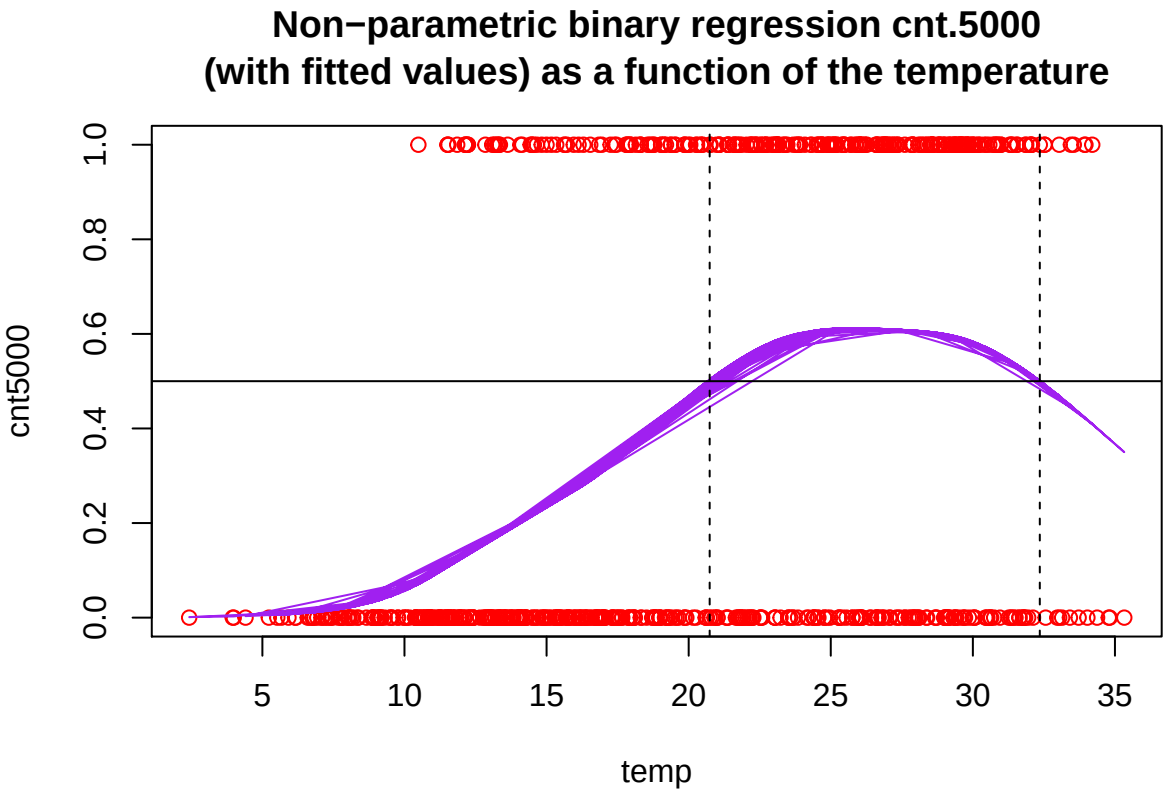
2

We define a new variable cnt.5000 taking the value 1 for days such that the number of total rental bikes is larger than or equal to 5000, or 0 otherwise.

cnt.5000	n
0	445
1	286

a)

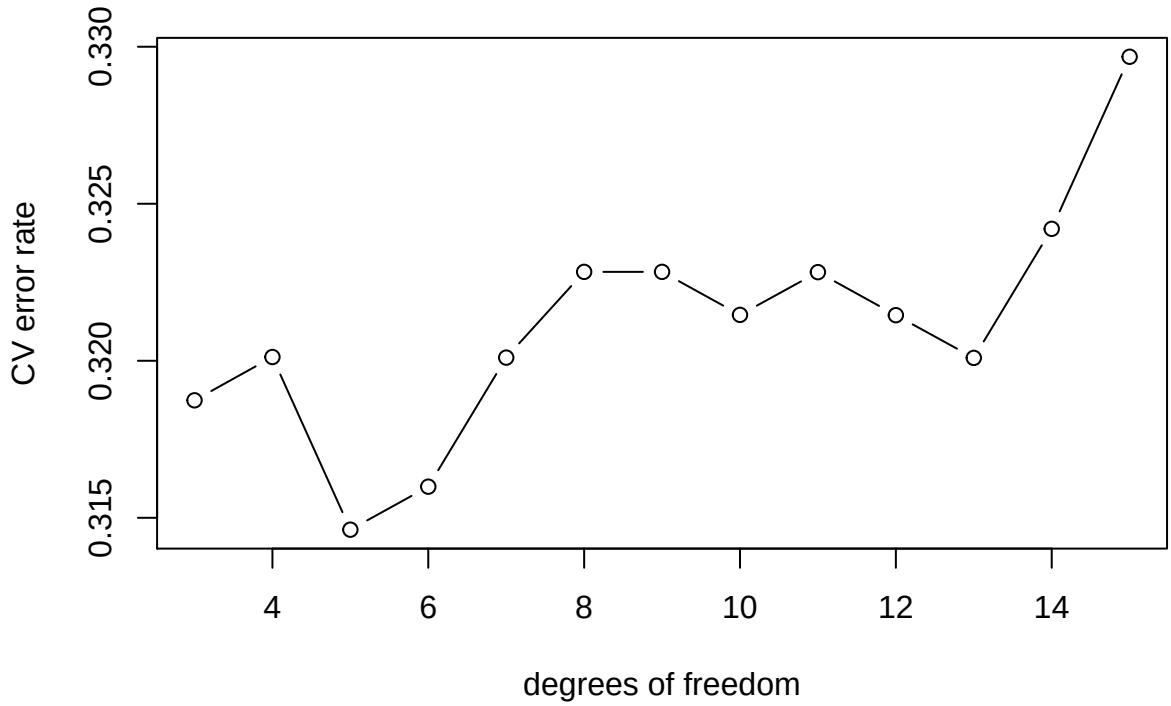
We use the function *logistic.IRWLS.splines* to fit the non-parametric binary regression *cnt.5000* as a function of the temperature, using *df=6*.



