Modelos aditivos generalizados

Machine Learning 1
DSLAB

Máster en Data Science. ETSII.

Móstoles, Madrid

January 15, 2021

Universidad
Rey Juan Carlos

¿Qué son los modelos lineales generalizados?

Universidad Rey Juan Carlos

Un modelo lineal tiene la forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

Al introducir una función de distribución y enlace en la regresión lineal, se han generalizado los modelos lineales, teniendo entonces Modelos Lineales Generalizados (GLM).

$$g(E(y|x)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

donde g es una función link que relaciona a la media con el predictor lineal $g(\mu) = y$.

¿Qué son los modelos aditivos generalizados?

Universidad Rey Juan Carlos

Un modelo aditivo tiene la forma:

$$y = \alpha + f_1(x_1) + f_2(x_2) + ... + f_p(x_p)$$

Al introducir una función de distribución y enlace en los modelos aditivos, se tienen los Modelos Aditivos Generalizados (GAM):

$$g(E(y|x)) = \alpha + s_1(x_1) + s_2(x_2) + ... + s_p(x_p)$$

donde los términos $s_1(x_1), ..., s_p(x_p)$ denotan funciones no paramétricas suaves.

Modelos aditivos generalizados

El marco GAM se basa en un modelo mental atractivo y simple:

- Las relaciones entre los predictores individuales y la variable dependiente siguen patrones suaves que pueden ser lineales o no lineales.
- Se pueden estimar estas relaciones suaves de forma simultánea y luego predecir $g(E(Y))$ simplemente sumándolas.

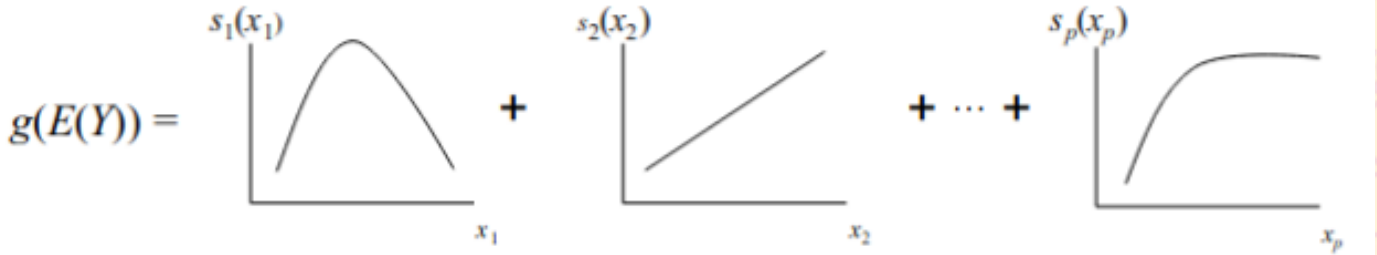

Figure: Definición de GAM.

Objetivos

Universidad Rey Juan Carlos

- Interpretabilidad: Cuando un modelo de regresión es aditivo, la interpretación del impacto marginal de una sola variable (la derivada parcial) no depende de los valores de las otras variables en el modelo. Por lo tanto, al observar simplemente el resultado del modelo, se pueden realizar afirmaciones simples sobre los efectos de las variables predictivas que tienen sentido para una persona no técnica.
- Automatización y Flexibilidad: GAM puede capturar patrones comunes no lineales que un modelo lineal clásico podría perder.
- Regularización: Al controlar la ondulación de las funciones de predicción, se puede abordar directamente el intercambio de sesgo/varianza.

