# Modelos Aditivos Generalizados

# Outline

# Regresión Polinomial I

- Permite estimar una curva no lineal usando un predictor lineal de la forma:

$$\eta_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_p X_i^p$$

- Impone una estructura global para la forma no lineal de la función de X.

# Regresión por Segmentos I

- El rango de X es dividido por los puntos de corte: $c_1, \ldots, c_\kappa$

$$
\begin{aligned}
C_0\left(X\right) &= I\left(X < c_1\right), \\
C_1\left(X\right) &= I\left(c_1 \leq X < c_2\right), \\
C_2\left(X\right) &= I\left(c_2 \leq X < c_3\right), \\
&\vdots \\
C_{\kappa-1}\left(X\right) &= I\left(c_{\kappa-1} \leq X < c_\kappa\right), \\
C_\kappa\left(X\right) &= I\left(c_\kappa \leq X\right),
\end{aligned}
$$

- El predictor lineal queda expresado como:

$$\eta_i = \beta_0 + \beta_1 C_1(X_i) + \beta_2 C_2(X_i) + \ldots + \beta_\kappa C_\kappa(X_i)$$

# Funciones Base I

- El predictor lineal queda expresado como:

$$\eta_i = \beta_0 + \beta_1 b_1(X_i) + \beta_2 b_2(X_i) + ... + \beta_\kappa b_\kappa(X_i)$$

  donde $b_j(.)$ son funciones base fijas y conocidas
- Regresión Polinómica: $b_j(x_i) = x_i^j$
- Regresión por segmentos: $b_j(x_i) = I\left(c_j \leq X_i < c_{j+1}\right)$
- Otras alternativas para $b_j(.)$: Ondeletas (*wavelets*), Series de Fourier, Splines, etc.

# Regresión por Splines I

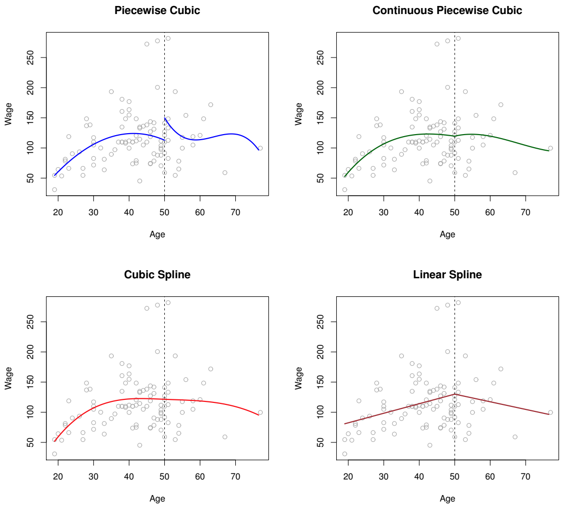
(a) **Polinomios por segmentos:**

$$\eta_i = \begin{cases} \beta_{01} + \beta_{11}X_i + \beta_{21}X_i^2 + \ldots + \beta_{d1}X_i^d & , \text{si } X_i < c_1 \\ \beta_{02} + \beta_{12}X_i + \beta_{22}X_i^2 + \ldots + \beta_{d2}X_i^d & , \text{si } c_1 \leq X_i < \\ \vdots & \vdots \\ \beta_{01} + \beta_{1\kappa+1}X_i + \beta_{2\kappa+1}X_i^2 + \ldots + \beta_{d\kappa+1}X_i^d & , \text{si } X_i \geq c_\kappa \end{cases}$$

donde $c_j$ son "nudos" (*knots*)

- Para $\kappa$ nudos, se tienen $\kappa + 1$ polinomios de grado $d$: gl $= (d+1)(\kappa+1)$
- Produce funciones discontinuas.

# Regresión por Splines II



# Regresión por Splines I

(b) **Spline polinómico:**
Una función $f : [a, b] \to \mathbb{R}$ es un spline polinómico de grado $d \geq 0$ con nudos $a = \xi_0 < \xi_1 < \ldots < \xi_\kappa < \xi_{\kappa+1} = b$ si cumple con las siguientes condiciones:

- $f(x)$ es $(d-1)$ veces continuamente diferenciable. (Excepto $d = 1$ solo debe asegurar continuidad)
- $f(x)$ es un polinomio de grado $d$ sobre el intervalo $[\xi_j, \xi_{j+1})$ definido por los nudos.

