

Como breviario histórico, los splines originales, los desarrolla el matemático I. J. Schoenberg como la solución al problema de encontrar la función h en el espacio de Sobolev W_M de funciones con $M - 1$ derivadas continuas y M -ésima derivada integrable al cuadrado que minimice:

$$\int_a^b (h^{(M)}(x))^2 dx,$$

sujeta a que interpole los puntos $h(x_i) = h_i \quad i = 1, 2, \dots, n$ (**schoenberg1964spline**).

Posteriormente, la teoría sobre los splines se expandió y estos fueron adoptados por ramas de la matemática tan diversas como los gráficos por computadora y, como es el caso, la estadística computacional. Bajo este contexto, los splines también surgen de forma orgánica pues el problema general de encontrar una función que ajuste a (??) se puede plantear como el problema de optimización:

$$\min_h \left[\sum_{i=1}^n (y_i - h(x_i))^2 + \lambda \int_a^b h''(t)^2 dt, \right] \quad (1)$$

para alguna $\lambda > 0$. Para este problema la solución se demuestra que son *splines cúbicos* ($M = 4$) con nodos en cada x_i , **green1994nonparametric**. Cabe mencionar, que esta formulación del problema engloba muchas de técnicas estadísticas interesantes además de conceptos de optimización. El lector reconocerá que el primer término es la *suma de residuales cuadrados (RSS)* y el segundo término del sumando es un caso particular de los métodos de regularización mencionados anteriormente. No es el enfoque de este trabajo entrar en los detalles pues cambios menores en la formu-

lación y diferentes elecciones de λ llevan a modelos que cada uno merece un trabajo extenso por si mismo. Sin embargo, es importante mencionar que la regularización y modelos de este tipo, son algunos de los más usados y útiles en el aprendizaje de máquina, pues logran captar patrones muy complejos al incluir muchos términos de orden superior e interacciones tratando de no sobreajustar los datos, al ir penalizando la complejidad a través del segundo término. Para ilustrar los métodos de regularización se retoma el ejemplo 4 del capítulo ?? (página ??). Esta base de datos se toma del curso en línea de **andrew2018ml** donde, se busca encontrar fronteras de clasificación circulares usando un modelo logístico en \mathbb{R}^2 al incluir todos los términos polinomiales y las interacciones posibles de las dos covariables hasta el orden 6 dando un total de 28 parámetros. Al correr el modelo regularizado, donde se penaliza la complejidad muchos de estos términos se desvanecen y se termina con un modelo más parsimonioso.

Lo esencial en la expresión (1) es la idea del *suavizamiento*, es decir, al tratar de minimizar el RSS se puede caer en problemas de sobreajuste en donde el modelo no se estén capturando efectos y patrones subyacentes, la llamada *señal* de los datos, sino sólo se trata de seguir (incluso interpolar) a los estos, dando resultados *ruidosos* muy dependientes de las observaciones. Para compensar la complejidad, se penaliza la función a minimizar con segundo termino que controla el número de parámetros y la suavidad deseada mediante λ . A este segundo término, se le conoce como termino de *penalización* y crece a medida que h se vuelve más compleja.¹

1. Si el lector tiene una intuición de análisis, notará que integrar la función al cuadrado, corresponde con el producto interno de las funciones pertenecientes al espacio de Hilbert $\mathcal{L}_2([a, b])$, **bergstrom1985estimation**.

Posterior a estas formulaciones, los splines vuelven a ser relevantes con el modelo aditivo de Hastie y Tibshirani. Ellos extienden el concepto de un espacio de covariables en una sola dimensión a $d \in \mathbb{N}$. La formulación del problema es prácticamente la misma que (1) pero ahora se busca estimar d funciones h , dando lugar al mismo número de parámetros λ :

$$y = \sum_{j=0}^d h_j(x_j) + \epsilon$$

$$\text{RSS}(h_0, h_1, \dots, h_d) = \sum_{i=1}^n [y_i - \sum_{j=0}^d h_j(x_{ij})]^2 + \sum_{j=1}^d \lambda_j \int h_j''(t_j) dt_j$$

con la convención de que h_0 es una constante. Ellos muestran que $h_j \quad j = 1, \dots, d$ son splines cúbicos. Sin embargo, sin restricciones adicionales, el modelo no sería *identificable*, es decir h_0 podría ser cualquier real y el modelo sería el mismo. Para asegurar la unicidad de la solución se añade la condición de que las funciones estimadas, promedien cero sobre los datos:

$$\sum_{i=1}^n h_j(x_{ij}) = 0 \quad \forall j \tag{2}$$

Esto lleva a la conclusión natural de que h_0 sea la media de las variables de respuesta, es decir: $h_0 = \bar{y}$. Por lo que si se ve cada dimensión j , se tiene que su función correspondiente h_j está centrada alrededor de la media \bar{y} .