Regresión spline - Modelos aditivos

Clase 6

Manuel Benjamín

November 18, 2023

Universidad de Buenos Aires

Regresión no lineal

Las siguientes técnicas permiten ajustar funciones no lineal utilizando las herramientas del modelo lineal

- Regresión polinomial.
- Regresión por funciones escalón.
- Regresión splines.
- Regresión por splines suaves.
- Regresión local.
- Modelos aditivos (generalizados)

```
> summary(Wage)
      year           age                      maritl              race                   education                      region
 Min.   :2003   Min.   :18.00   1. Never Married: 648   1. White:2480   1. < HS Grad       :268   2. Middle Atlantic   :3000
 1st Qu.:2004   1st Qu.:33.75   2. Married      :2074   2. Black: 293   2. HS Grad         :971   1. New England       :   0
 Median :2006   Median :42.00   3. Widowed      :  19   3. Asian: 190   3. Some College    :650   3. East North Central:   0
 Mean   :2006   Mean   :42.41   4. Divorced     : 204   4. Other:  37   4. College Grad    :685   4. West North Central:   0
 3rd Qu.:2008   3rd Qu.:51.00   5. Separated    :  55                   5. Advanced Degree:426    5. South Atlantic    :   0
 Max.   :2009   Max.   :80.00                                                                     6. East South Central:   0
                                                                                                  (Other)              :   0
           jobclass               health      health_ins      logwage           wage
 1. Industrial :1544   1. <=Good     : 858   1. Yes:2083   Min.   :3.000   Min.   : 20.09
 2. Information:1456   2. >=Very Good:2142   2. No : 917   1st Qu.:4.447   1st Qu.: 85.38
                                                           Median :4.653   Median :104.92
                                                           Mean   :4.654   Mean   :111.70
                                                           3rd Qu.:4.857   3rd Qu.:128.68
                                                           Max.   :5.763   Max.   :318.34
```

Figure 1: Conjunto de datos *Wage*. Salarios de personas y variables explicativas.

Regresión polinomial

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots \beta_d X^d + \varepsilon.$$

El modelo sigue siendo lineal en los parámetros por lo que la metodología es por minimización del RSS.

Para un grado suficientemente alto el modelo es capaz de producir funciones fuertemente no lineales.

Regresión polinomial

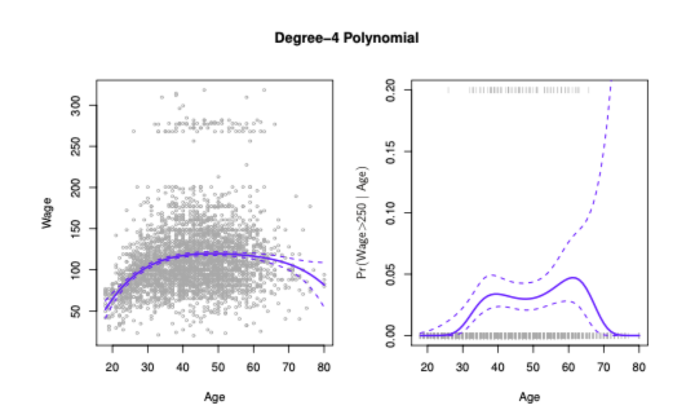

Figure 2: Regresión lineal y logística con un polinomio de grado 4. Las lineas solidas son los valores predichos y las punteadas intervalos construidos con dos desvíos estandard. Se puede ver que en los borde d los intervalos la variabilidad aumenta.

Regresión por funciones escalón.

Partimos el dominio en $K+1$ puntos o nodos $c_0 < c_1 < \ldots < c_K$ que forman intervalos disjuntos

$$\begin{aligned}
C_0(X) &= I(X < c_0) \\
C_1(X) &= I(c_0 \leq X < c_1) \\
C_2(X) &= I(c_1 \leq X < c_2) \\
&\vdots \qquad \vdots \\
C_{K-1}(X) &= I(c_{K-1} \leq X < c_K) \\
C_K(X) &= I(c_K \leq X)
\end{aligned}$$

Los puntos de corte se eligen de antemano. Usualmente se toman nodos equidistantes o que contengan una cantidad parecida de observaciones.