Universidad Rey Juan Carlos

Los *smoothers* son las piedras angulares de GAM. A alto nivel, hay dos clases de *smoothers* utilizados para GAM:

- Suavizado de *splines*: Estima la función suave minimizando la suma de cuadrados penalizada

$$RSS = \sum_{i=1}^{n} (y_i - s(x_i))^2 + \lambda \int (s''(x))^2 dx$$

  donde la suma residual de los cuadrados garantiza que se ajustan los datos observados, mientras que el término de penalización impone la suavidad (es decir, penaliza la ondulación). La compensación entre el ajuste del modelo y la suavidad se controla mediante el parámetro de regularización, λ.
- Regresión de *splines* (*B-splines*, *P-splines*, *thin plate splines*): Ofrecen una alternativa más práctica para suavizar las *splines*. La principal ventaja es que pueden expresarse como una combinación lineal de un conjunto finito de funciones básicas que no dependen de la variable Y dependiente. Se puede escribir una *spline* de regresión de orden q como $s(x) = \sum_{l=1}^{K} B_{l,q}(x)\beta_l = B'\beta$.

En general un *B-splines* de grado p:

- Consiste en $p+1$ trozos de polinomio de orden p.
- Se unen en p nodos internos
- En los puntos de unión las derivadas hasta el orden $p1$ son continuas.
- El *B-splines* es positivo en el dominio expandido por $p+2$ nodos y 0 en es resto.
- Excepto en los extremos, se solapa con $2p$ trozos de polinomios de sus vecinos.
- Para cada valor de x, $p+1$ *B-splines* son no nulos.

