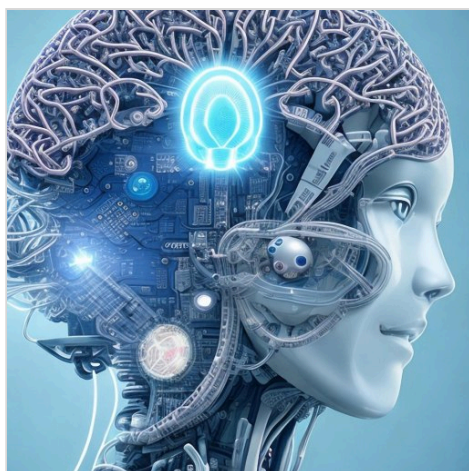


L U M B R E



Transformaciones de variables

Grado: 5

Materia: Feature Engineering · Machine Learning

11/10/2026

Antes de empezar

Raras veces la relación entre causa y efecto es recta y con ruido amable. Transformar variables —comprimir colas, partir en tramos, curvar— acerca los datos a las suposiciones del modelo, y suele valer más que cambiar de algoritmo.

QUÉ APRENDERÁS

- Aplicar transformaciones logarítmicas, de agrupamiento en tramos y polinómicas para revelar estructura que un modelo lineal no capta de entrada.

Transformaciones de variables

Reformular la variable, no solo el modelo

Antes de un modelo más complejo, prueba una variable mejor contada. Una simple raíz cuadrada o un logaritmo pueden desbloquear un patrón que la versión original ocultaba por pura torpeza de escala.

VIDEO

La ingeniería de características: guía completa

Vídeo en español sobre cómo transformar y construir variables para mejorar los modelos.



Escanea para verlo en Lumbre

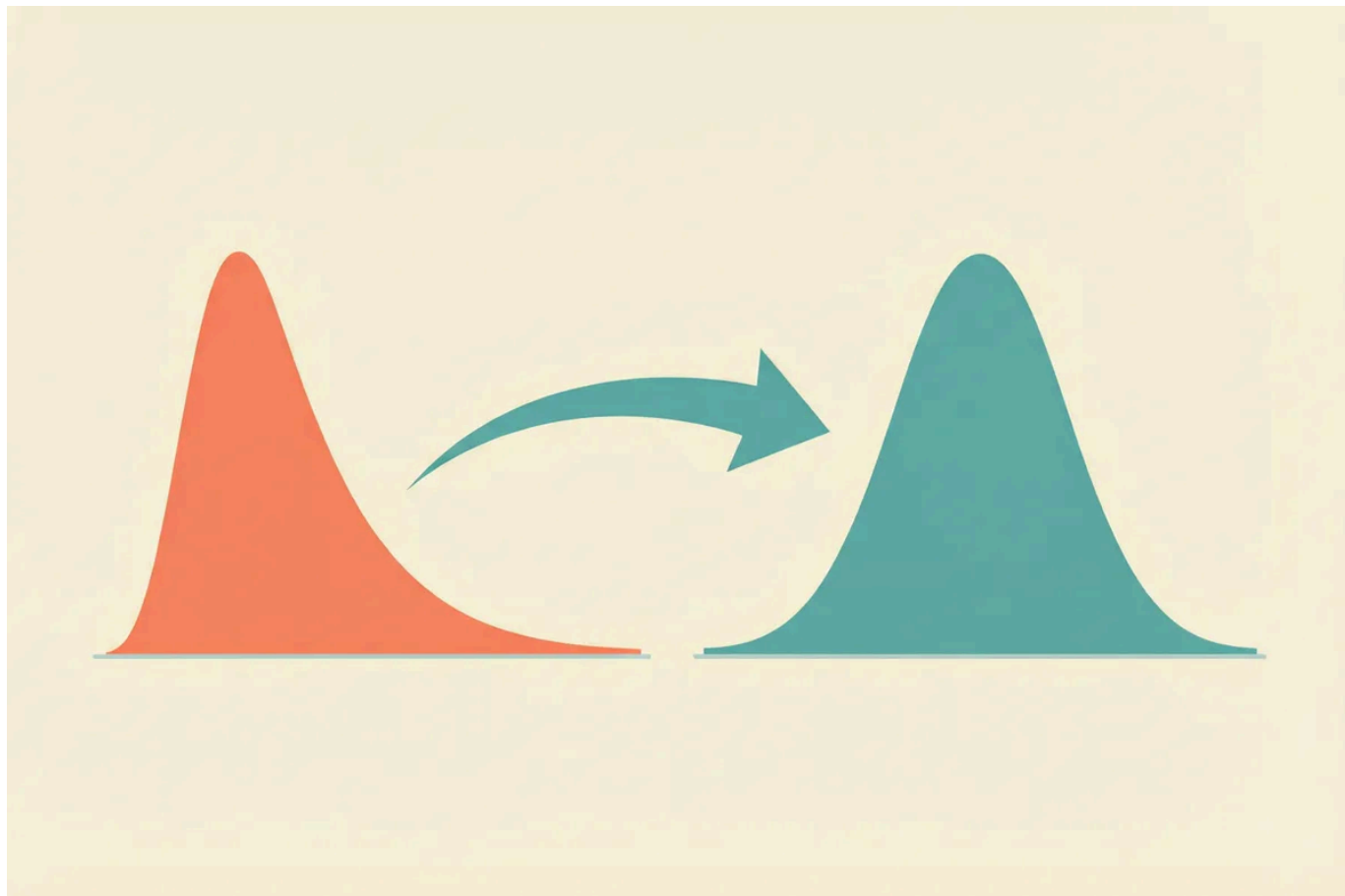
<https://www.youtube.com/watch?v=H7XkA5pSaps>

Para qué transformar

Muchos modelos asumen relaciones lineales, residuos simétricos o escalas comparables. Cuando los datos las violan, el modelo sufre. Transformar acerca los datos a esas suposiciones y realza la señal que estaba comprimida o desbordada.

