

L U M B R E



Regresión logística

Grado: 4

Materia: Regresión y Modelado Estadístico · Probabilidad y Estadística

11/10/2026

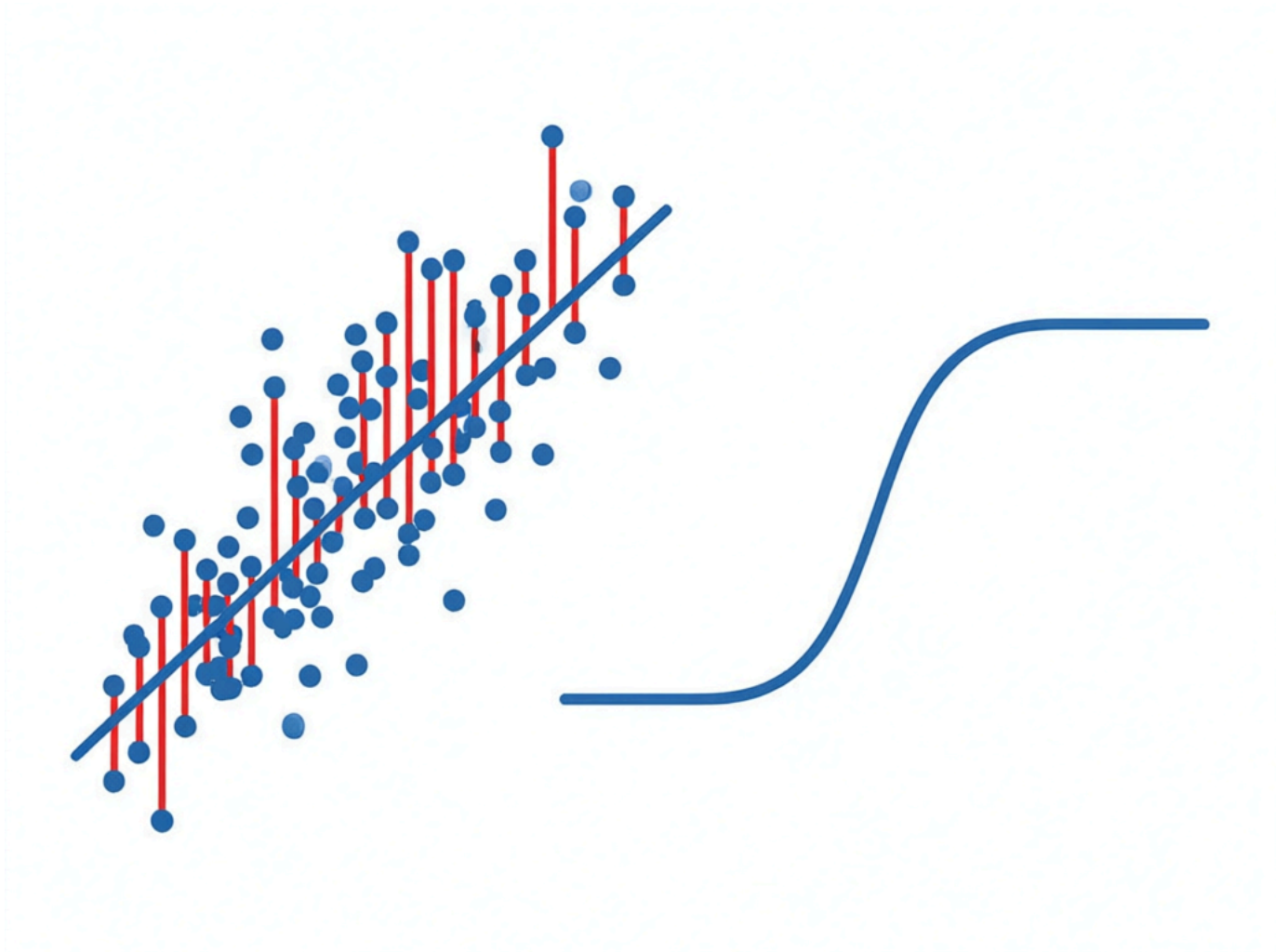
Antes de empezar

Cuando la respuesta no es una cantidad sino un sí o no, la recta queda fuera de lugar: puede predecir probabilidades mayores que uno. La regresión logística resuelve el problema modelando el logaritmo de los odds, y es el caballo de batalla de la clasificación.

QUÉ APRENDERÁS

- Modelar la probabilidad de un resultado binario con la regresión logística, interpretar los odds ratio y evaluar el clasificador con curvas y umbrales.

Regresión logística



Regresión y modelado estadístico: lineal, múltiple y logístico.

Mapa conceptual

- **Regresión**
 - Lineal simple

- Beta 0 y 1
- R-cuadrado
- Residuos
- Supuestos
 - Normalidad
 - Homocedasticidad
 - Independencia
- Multiple
 - Multicolinealidad
 - Ajuste R²
 - AIC/BIC
- Logística
 - Odds ratio
 - Sigmoides
 - Regularización

El mapa mental interactivo es interactivo en la versión online.



Una S que acota la probabilidad

Predecir una probabilidad con una recta se pasa de rango: antes o después predecirás un ciento y veinte por ciento de certeza. La logística dobla la recta en una S que nunca sale de cero a uno, y convierte el modelado de probabilidades en un problema lineal por dentro.