Gráfica con *B-splines*

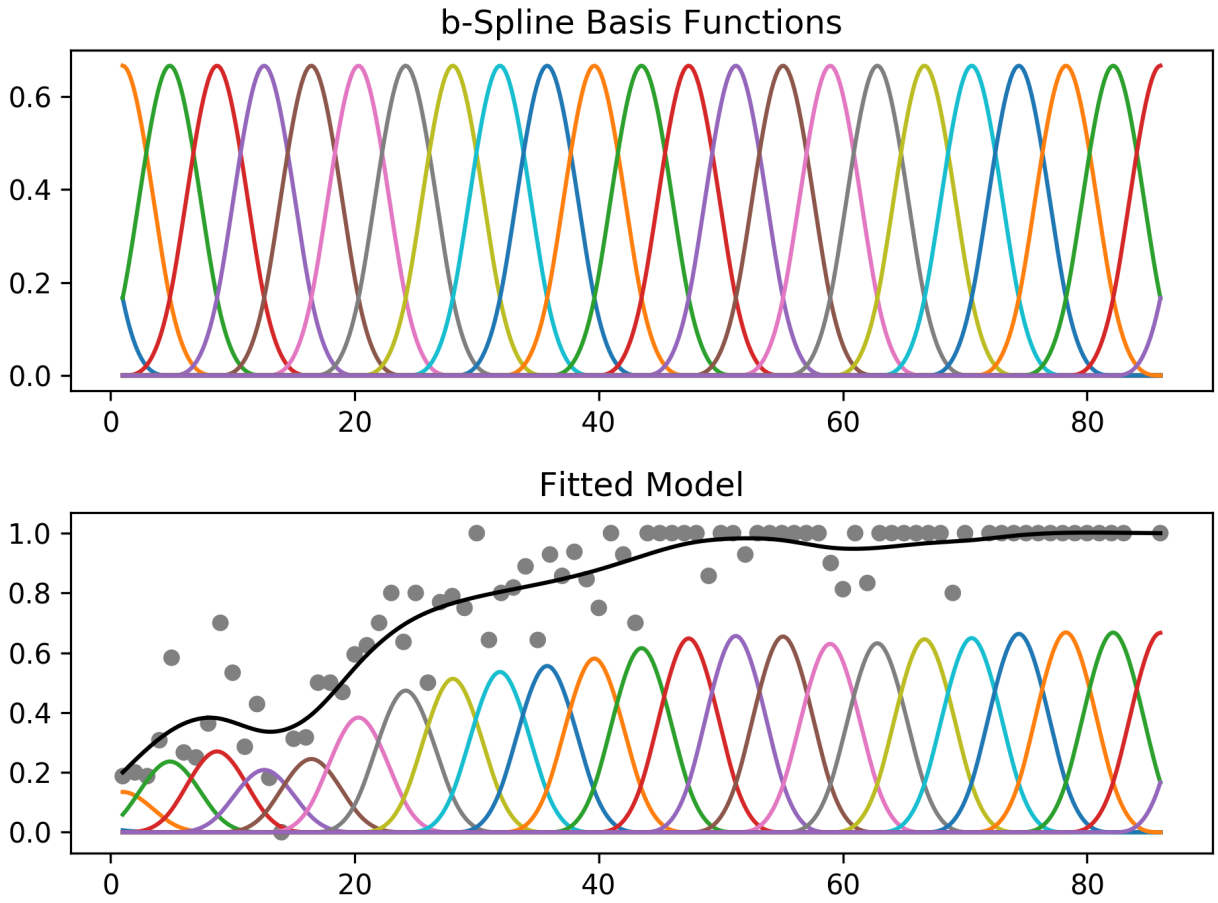

Figure: Efectos de cada variable independiente sobre el modelo.

La novedad que introducen los *P-splines* es que la penalización es discreta y que se penalizan los coeficientes directamente, en vez de penalizar la curva, lo que reduce la dimensionalidad del problema.
En el caso de datos normalmente distribuidos tenemos el modelo de regresión $y = Ba + \epsilon$, donde $\epsilon N(0, \sigma 2I)$, y $B = B(X)$ es la base splines de la regresión obtenidos a partir de X. Los coeficientes se estiman mediante mínimos cuadrados penalizados:

$$S(a; y, \lambda) = (yBa)^t(yBa) + \lambda a^t Pa$$

Además, los *P-splines* ajustan de forma exacta los polinomios, es decir, si la curva es polinomica, un *P-splines* la recuperara exactamente.

