Modelos de Clasificación

Doctorado UCA

Ricardo Pasquini

Julio 2024

Modelos de Clasificación

- Hasta aquí hemos trabajado con variables continuas como variables a explicar/predecir, pero en una gran cantidad de aplicaciones las predicciones son discretas.
- Estos modelos también son conocidos como modelos de clasificación

Modelos de Clasificación Binaria

- Ejemplo: Elección de transporte
- Variable objetivo: 1 si toma transporte público, 0 de otro modo.

$$P(\text{Transporte Publico}) = \beta_0 + \beta_1 \text{Ingreso}_i + \beta_2 \text{Genero}_i + \epsilon_i$$

Modelos de Clasificación Binaria

- En general la variable $y_i \in \{0, 1\}$, representando un resultado "Sí" / "No".
- El objetivo es describir $P(y_i = 1|x_i)$ (Notar que esto determinaría la completa distribución condicional)
- Si proponemos un modelo lineal:

$$P(y_i = 1|x_i) = x_i'\beta$$

  - Notemos también que $P(y_i = 1|x_i) = E(Y_i|x_i)$ por lo tanto podríamos estimar

$$y_i = x'\beta + e_i$$

Modelos de Clasificación Binaria

- Sin embargo, notemos que en un caso de observaciones binarias tendríamos algo como:

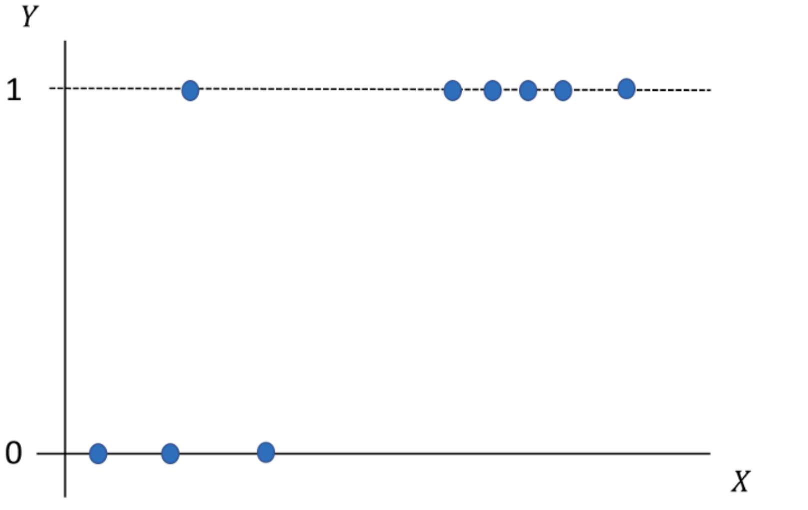


Figure 1: Observaciones Binarias

Modelo Lineal para Clasificación

- El resultado devolvería:

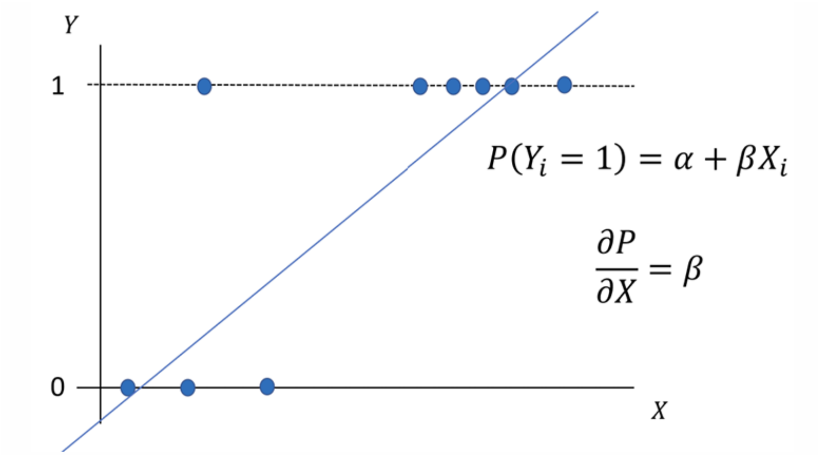


Figure 2: Modelo Lineal

Modelos de Clasificación

- Es deseable imponer una restricción adicional, que $0 \leq P(y_i = 1|x_i) \leq 1$
- Por eso la alternativa estándar es proponer un modelo como:

$$y_i = F(x'\beta) + e_i$$

  donde $F(\cdot)$ es una función que cumple la restricción, como por ejemplo una Función de Densidad Acumulativa.