(b) **Splines polinómicos y series de potencia truncadas:**

$$\eta_i = \beta_o + \beta_1 X_i + \ldots + \beta_d X_i^p + \beta_{d+1} h\left(X_i, \xi_1\right) + \ldots + \beta_{d+\kappa} h\left(X_i, \xi_\kappa\right)$$

con $gl = (d+1)(\kappa+1) - d * \kappa = (d+1) + \kappa$, donde:

$$h\left(X, \xi_j\right) = \left(X - \xi_j\right)_+^d = \begin{cases} \left(X - \xi_j\right)^d & , \text{si } X \geq \xi_j \\ 0 & , \text{de otro modo} \end{cases}$$

# Regresión por Splines II

(b) **Splines naturales:**

- Adiciona restricciones de modo que en la función sea lineal más allá de los nodos de frontera

# Splines suavizados I

- Considerando un problema de regresión, asumamos que se se desea estimar la media por una función $g(x)$.
- Como es usual se debería minimizar la suma de cuadrados de los errores (SCE)

$$\sum_{i=1}^{n}(y_i - g(x))^2$$

sin embargo esto puede llevar a un sobreajuste.
- Lo que en realidad se desea es encontrar una función $g(x)$ que haga la SSE pequeña pero que sea suave.
- Una forma de de conseguir que $g$ sea suave es encontrar $g$ que minimize el siguiente criterio

$$\sum_{i=1}^{n}(y_i - g(x))^2 + \lambda \int g''(t)^2 dt$$

# Splines suavizados II

- Se puede demostrar que la función que minimiza esta expresión es un spline cúbico natural con nodos en las mismas observaciones.
- Por lo tanto $g$ se puede escribir como

$$g(x) = \sum N_j(x)\theta_j$$

donde $N_j(x)$ son las bases del spline cúbico natural.
- El criterio anterior se reduce a

$$(y - N\theta)^T(y - N\theta) + \lambda\theta^T\Sigma\theta$$

cuya solución es dada por

$$\widehat{\theta} = (N^TN + \lambda\Sigma)^{-1}N^Ty$$

# Splines suavizados III

- El valor ajustado por este modelo será dado por

$$\begin{aligned}\widehat{g} &= N(N^T N + \lambda\Sigma)^{-1} N^T y \\ &= S_\lambda y\end{aligned}$$

- En analogía con el modelo lineal usual $tr(S_\lambda)$ se le denomina los grados de libertad efectivos del spline suavizado.
- El parámetro de suavizado $\lambda$ suele ser reparametrizado en términos de los grados de libertad efectivo.
- Estos grados de libertad efectivos deben ser prefijado o estimados a través de los datos.

# Modelos Aditivos Generalizados I

- Los modelos aditivos generalizados (GAM) presentan un enfoque general para extender el modelo lineal general permitiendo formas funcionales no-paramétricas para cada una de las covariables mientras se mantiene la propiedad de aditividad.
- Al igual que en los MLG, la variable respuesta sigue una distribución perteneciente a la familia exponencial.
- El predictor lineal es expresado de la siguiente manera:

  $$\eta_i = \beta_0 + s_1(X_1) + s_2(X_2) + ... + s_p(X_p)$$

  donde $s_j(.)$ son funciones no-lineales de suavizamiento
- Usualmente para las funciones $s_j(.)$ se modelan a través de splines.

# Modelos Aditivos Generalizados II

- Por ejemplo,en el caso de modelos para clasificación podemos extender el usual modelo logístico

$$Y \sim Bernoulli(\pi)$$

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$$

- Considerando reemplazar cada componente lineal $(\beta_j X_j)$ con una función no lineal $s_j(X_j)$,

$$Y \sim Bernoulli(\pi)$$

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + s_1(X_1) + s_2(X_2) + ... + s_p(X_p)$$

- Para estimar estos modelos se suele minimizar la verosimilitud del modelo con regularización para evitar sobreajustes.

# Ventajas y Desventajas de los modelos GAM I

- Los modelos GAM permiten ajusta una función no lineal para cada covariable $X_j$ de manera que podemos modelar automáticamente las relaciones no lineales que los modelos de regresión usuales no tomarían en cuenta. Esto siginifica que no necesitamos manualmente intentar diversas transformaciones en cada variable individualmente.
- Los ajuste no lineales pueden pontencialmente llevarnos a mejores predicciones de la respuesta $Y$.
- Debido a que el modelo es aditivo, todavia podemos examinar el efecto de cada $X_j$ en $Y$ individualmente mientras consideramos las otras covariables fijas. Por lo tanto, si estamos interesados en realizar inferencia, los modelos GAM nos dan una representación útil para este fin.

# Ventajas y Desventajas de los modelos GAM II

- La principal limitación de los modelos GAM es la restricción que el modelo es aditivo. Con muchas variables, algunas interacciones importantes pueden ser no consideradas. Sin embargo, como en los modelos de regresión lineales, se pueden adicionar manualmente términos de interacción incluyendo predictores adicionales de la forma $X_j \times X_k$ en el modelo.

# Referencias