B-*splines* vs. P-*splines*

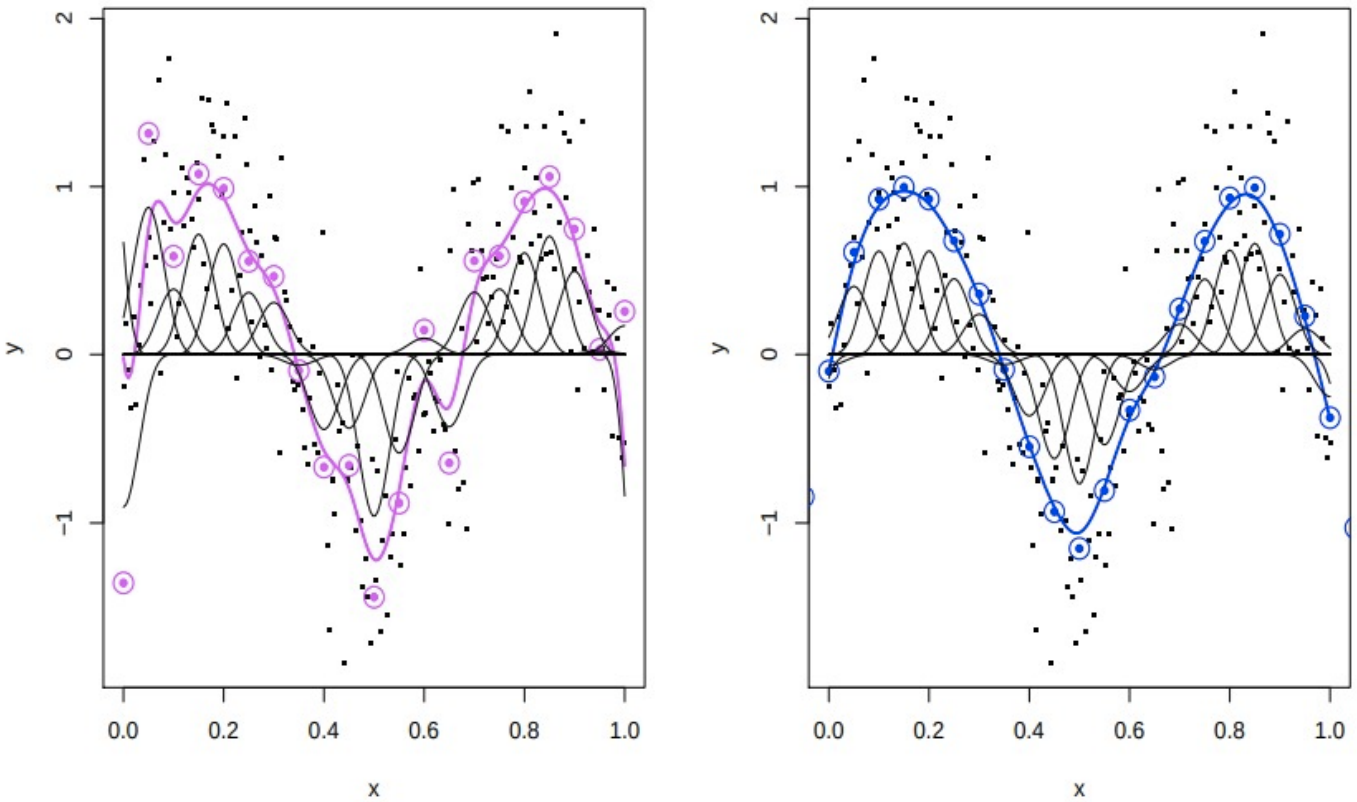

Figure: Curva estimada con 20 nodos, sin penalizar los coeficientes (izquierda) y penalizando los coeficientes (derecha).

Estimación de GAMs: *smoothers*

Universidad Rey Juan Carlos

Al estimar los GAM, el objetivo es estimar simultáneamente todos los *smoothers*, junto con los términos paramétricos (si los hay) en el modelo, mientras se tiene en cuenta la covarianza entre los *smoothers*. Existen dos vías para realizar esta estimación:

- Algoritmo de puntuación local: Extensión del algoritmo de *backfitting*, que se basa en el proceso de Gauss-Seidel para resolver sistemas lineales.
- *Penalized Re-weighted Iterative Least Square* (PIRLS): Generalización del IRLS usado para la resolución de GLMs.

En general, el algoritmo de puntuación local es más flexible en el sentido de que puede usar cualquier tipo de *smoothers* en el modelo, mientras que el enfoque GLM sólo funciona para regresión de *splines*. Sin embargo, el algoritmo de puntuación local es computacionalmente más costoso y no se presta tan bien a la selección automatizada de parámetros de regularización como el enfoque GLM.

Estimación de GAMs: regularización

Universidad Rey Juan Carlos

Cuando se ajusta un GAM, la elección de los parámetros de regularización, es decir, los parámetros que controlan la suavidad de las funciones predictivas, es clave para el ajuste del modelo. Se pueden preseleccionar los parámetros de regularización o estimar a partir de los datos mediante alguna de las dos siguiente formas:

- Criterios de validación cruzada generalizada (GCV).
- Enfoque de modelo mixto a través de máxima verosimilitud restringida (REML).

REML sólo se aplica si se está lanzando GAM como un GLM grande. En general, el enfoque REML converge más rápido que el GCV y el GCV tiende a ser menos suave.

Parámetros de GAM: Funciones de suavizado

Universidad Rey Juan Carlos

- Términos lineales.
- *Splines* cúbicos.
- Regresión de *splines*: *B-splines* y *P-splines*.
- Términos factor.
- Funciones de suavizado multidimesionales: Tensores y *thin plate splines*.

Parámetros de GAM: Distribuciones

Universidad Rey Juan Carlos

- Normal: $f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Binomial: $f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$
- Poisson: $f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$
- Gamma: $f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)}$
- Gaussiana inversa: $f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} exp[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}]$

Parámetros de GAM: Funciones link

Universidad Rey Juan Carlos

Las funciones de enlace llevan la media de la distribución a la predicción lineal. Se dispone de las siguientes:

- Identity: $f(x) = x$
- Logit: $f(x) = log(\frac{x}{1-x}) = log(x) - log(1-x)$
- Inverse: $f(x) = \frac{1}{x}$
- Log: $f(x) = log(x)$
- Inverse-squared: $f(x) = \frac{1}{x^2}$

Gráficos de dependencias parciales

Universidad Rey Juan Carlos

Este tipo de gráficas muestran el efecto marginal que una o dos características tienen sobre el resultado predicho de un modelo de aprendizaje automático. Un gráfico de dependencia parcial puede mostrar si la relación entre la variable objetivo y una variable de entrada es lineal, monótona o más compleja.