- Dos funciones comúnmente utilizadas son:
  - Logística: $F(u) = \frac{1}{(1+e^{-u})} = \frac{e^u}{1+e^u}$ (resultando en el modelo llamado **Logit**)
  - Normal: $F(u) = \Phi(u)$ (resultando en el modelo llamado **Probit**)

Ejemplo: Logit

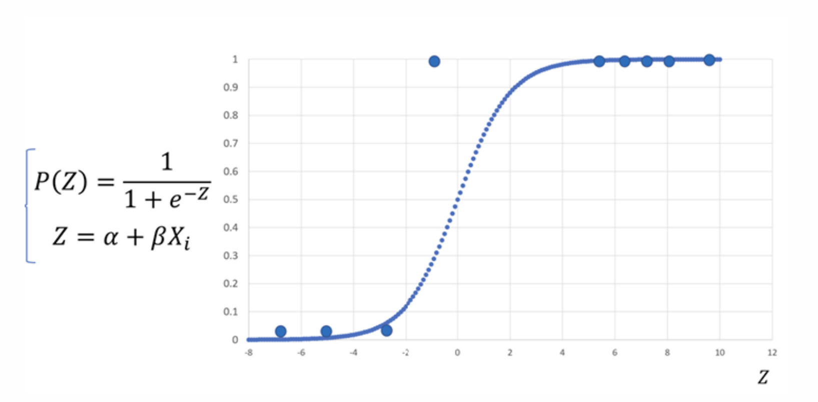


Figure 3: Logit

Logit

- Logit también es el nombre que toma el logaritmo del ratio de chances entre dos resultados.
  - En este caso, tal como fue definido, el modelo logit propone que el logaritmo del ratio de chances es la función lineal $\alpha + \beta X$. Esto se puede ver facilmente:

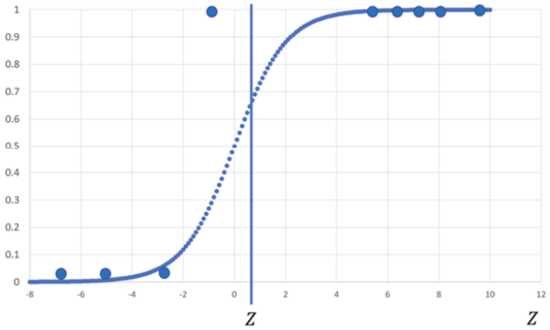
$$\ln \frac{P(y_i = 1|x_i)}{P(y_i = 0|x_i)} = \beta X$$

Efectos marginales

- Comúnmente nos interesará conocer $\frac{\partial p}{\partial x}$.
- Notar que debido a la transformación funcional de $F(\cdot)$ el efecto marginal de un incremento en x_i no es constante. Por lo tanto, para conocer el efecto de un incremento marginal en x adicionalmente tendremos que especificar el nivel (de Z) de interés.

$$\begin{cases} P(Z) = \dfrac{1}{1+e^{-Z}} \\ Z = \alpha + \beta X_i \end{cases}$$

$$\frac{\partial p}{\partial X} = \frac{e^{-Z}}{(1+e^{-Z})^2}\beta$$



Efectos marginales

- Típicamente los paquetes estadísticos devuelven un efecto marginal que :
  - Podría estar calculado por default en el *valor promedio de las variables* (esto podria ser de interés o no).
  - Nos permiten determinar los valores de las variables explicativas donde calcular el efecto.
  - Permiten calcular un *Efecto Marginal Promedio* entendido como el promedio de los efectos marginales en los valores definidos por las observaciones en nuestra muestra.

