

Ejemplo de Modelos GAM en R

Inferencia Estadística - UAO

1. Introducción

Los modelos aditivos generalizados (GAM) permiten capturar relaciones no lineales entre las variables predictoras y la variable respuesta usando funciones suaves (splines).

En este ejemplo se utiliza el dataset `airquality`, que contiene medidas de calidad del aire en Nueva York.

2. Preparación de los datos

```
data("airquality")
airquality <- na.omit(airquality) # Eliminar NA
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 7    23     299  8.6   65     5   7
## 8    19      99 13.8   59     5   8
```

Nos interesa modelar la variable `Ozone` como función de `Temp`, `Wind` y `Solar.R`.

3. Ajuste del modelo GAM

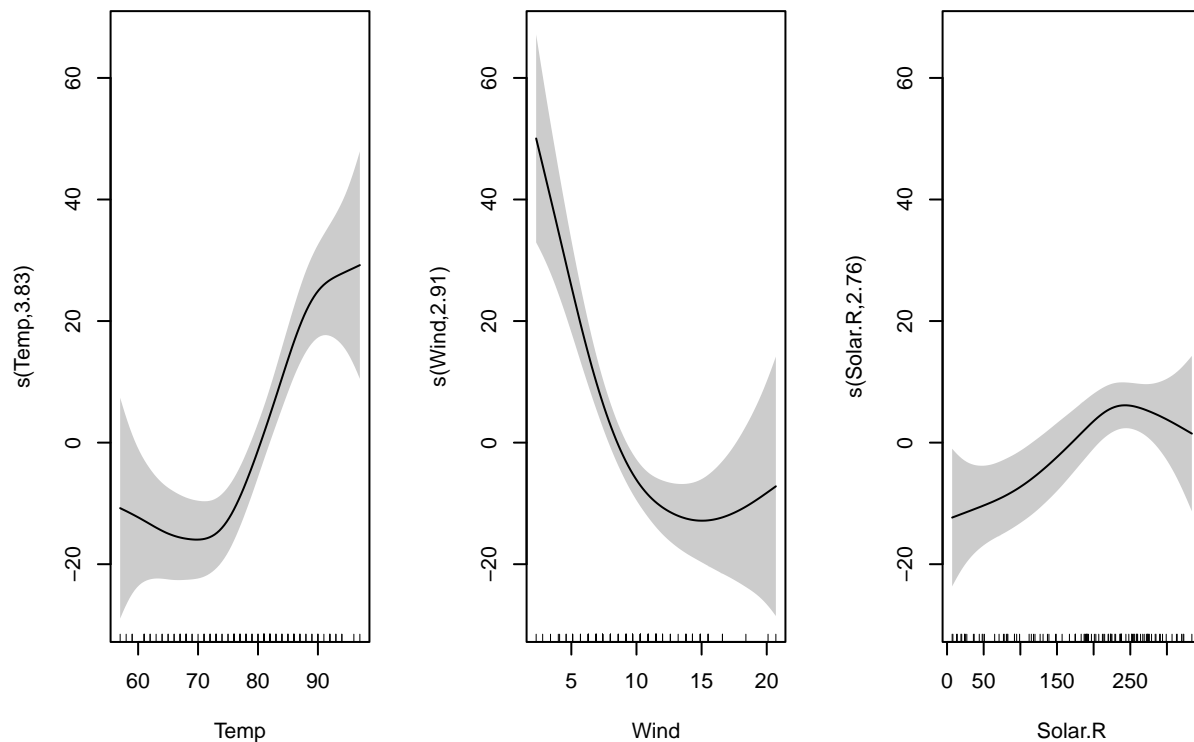
```
# Modelo GAM con funciones suaves para cada predictor
gam_ozone <- gam(Ozone ~ s(Temp) + s(Wind) + s(Solar.R), data = airquality)
summary(gam_ozone)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Ozone ~ s(Temp) + s(Wind) + s(Solar.R)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  42.099      1.663    25.32  <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(Temp)   3.833  4.753 11.613 < 2e-16 ***
## s(Wind)    2.910  3.657 13.695 < 2e-16 ***
## s(Solar.R) 2.760  3.447  3.967 0.00858 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.723   Deviance explained = 74.7%
## GCV = 338.9   Scale est. = 306.83    n = 111
```

4. Visualización de los efectos

```
par(mfrow = c(1, 3))
plot(gam_ozone, se = TRUE, shade = TRUE)
```



Cada gráfico muestra cómo varía Ozone con cada predictor, manteniendo los demás constantes.