Regresión por funciones escalón

$$Y = \beta_0 + \beta_1 I_1(X) + \beta_2 I_2(X) + \ldots \beta_d I_K(X) + \varepsilon.$$

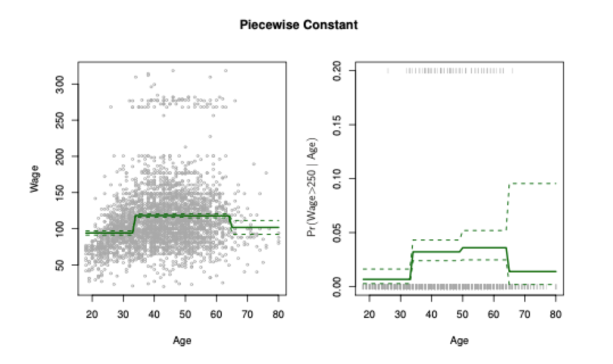

Figure 3: Ajuste lineal y regresión logística con 4 escalones ($K = 3$).

Regresión por funciones escalón

$$Y = \beta_0 + \beta_1 I_1(X) + \beta_2 I_2(X) + \dots \beta_d I_K(X) + \varepsilon.$$

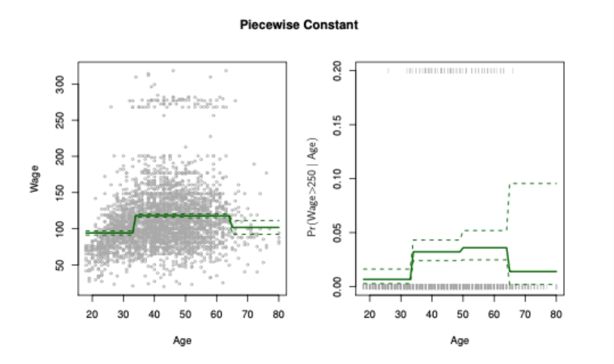

Figure 3: Ajuste lineal y regresión logística con 4 escalones ($K = 3$).

¿Por qué no incluimos C_0 en la fórmula?

Regresión en espacios finito dimensionales - Bases de funciones

Uno puede elegir un espacio de funciones de dimensión d y $b_1(X), \ldots, b_d(X)$ una base de funciones que general el espacio.

$$Y = \beta_0 + \beta_1 b_1(X) + \ldots + \beta_d(X) + \varepsilon$$

Las funciones $b_1(), \ldots, b_d()$ son fijas y conocidas de antemano.

Los ejemplos anteriores son casos particulares

- Regresión polinomial $b_j(X) = X^j$.
- Regresión por escalones $b_j(X) = I_j(c_{j-1} < X \leq c_j)$.

Observaciones

- Hallar los β que minimizan el RSS equivale a hallar la función del espacio que minimiza la suma de los residuos cuadrados.

$$\arg\min_{g \in F} \sum_{i=1}^{n} (Y_i + g(X_i))^2$$

  Donde

$$F = \langle 1, b_1, \ldots, b_d \rangle$$

- No existe una única base posible. En terminos de predicción todas dan lo mismo.
- Las diferencia entre una base u otra para el espacio puede importar por temas computacionales.

Polinomios a trozos

Elegimos un grado de polinomio d y nodos $c_0 < c_1, \ldots, c_K$ dando lugar a un espacio de polinomios a trozos.

Observación

En regresión por escalones es un polinomio a trozo con $d = 0$

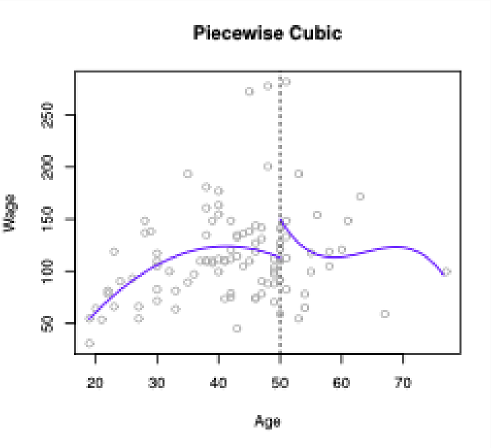

Figure 4: Ajuste polinomio grado tres a trozos con un único nodo. Se observa una discontinuidad en el nodo