Estimación

- La estimación se realiza mediante Máxima Verosimilitud.
  - Recordemos que la estimación mediante Máxima Verosimilitud propone buscar el valor del parámetro que maximiza la probabilidad conjunta de ocurrencia de los datos que sugieron en la muestra.

$$\max_{\beta} \prod_{i=1}^{n} P(Y_i, \beta)$$

    donde $p(Y_i, \beta)$ es la probabilidad de la observación i-esima en cuestión.
  - Notemos en este caso las probabilidades son

$$\begin{cases} P(Y = 0|X) \text{ si } Y_i = 0 \\ P(Y = 1|X) \text{ si } Y_i = 1 \end{cases}$$

Estimación

- Es decir los términos de la productoria serán

$$\begin{cases} 1 - F(x_i'\beta) \text{ si } Y_i = 0 \\ F(x_i'\beta) \text{ si } Y_i = 1 \end{cases}$$

- Siguiendo el método de máxima verosimilitud, será conveniente maximizar la función en logaritmos

$$\max_{\beta} L = ln(\prod_{i=1}^{n} P(Y_i, \beta)) = \max_{\beta} \sum ln(P(Y_i, \beta))$$

  - Típicamente se utilizarán algoritmos de maximización para encontrar el máximo.

Multinomial Logit para elecciones múltiples

- Un caso de interés es la elección entre alternativas múltiples. Por ejemplo:
  - Elección medio de transporte (auto, colectivo, tren, etc.).
  - Elección residencial (barrio 1, barrio 2,...,barrio n).
- Analizaremos aquí un modelo con alternativas que no tienen un orden implícito. Para alternativas ordenadas, existen otros modelo disponibles.

Multinomial Logit para elecciones múltiples

- El modelo multinomial se puede fundamentar suponiendo que cada individuo tiene una utilidad por la j-esima opcion dada por

$$U_{ij} = \beta' z_{ij} + \epsilon_{i,j}$$

- Si el consumidor elige la j-esima opción sobre las $J - 1$ restantes, es porque esa opción le dió mayor utilidad. La idea de la modelización probabilística es

$$P(Y_i = j) = P(U_{ij} > U_{ik}) \ \forall k \neq j$$

- Y se puede obtener que la probabilidad de elegir la categoría j-esima es:

$$P(Y_i = j) = \frac{e^{\beta' z_{ij}}}{\sum_{j=1}^{J} e^{\beta' z_{ij}}}$$

Multinomial Logit para elecciones múltiples

- Un supuesto implícito del modelo es la independencia de alternativas irrelevantes.
- Este supuesto pide que la preferencia por una categoría sobre otra no cambie si se considera una tercera como parte de la comparación.
- El uso de este supuesto permite que este modelo pueda implementarse como la comparación de K-1 categorías contra una K-esima. Conociendo estas preferencias puedo derivar todas las demás.

Derivación como un conjunto de elecciones binarias

$$ln \frac{Pr(Y_i = 1)}{Pr(Y_i = K)} = \beta_1 \cdot X_i$$

$$ln \frac{Pr(Y_i = 2)}{Pr(Y_i = K)} = \beta_2 \cdot X_i$$

$$\dots$$

$$ln \frac{Pr(Y_i = K - 1)}{Pr(Y_i = K)} = \beta_{K-1} \cdot X_i$$

Multinomial Logit para elecciones múltiples

$$Pr(Y_i = 1) = Pr(Y_i = K)e^{\beta_1 \cdot X_i}$$

$$Pr(Y_i = 2) = Pr(Y_i = K)e^{\beta_2 \cdot X_i}$$

$$\dots$$

$$Pr(Y_i = K-1) = Pr(Y_i = K)e^{\beta_{K-1} \cdot X_i}$$

$$Pr(Y_i = K) = 1 - Pr(Y_i = K) \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}$$

$$\rightarrow Pr(Y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

Multinomial Logit para elecciones múltiples

Reemplazando para obtener expresiones para $Pr(Y_i = 1)$, $Pr(Y_i = 2)$ ..

$$Pr(Y_i = 1) = \frac{e^{\beta_1 \cdot X_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

$$Pr(Y_i = 2) = \frac{e^{\beta_2 \cdot X_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

$$...$$

$$Pr(Y_i = k) = \frac{e^{\beta_k \cdot X_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

Estimación

- Como en el caso del Logit, se aplicará un método similar a Máxima Verosimilitud para identificar las condiciones del óptimo, y se utilizarán algoritmos numéricos iterativos para encontrar la solución.

Resumen

- Los modelos de clasificación binaria incorporan una transformación no-lineal que garantizar que las predicciones vivan en el rango apropiado.
- Esto implica que los efectos marginales de un factor explicativo serán especificos a un nivel deseado.
- El modelo puede generalizarse

Referencias

- Hansen, B. 2018, Econometrics , ch. 21
- Greene Análisis Econométrico ch.19