Logaritmos y colas pesadas

Variables como ingresos o poblaciones se amontonan a la izquierda con una cola larguísima. El logaritmo aplasta la cola, reparte el espacio y convierte crecimientos multiplicativos en aditivos, haciendo el patrón legible para modelos lineales.



Una transformación como el logaritmo toma una variable con cola pesada y la reordena cerca de la simetría, acercándola a lo que suponen muchos modelos.

Tramos y discretización

Partir una continua en tramos —edades en franjas, temperatura en intervalos— captura umbrales y relaciones no monótonas que una recta jamás dirá. Paga con pérdida de detalle fino y elección de cortes, que conviene hagan con criterio.

Polinomios y curvatura

Añadir potencias de una variable deja que un modelo lineal dibuje curvas: una parábola encaja relaciones con vuelta. Es barato y expresivo, pero los grados altos se descontrolan en los bordes; hay que graduarlo con validación.

Escalado y centrado

Restar la media y dividir por la desviación deja cada variable en escala comparable, imprescindible para regularización, vecinos más cercanos o descenso de gradiente. No cambia relaciones, pero cambia quién pesa en modelos sensibles a la magnitud.

Tratar extremos

Recortar colas o usar transformaciones resistentes evita que unos pocos valores descomunales arrastren el ajuste. La elección depende: ¿importa el valor exacto del extremo o solo su orden? No es lo mismo en ingresos que en temperaturas.

Monotonía y su efecto

Las transformaciones crecientes preservan el orden de los datos, así que no alteran métricas basadas en ranking como el área bajo la curva. Cambian, en cambio, la media, la varianza y la forma de los residuos, que es lo que interesa a muchos modelos.

Conocimiento del dominio primero

La mejor transformación nace de entender la variable, no de probar a ciegas. Saber que una magnitud crece exponencialmente, o que tiene un umbral fisiológico, dicta el logaritmo o el tramo. La intuición del experto ataja cien experimentos.

Errores frecuentes

- Aplicar transformaciones sin registrar la inversa, y no poder interpretar en unidades reales.
- Discretizar con demasiados tramos, desperdiciando la ventaja de la transformación.
- Transformar para inflar la métrica de validación sin causa que lo justifique.

Ejemplo resuelto: transformar ingresos para un modelo lineal

Una aseguradora ajusta una regresión para predecir el importe de sus pólizas, pero los ingresos, muy asimétricos, deforman el ajuste. El protocolo:

1. Diagnosticar: observar el histograma de ingresos y confirmar la cola larga hacia la derecha.
2. Aplicar logaritmo para comprimir la cola y repartir la distribución.
3. Reajustar el modelo y comparar la banda de residuos con la versión sin transformar.

4. Guardar la inversa del logaritmo para volver a leer en euros reales.
5. Verificar en validación que la mejora es estable y no un artefacto del azar.

¿Para qué sirve en la realidad?

Ingresos, poblaciones, tiempos de espera o ventas crecen multiplicativamente y desbordan a los modelos lineales. El logaritmo y los tramos, guiados por el dominio, suelen valer más que cambiar de algoritmo: hacen legible un patrón que la escala original escondía.

1. ¿Qué logra típicamente aplicar un logaritmo a una variable con cola pesada?

- ☐ Aumentar la importancia de los extremos
- ☐ Comprimir la cola y repartir la distribución, estabilizando la varianza
- ☐ Volverla lineal siempre
- ☐ Eliminar sus valores atípicos por completo

2. Una transformación creciente altera el orden relativo de las observaciones.

- ☐ Verdadero
- ☐ Falso

Transformar dentro del pipeline

Toda transformación que aprenda parámetros —el escalado, los cortes— debe vivir dentro del pipeline para ajustarse solo con entrenamiento. Aplicarla antes de dividir abre la puerta a la fuga.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Transformaciones



4. Ordena el criterio para transformar una variable:

Elementos:

- entender el dominio y la escala de la variable
- observar su distribucion: colas, umbrales, curvatura
- elegir la transformacion que revele la estructura
- validar que mejora sin introducir fuga

Conocer	. . .
Diagnosticar	. . .
Elegir	. . .
Verificar	. . .

5. Une cada transformación con lo que consigue.

5.	comprime la cola y reparte la asimetría	-----
5.	parte una continua en franjas con umbral	-----
5.	añade potencias para dibujar curvas	-----
5.	resta la media y divide por la desviación	-----

Opciones: logaritmo · discretizacion · polinomio · escalado

Reformular la variable, no solo el modelo

Transformaciones

Logaritmo

colas pesadas

crecimiento multiplicativo

Tramos

umbrales

no monótono

Polinomio

curvatura

cuidado en bordes

Escalado

comparar magnitudes

sin cambiar forma

El mapa mental interactivo es interactivo en la versión online.



6. Una columna de ingresos mejora tu modelo lineal un salto enorme al transformarla. ¿Qué transformación probarías primero y cómo verificarías que no es casualidad?

Referencia curricular

Grado:	5
Materia:	Feature Engineering · Machine Learning

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



<https://univ.lumbre.lat/object/cdia-feng-001-transformaciones-de-variables>

Material educativo generado desde Lumbre · lumbre.lat