Hwk 07 Solutions L11

Due on Oct 30, 2020

```
## Warning: package 'nlme' was built under R version 4.0.3
```

Applications

Question 1

Part (a)

(i) See below for a summary of the GAM model fit to predict ozone using the other variables in the air quality dataset.

```
##
## 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  0.01453 *
## s(Wind)    1.000  1.000 6.611  0.01177 *
## s(Temp)    4.986  6.080 6.232 1.63e-05 ***
## s(TWcp)    4.679  5.715 4.451  0.00109 **
## s(TWrat)   7.210  7.977 7.874  < 2e-16 ***
## ---
## 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) All predictors look important, judging by the tests for each term (all p-values are <0.05).
(iii) The ratio of temperature to wind appears to have the most nonlinear relationship with ozone, using 7.2 DF for the spline. followed by temperature and the temperature-wind cross-product.

Part (b)

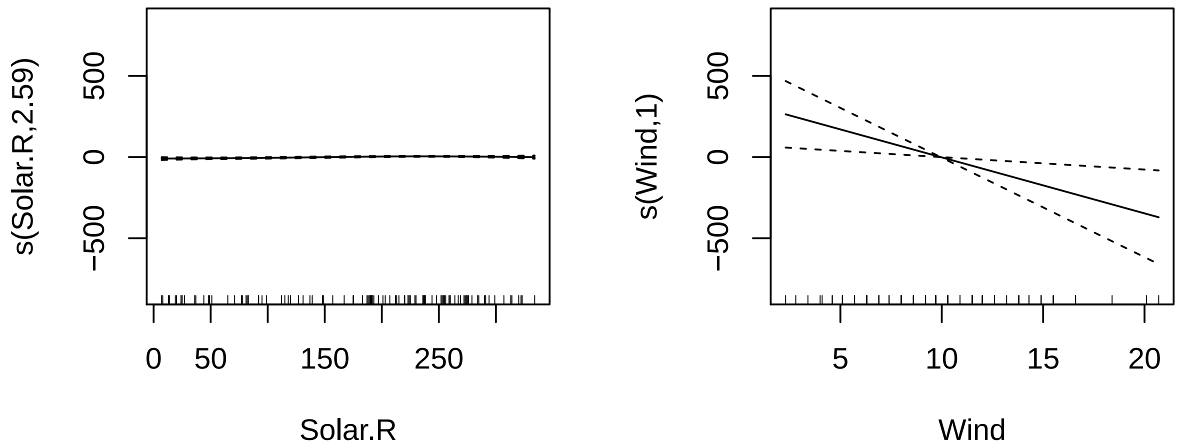

Figure 1: GAM Marginal Fits - I

(i) See Figures 1, 2 and 3 for the marginal splines from our GAM fit.
(ii) The fit for the temperature-wind speed ratio is clearly nonlinear, although it is almost exclusively increasing over the range of the data. The fit for temperature alone looks approximately linear, other than some minor curvature at high temperatures. The relationship between temperature and ozone is monotonically decreasing.

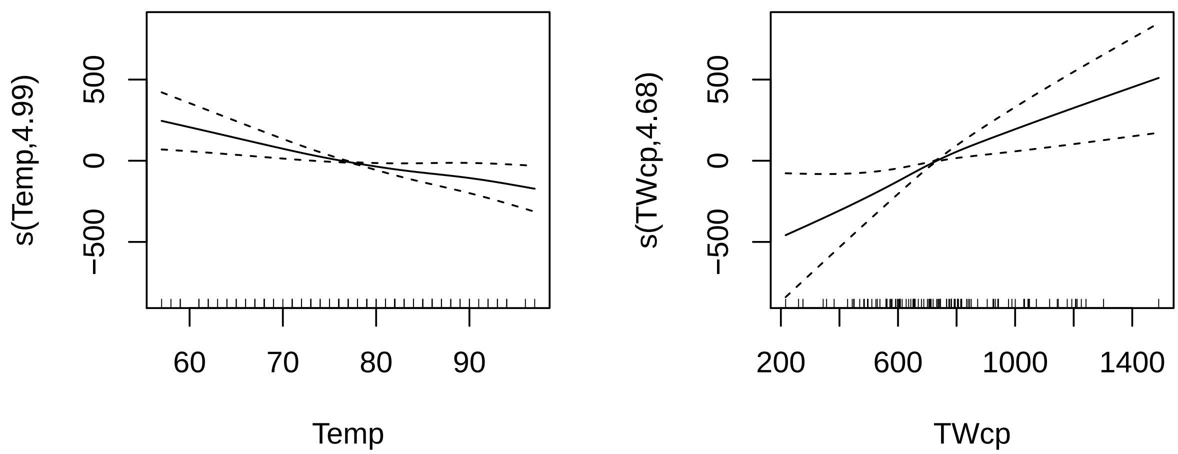

Figure 2: GAM Marginal Fits - II

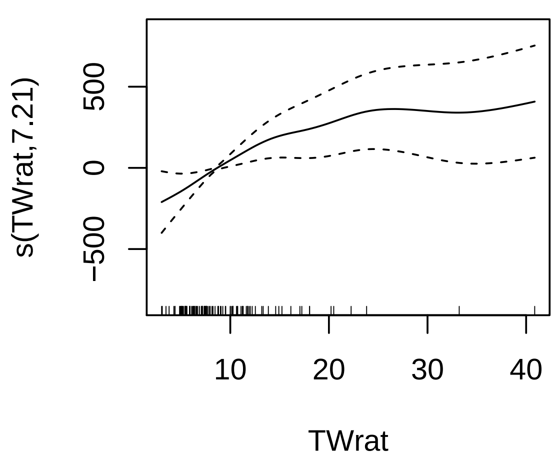

Figure 3: GAM Marginal Fits - III

Table 1: CV MSPEs

	1	2	3	4	5	6	7	8	9	10	Full
LS	260	323	517	106	192	374	213	593	1056	729	436
Step	328	404	577	107	178	369	233	635	1074	787	469
Ridge	281	364	539	108	185	372	220	588	1043	725	442
LAS-Min	314	317	584	118	190	485	239	591	1037	661	454
LAS-1se	416	233	769	157	306	803	339	621	1232	416	529
PLS	260	323	514	138	146	417	266	656	1056	699	447
GAM	108	1103	450	115	272	263	161	452	258	722	390

Question 2

