

Práctica 2: Regresión polinómica

Modelos Lineales

Grados en Estadística y Matemáticas

- 1 Regresión Polinómica con una variable regresora
- 2 Regresión Polinómica con varias variables regresoras

Regresión Polinómica con una variable regresora

Queremos determinar si una variable X influye en la distribución de probabilidad de otra variable aleatoria Y .

En principio, el marco adecuado para este estudio es Modelo de Regresión Lineal. Pero a veces la posible relación entre X e Y no es de carácter lineal, sino que viene marcada por otra función $f()$.

- 1 En general es complicado realizar inferencia si no conocemos la expresión funcional de $f()$.
- 2 Es conveniente que $f()$ tenga una expresión funcional conocida, salvo por una serie de parámetros.
- 3 Los polinomios aproximan (uniformemente) cualquier función continua en un intervalo (Teorema de aproximación de Weierstrass).

Por todo ello, el marco adecuado para el estudio es la **Regresión Polinómica**

Regresión Polinómica con una variable regresora

Ejemplo 1

Una compañía observa que la demanda de uno de sus productos cambió debido a una variación rápida de su precio por unidad. Se observa la demanda del producto (unidades) en una región en particular sobre un intervalo bastante amplio de precios (dolares). Los datos que se encuentran en el archivo `company.dat`.

Se trata de realizar un análisis de dichos datos, indicando una ecuación que marque la relación entre el precio del producto por unidad y la demanda del mismo.

Regresión Polinómica con una variable regresora

Modelo de Regresión Polinómica

El modelo que consideraremos es el de Regresión Polinómica, que matemáticamente se formula del siguiente modo:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \cdots + \beta_p X^p + \varepsilon$$

Se trata de un modelo lineal de rango completo, por tanto, es válido todo lo dicho en esta práctica.

El problema aquí es **determinar el grado p del polinomio** que mejor define la posible relación entre X e Y .

La forma de hacerlo es introduciendo los términos $\beta_i X^i$, mediante un método de pasos hacia adelante, hasta que introduzcamos un término que no resulte significativo, es decir, que al hacer el contraste

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

aceptemos la hipótesis nula H_0 .

Regresión Polinómica con una variable regresora

Ejemplo 1 (predicciones)

Para estimar la demanda para un precio del producto de 14 dólares.

```
predict (LinearModel.x, data.frame(dolares=14), interval="confidence")  
predict (LinearModel.x, data.frame(dolares=14), interval="prediction")
```

Ejemplo 2

Las puntuaciones de cierto examen pueden depender de las horas de estudio y las horas de ejercicio previas al examen. Para establecer una ecuación que relacione estas variables, se realiza un experimento con 100 estudiantes, que registran cuántas horas de estudio y ejercicio tuvieron en los días previos al examen. Posteriormente se registró también la puntuación del examen. Los datos se proporcionan en el archivo `examen.dat`.

Modelo de Regresión Lineal

Inicialmente, en este problema, consideraremos el modelo Regresión Lineal con Y : puntuación en el examen, X_1 : Horas de estudio y X_2 : Horas de ejercicio Polinómica:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mathcal{E}$$

Al hacer el diagnóstico de este modelo se aprecia la ausencia de un término cuadrático. El problema es que en modelos con más variables, los términos polinómicos de mayor grado aumentan excesivamente el número de parámetros. En este caso, el término cuadrático está compuesto de tres sumandos, quedando el siguiente modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \mathcal{E}$$

La significación estadística del término cuadrático se determinará con el siguiente contraste:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1 &: \beta_i \neq 0 \text{ para algún } i=1,2,3 \end{aligned}$$