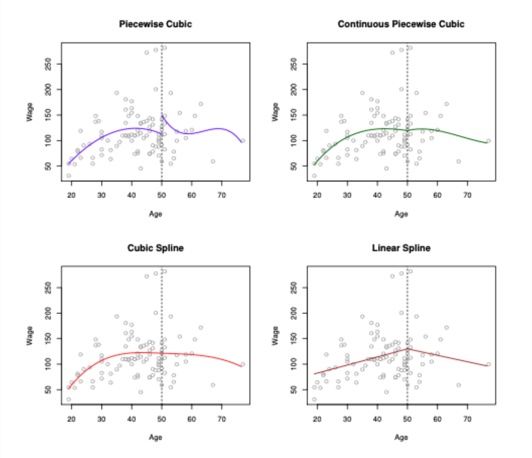

Figure 5: Ajustes por espacios de splines que imponen distintos grados de suavidad en el los nodos y distintos grados en cada intervalo

Espacios de splines

Los espacios de splines son polinomios a trozos con restricciones en la continuidad de la función y su derivadas.

Los splines mas comunes:

- Splines cúbicos.
  Polinomios a trozos cúbicos con segunda derivada continua
  La dimensión del espacio con $K + 1$ nodos es es K + 4
- Splines naturales
  Splines cúbicos con una restricción adicional en los nodos exteriores (derivada segunda nula)
  La dimensión del espacio con $K + 1$ nodos es K + 2

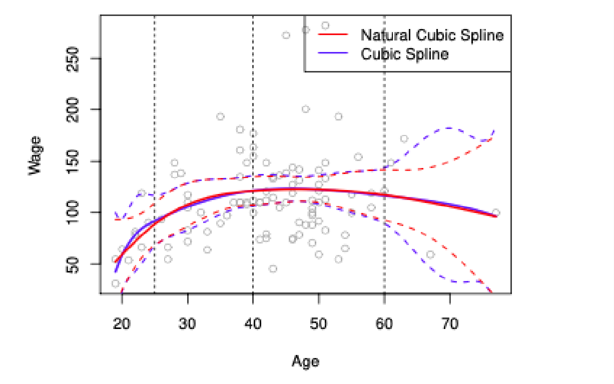

Figure 6: Spline natural y cúbico para los datos de *Wage*.
La restricción de la derivada segunda en los bordes disminuye la variabilidad.

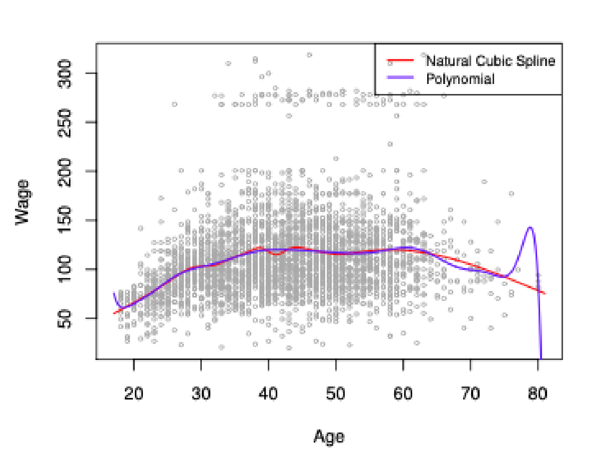

Figure 7: Comparación para splines natural con 15 grados de libertad y un polinomio de grado 15. Se observa un comportamiento errático en el extremo derecho para el spline cúbico.

Elección de la cantidad y la locación de los nodos

- Los regiones con mayor concentración de nodos permiten mas flexibilidad a la función.
- Un uso excesivo de nodos implicaría una interpolación perfecta ($RSS = 0$) incrementando la variabilidad.
- Poca cantidad de nodos lleva a un mayor sesgo.
- Validación cruzada es posible para la cantidad de nodos pero la ubicación se complica.
- En general se utilizan nodos equidistantes en el dominio o de manera que agarren una proporción de datos parecida en cada intervalo..

Splines de suavizado

Idealmente uno quiere encontrar una función suave $g(X)$ de manera que ajuste bien los datos. El problema es que el espacio de funciones suaves (2 derivada continuas) es muy grande (infinito dimensional) y siempre (casi) podemos encontrar una función que interpole interpole todos los datos. Necesitamos penalizar la suavidad de la función

Perdida penalizada

$$\sum_{i=1}^{n}(Y_i - g(X_i))^2 + \lambda \int (g''(t))^2 \, dt.$$

Perdida penalizada

$$\sum_{i=1}^{n}(Y_i - g(X_i))^2 + \lambda \int (g''(t))^2 \, dt.$$

1. Hallar g entre todas las funciones con segunda derivada continua es un problema de optimización sobre un espacio de dimensión infinita.