```
## Warning: package 'kableExtra' was built under R version 4.0.3
```

(a) The MSPEs for each fold and for the full data are given in Table 1.

(b)(c) Boxplots of MSPEs and RMSPEs are given in Figures 4 and 5 respectively. We see that GAM performs very well relative to the other models, both on MSPE and RMSPE. (d) The reason that GAM does this well is that it models nonlinear relationships between variables, and we have seen that the air quality data has clear nonlinearity.

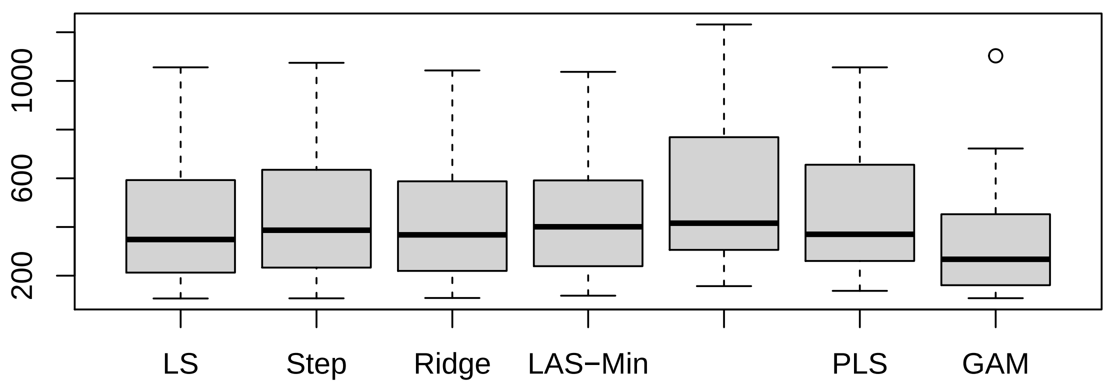

Figure 4: MSPE Boxplots

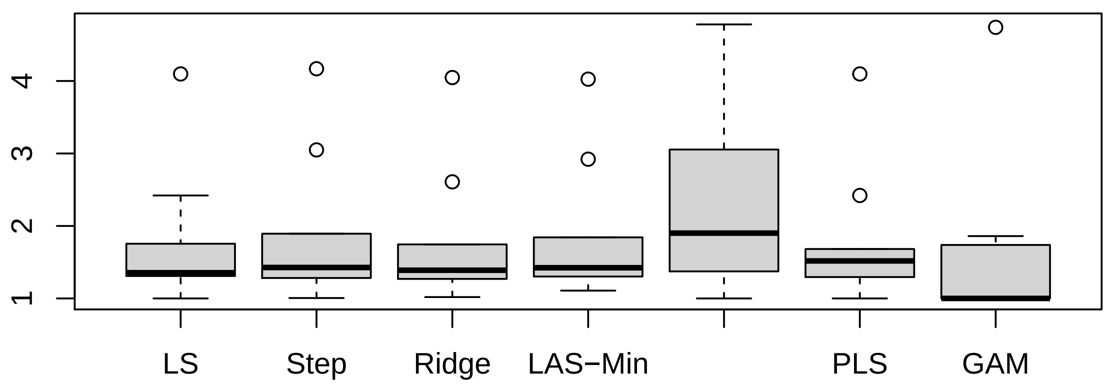

Figure 5: RMSPE Boxplots