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 GAM to model the relationship between Ozone and all five explanatories:
(a) Print out and look at the summary from the gam object.

i. **Show the summary**

```
# Title: STAT 452 Exercise 7 L11Q11
# Author: Injun Son
# Date: October 25, 2020

library(dplyr)
library(MASS)   # For ridge regression
library(glmnet) # For LASSO
library(mgcv)
source("Helper Functions.R")
data = na.omit(airquality[, 1:4])
data$TWcp = data$Temp*data$Wind
data$TWrat = data$Temp/data$Wind

### The gam() function has similar syntax to lm(). Specify a model formula
### and a data frame, but for each predictor, you can optionally put it
### inside a function called s(). See below for a demonstration.
fit.gam = gam(Ozone ~ s(Solar.R) + s(Wind) + s(Temp) + s(TWcp) + s(TWrat),
              data = data)

### Get information about the GAM fit using the summary() function.
summary(fit.gam)
```

```
> summary(fit.gam)

Family: gaussian
Link function: identity

Formula:
Ozone ~ s(Solar.R) + s(Wind) + s(Temp) + s(TWcp) + s(TWrat)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   42.099      1.251   33.66   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
             edf Ref.df     F  p-value
s(Solar.R) 2.595  3.236 3.612  0.01447 *
s(Wind)    1.000  1.000 6.611  0.01175 *
s(Temp)    4.986  6.080 6.232 1.37e-05 ***
s(TWcp)    4.679  5.715 4.451  0.00104 **
s(TWrat)   7.210  7.977 7.874 1.44e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.843   Deviance explained = 87.2%
GCV = 215.32  Scale est. = 173.67    n = 111
```

ii. **According to the summary, are there any variables that seem unimportant? If so which ones?**

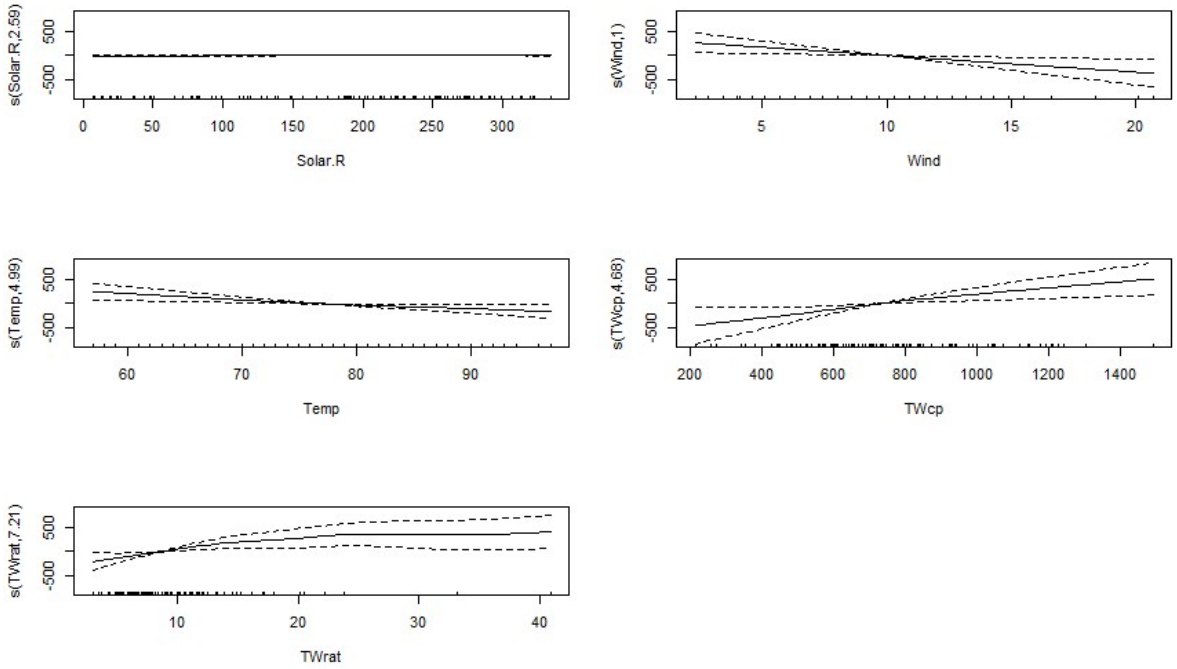
➔ No, all the p-values are lower than 0.05 which means they are significant.

iii. Which variables seem to have the most nonlinear influence on Ozonw, according to their degrees of freedom?

-> TWrat has the highest edf so I guess TWrat would have the most nonlinear influence on Ozone.

(b) Plot the marginal splines for each variable, making sure that the plot is large enough for you to see the patterns and the error bounds

i. **Present these plots.**



ii. For the two most nonlinear patterns identified in part (a), comment on the shape of the patterns. **Does the nonlinearity suggest a clear nonmonotone relationship, or mostly just vary the rates of increasing and decreasing trends?**

➔ Temp and TWrat have the highest edf, so two will be most non linear. For Temp, the nonlinearity suggests just the increasing and decreasing trends but for TWrat the nonlinearity suggests a nonmonotone relationship.