El problema de predecir probabilidades

Queremos la probabilidad de que ocurra un evento dadas unas covariables. Una recta lineal puede dar valores fuera del intervalo válido de cero a uno, y además forzaría una relación lineal entre predictor y probabilidad que casi nunca es realista.

Probabilidad, odds y logit

El odds es la razón entre probabilidad de ocurrencia y de no ocurrencia; vale de cero a infinito. El logaritmo de los odds, el logit, recorre toda la recta real. Modelar el logit linealmente y luego invertir la transformación acota la probabilidad automáticamente.

La función logística

Al invertir el logit, la probabilidad resulta ser una curva sigmoide: una S que se aplanan cerca de cero y de uno y crece deprisa en el medio. Esa forma refleja que cambiar un predictor importa mucho en la zona media y poco en los extremos ya saturados.

Estimación por verosimilitud

La logística no se ajusta por mínimos cuadrados sino maximizando la verosimilitud: busca los coeficientes que hacen más probables las etiquetas observadas. No hay fórmula cerrada; se resuelve iterativamente con métodos numéricos como el descenso de gradiente o Newton.

Interpretar los coeficientes

Un coeficiente actúa sobre el logaritmo de los odds. Al exponentarlo, resulta un odds ratio: el factor por el que se multiplican los odds al subir el predictor una unidad, manteniendo los demás. Un odds ratio mayor que uno aumenta la probabilidad; menor que uno, la reduce.

Por qué odds ratio y no probabilidad

El efecto sobre la probabilidad no es constante: depende de dónde esté uno en la S. En cambio, el efecto sobre los odds es multiplicativo y uniforme, lo que da a los coeficientes una interpretación estable en todo el rango.

La trampa de no colapsar

Con pocas observaciones por predictor, la máxima verosimilitud puede llevar los coeficientes al infinito: separación perfecta, donde alguna combinación de predictores clasifica sin error. El modelo «explota». La regularización o más datos lo remedian.

Regularizar la logística

Penalizar los coeficientes —con Ridge o Lasso— controla el sobreajuste y estabiliza la estimación, especialmente con muchos predictores o separación casi perfecta. La logística penalizada es casi siempre preferible a la ingenua en muestras finitas.

Evaluación del clasificador

Una probabilidad no es una decisión: hay que elegir un umbral para clasificar. Bajarlo captura más positivos pero a costa de más falsos positivos. La matriz de confusión a un umbral fijo esconde este trueque; hay que mirarlo en toda su gama.

Curva ROC y área

La curva ROC dibuja la tasa de verdaderos positivos frente a la de falsos positivos al variar el umbral. Su área resume la capacidad de discriminación del modelo con independencia del umbral: medio es azar, uno es clasificación perfecta.

Precision, exhaustividad y F1

En clases desbalanceadas, la exactitud engaña: predecir siempre «mayoría» acierta mucho. La precisión y la exhaustividad, y su media armónica, exponen el comportamiento sobre la clase minoritaria, que suele ser la que importa.

Calibración frente a discriminación

Un modelo puede ordenar bien los casos (buena discriminación) pero dar probabilidades mal ajustadas (mala calibración). Si se necesitan probabilidades creíbles para decidir costes, hay que calibrarlas, pues la logística las estima en bruto.

Errores frecuentes

- Reportar la exactitud global en un problema con clases muy desbalanceadas.
- Interpretar un coeficiente como cambio de probabilidad cuando es un cambio de odds.
- Fijar el umbral en cero coma cinco por costumbre sin mirar el coste real de cada error.

1. En regresión logística, exponer un coeficiente de un predictor proporciona:

- ☐ El cambio en la probabilidad de la respuesta
- ☐ El odds ratio: el factor que multiplica los odds al subir el predictor una unidad
- ☐ La R cuadrada de ese predictor
- ☐ El número de observaciones

2. La regresión logística puede predecir valores fuera del intervalo de cero a uno.

- ☐ Verdadero
- ☐ Falso

Extensiones multiclase

Para más de dos clases, se generaliza con un modelo de softmax (logística multinomial) que asigna una probabilidad a cada categoría, o con conjuntos de comparaciones de una contra el resto. La intuición de los odds se mantiene clase a clase.

Interpretar en la práctica

Un odds de «1,8 por cada año adicional» debe traducirse a lenguaje de decisión: ¿cuánto sube la probabilidad para alguien con tu perfil base? Los odds distorsionan la intuición cuando la probabilidad no es pequeña; conviene presentar riesgos absolutos.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Logística



4. Ordena el pipeline de un clasificador logístico:

Elementos:

- estima coeficientes por verosimilitud
- usa las probabilidades para ordenar casos
- elige el umbral según costes del error
- evalúa con ROC, precisión y exhaustividad

Ajustar	. . .
Ordenar	. . .
Umbral	. . .
Medir	. . .

Interpretable ML <https://christophm.github.io/interpretable-ml-book/> Interpretación de odds ratio y calibración en modelos logísticos.

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=1wK1K2M7tXI>

5. Un modelo de fraude marca una exhaustividad excelente pero una precisión ínfima. ¿Qué significa ese desequilibrio y cómo elegirías el umbral para producción?

NIST: Regression <https://www.itl.nist.gov/div898/handbook/pmd/section1/pmd16.htm>
Referencia completa de metodos de regresion.

Referencia curricular

Grado:	4
Materia:	Regresión y Modelado Estadístico · Probabilidad y Estadística

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-reg-004-regresion-logistica>

Material educativo generado desde Lumbre · lumbre.lat