

Application

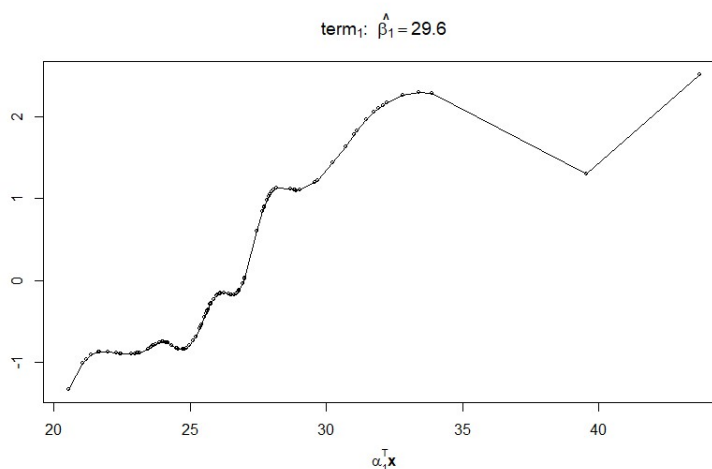
Refer to the Air Quality data described previously, and the analyses we have done with Ozone as the response variable, and the five explanatory variables (including the two engineered features).

1. Use PPR to model the relationship between Ozone and all five explanatories (you don't need to use the `scale()` function in the formula). Use `max.terms=5`. Also use the `gcv.spline` smoothing method as shown in the example.

```
2 # Author: Injun Son
3 # Date: October 25, 2020
4
5 library(dplyr)
6 library(MASS) # For ridge regression
7 library(glmnet) # For LASSO
8 library(mgcv)
9 source("Helper Functions.R")
10 data = na.omit(airquality[, 1:4])
11 data$TWcp = data$Temp*data$Wind
12 data$Twrat = data$Temp/data$Wind
13
14 ### We fit PPR models using the ppr() function. This function uses data frame-
15 ### model formula syntax. We also need to specify the number of terms to
16 ### include in our model, and what type of smoother to use. Number of terms is
17 ### set using the nterms input, and type of smoother is set using sm.method.
18 ### Options for smoothers include "spline" (smoothing spline where you have
19 ### to also specify the number of degrees of freedom using df) and "gcv.spline"
20 ### (smoothing spline with degrees of freedom chosen by GCV).
21 fit.ppr.1 = ppr(Ozone ~ ., data = data,
22                 max.terms = 5, nterms = 1, sm.method = "gcv.spline")
23 plot(fit.ppr.1)
24
25 fit.ppr.2 = ppr(Ozone ~ ., data = data,
26                 max.terms = 5, nterms = 2, sm.method = "gcv.spline")
27 plot(fit.ppr.2)
```

(a) Specify `nterms=1`.

- Show the plot of the spline for the selected projection
- Report the training SSE from the summary



```
> summary(fit.ppr.1)
Call:
ppr(formula = Ozone ~ ., data = data, max.terms = 5, nterms = 1,
     sm.method = "gcv spline")

Goodness of fit:
 1 terms  2 terms  3 terms  4 terms  5 terms
24344.08 17849.81 16618.96 16359.43 16672.41

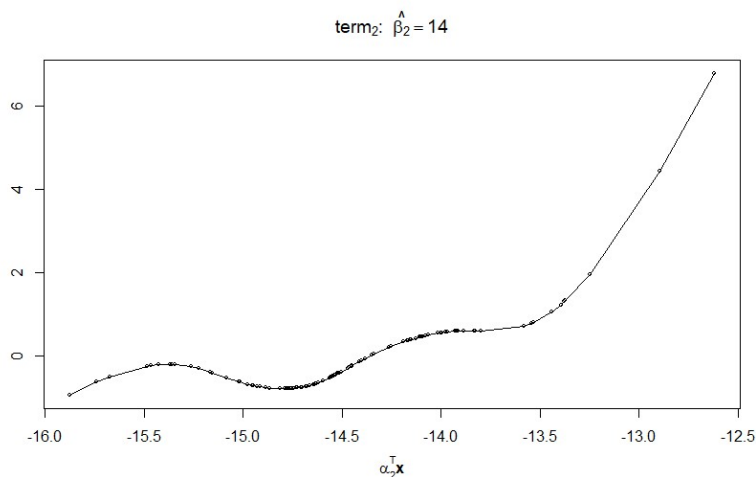
Projection direction vectors ('alpha'):
      Solar.R      Wind      Temp      Twcp      Twrat
0.005480085  0.838706310  0.236497771 -0.007993799  0.490455500

Coefficients of ridge terms ('beta'):
 term 1
29.63104

Equivalent df for ridge terms:
 term 1
11.65
```

(b) Repeat with nterms=2.

i. **Show the plot** of the spline for the selected projection



ii. Report the training SSE from the summary

```
> summary(fit.ppr.2)
Call:
ppr(formula = Ozone ~ ., data = data, max.terms = 5, nterms = 2,
     sm.method = "gcv spline")

Goodness of fit:
 2 terms  3 terms  4 terms  5 terms
17849.81 16618.96 16359.43 16672.41

Projection direction vectors ('alpha'):
      term 1      term 2
Solar.R  5.225461e-03  8.417329e-05
Wind     8.771566e-01 -9.636819e-01
Temp     2.565199e-01 -2.118937e-01
Twcp     -8.689359e-03  1.331013e-02
Twrat     4.058215e-01  1.619910e-01

Coefficients of ridge terms ('beta'):
 term 1  term 2
31.35046 13.99775

Equivalent df for ridge terms:
 term 1  term 2
11.78   6.74
```