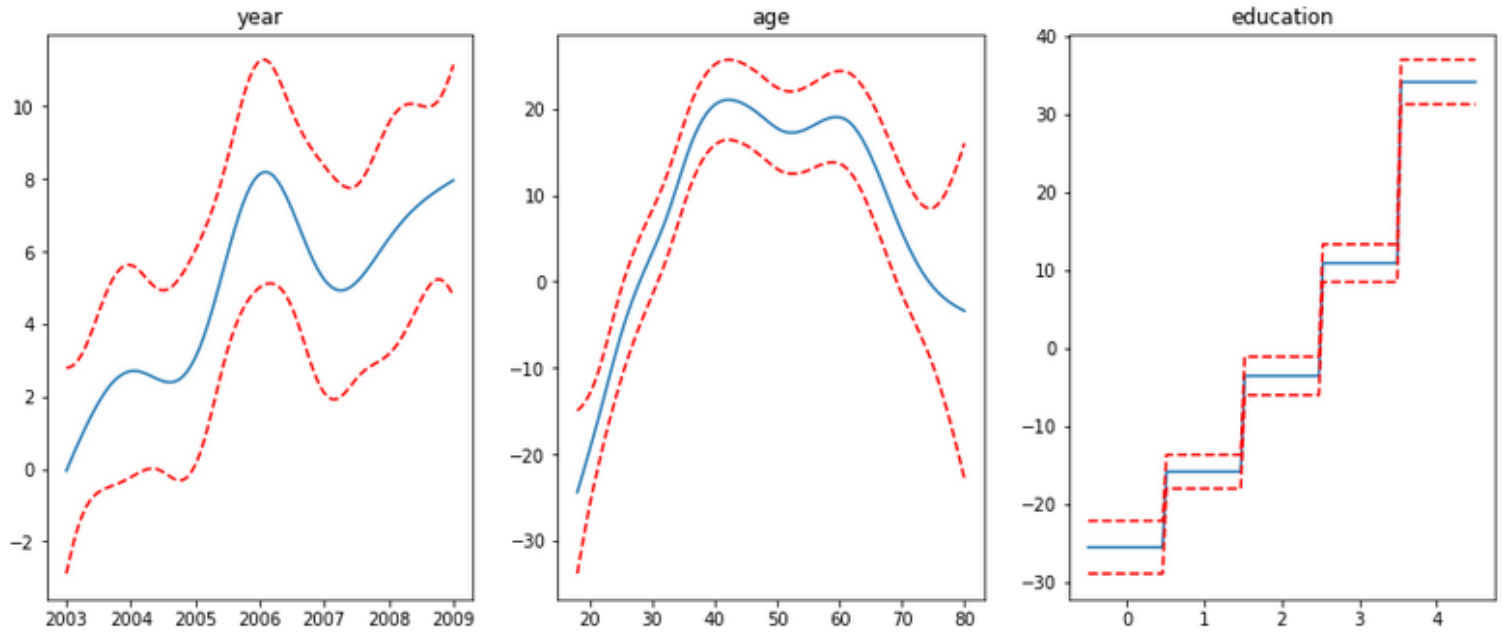

Figure: Ejemplo de gráficos de dependencias parciales.

Documentación: Python y R

Universidad Rey Juan Carlos

R

- *mgcv*: https://cran.r-project.org/web/packages/mgcv/mgcv.pdf

Python

- *PyGAM*: https://pygam.readthedocs.io/en/latest/notebooks/quick_start.html

Artículos

- *GAM: The Predictive Modeling Silver Bullet*: https://pdfs.semanticscholar.org/dea4/adaaf06e6fc99179a2620b7a031188c6e532.pdf