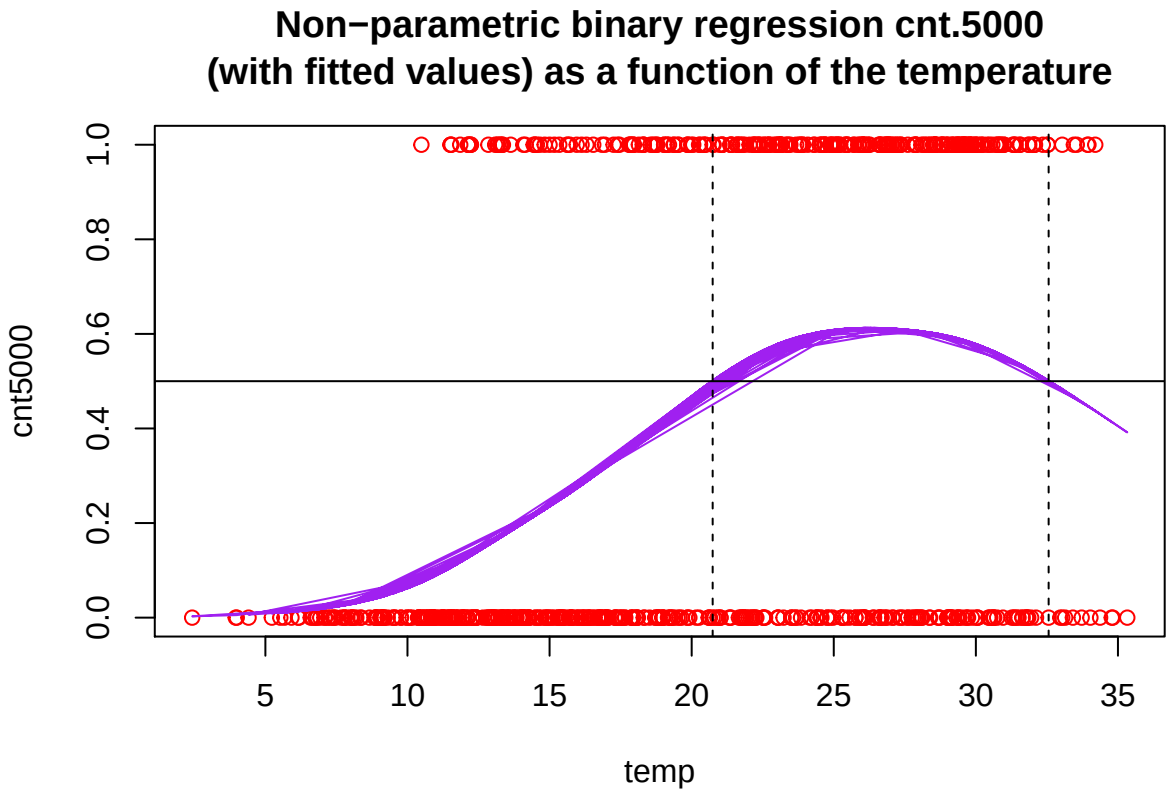
The range of temperatures that the $Pr(cnt >= 5000|temp)$ is larger than 0.5 is from $20.7°C$ to $32.4°C$.

b)

We now choose the parameter df by *k-fold* cross validation with $k = 5$ and using $df.v = 3 : 15$ as the set of possible values for *df*.



The minimum is obtained at 5 degrees of freedom. We now refit the non-parametric binary regression cnt.5000 as a function of the temperature using the obtained df.



The range of temperatures that the $Pr(cnt >= 5000|temp)$ is larger than 0.5 is from $20.7°C$ to $32.6°C$.