Perdida penalizada

$$\sum_{i=1}^{n}(Y_i - g(X_i))^2 + \lambda \int (g''(t))^2\, dt.$$

1. Hallar g entre todas las funciones con segunda derivada continua es un problema de optimización sobre un espacio de dimensión infinita.
2. **Teoremón**
   La función g que minimiza la pérdida entre todas las funciones con 2 derivada continua es un spline cúbico con nodos en todos los puntos $X_1, \ldots X_n$.
3.

- La derivada segunda es la aceleración de una partícula en movimiento. Mientras menor sea su aceleración su función de posición respecto del tiempo es más suave.
- Si $\lambda = 0$ la función interpola los puntos y la función será poco suave.
- Si $\lambda \to +\infty$ quien será $g(X)$?
- λ es un hiperparámetro que puede ser elegido mediante validación cruzada.
- Existe una relación entre λ y los grados de libertad del ajuste que muchas veces se utiliza para comparar metodologías.

Modelos aditivos - GAMs

Se extiende del modelo lineal multivariado con p variables explicativas

$$Y = \beta_0 + \beta_1 X_1 + \ldots \beta_p X_p + \varepsilon.$$

A un modelo **aditivo**

$$Y = \beta_0 + f_1(X_1) + \ldots f_p(X_p) + \varepsilon.$$

$$\mathtt{years} = \beta_0 + f_1(\mathtt{years}) + f_2(\mathtt{age}) + f_3(\mathtt{education})$$

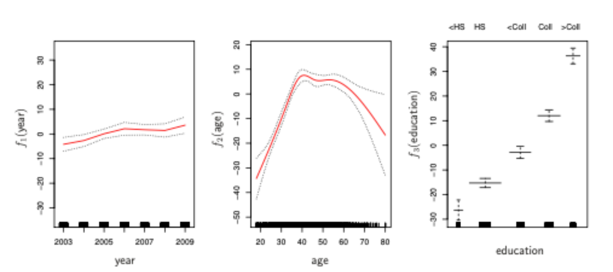

Figure 8: Modelo ajustado para los datos de *Wage* estimando el efecto de *year*, *Age* y *education*. Las primeras dos se ajustaron con un splines naturales de 4 y 5 grados de libertad. Educación se uso una base de escalones.

Interpretación

- Manteniendo fijo años y educación, el salario aumenta levemente con los años. Esto puede mostrar el efecto de la inflación.
- Fijo el año y la educación los mejores salarios se obtienen entre los 35 y 60 años.
- Fijo el año y la edad, a mayor educación mayor salario.

Pros y contras de GAMs

- Encuentra relaciones no lineales de forma automática sin tener que proponer uno transformaciones del as variables explicativas.
- Pueden incrementar el poder predictivo respecto del modelo lineal.
- Gracias a que el modelo es aditivo podemos examinar el efecto de cada variable explicativa sobre la respuesta.
- Se pueden parametrizar la suavidad de las funciones a través de los grados de libertad.
- No estamos obligado a usar splines, podemos usar cualquier metodología.
- La aditividad puede ser una restricción ya que pueden existir interacciones entre variables. Una forma de resolverlo es incluyendo $x_1 * x_2$ como covariable o incluso $f_{ij}(x_i, x_j)$.

Modelos aditivos generalizados

Los modelos lineales generalizados se pueden extender a modelos aditivos generalizados.

Regresión logística aditiva

$$\log\left(\frac{p(Y=1|X_1,\ldots,X_p)}{1-p(Y=1|X_1,\ldots,X_p)}\right) = \beta_0 + f_1(X_1) + \ldots f_p(X_p).$$

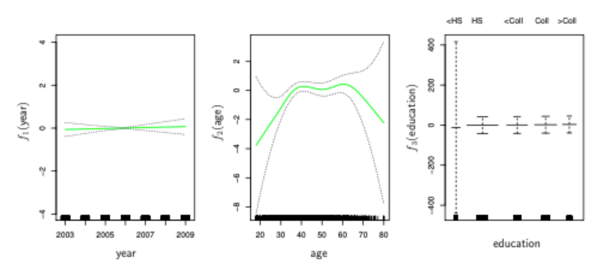

Figure 9: Regresión logística aditiva