5. Comparación con un modelo lineal

```
lm_ozone <- lm(Ozone ~ Temp + Wind + Solar.R, data = airquality)
AIC(gam_ozone, lm_ozone)
```

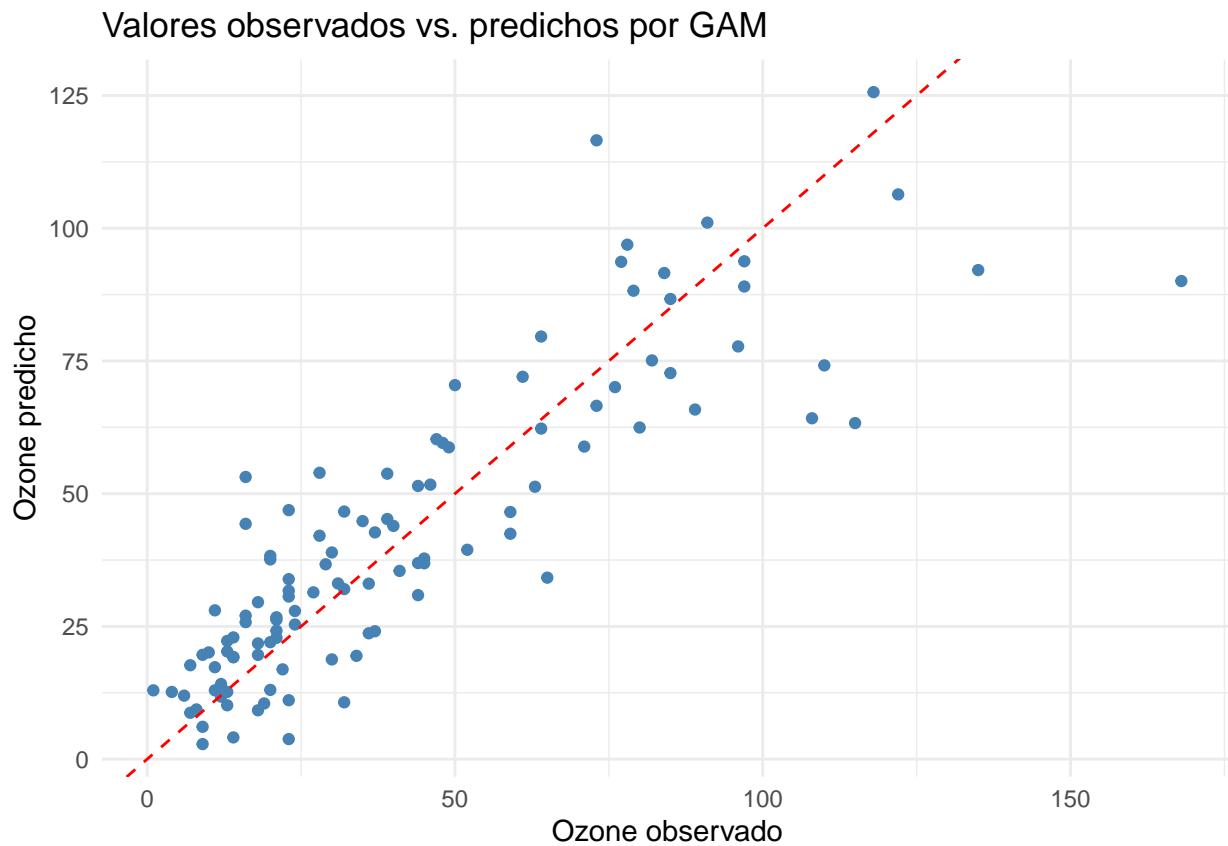
```
##           df      AIC
## gam_ozone 11.50254 962.5960
## lm_ozone   5.00000 998.7171
```

Interpretación: un AIC más bajo indica mejor ajuste. Si el modelo GAM tiene un AIC inferior, mejora la predicción al capturar no linealidad.

6. Predicción y visualización

```
airquality$pred_gam <- predict(gam_ozone)

ggplot(airquality, aes(x = Ozone, y = pred_gam)) +
  geom_point(color = "steelblue") +
  geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Valores observados vs. predichos por GAM",
       x = "Ozone observado", y = "Ozone predicho") +
  theme_minimal()
```



7. Conclusión

- Los modelos GAM permiten mayor flexibilidad frente a modelos lineales clásicos.
- Son útiles cuando se sospechan relaciones no lineales entre las variables.
- En este caso, Ozone muestra una dependencia no lineal con Temp, Wind y Solar.R, bien capturada por el GAM.