

Un modelo de regresión lineal aplicando lógica difusa

Alex Gutiérrez, Wilfrido Ferreira

^{1,2}Departamento de Ciencias Básicas, Universidad Autónoma del Caribe, Barranquilla, Colombia

Article History

Received Dec , 13th , 2020

Accepted Dec ,14th , 2020

Published Dec, 15th,2020



© 2020 The Authors. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution(CC-BY)license

<https://creativecommons.org/licenses/by/4.0>

Abstract

En este artículo se hace una aplicación de la lógica difusa a modelo de regresión lineal con parámetros difusos, utilizando los criterios que sugiere Bo Yuan y Klir [2], comparando los resultados obtenidos con el modelo de regresión lineal probabilístico.

Keywords

Conjuntos difusos, Lógica difusa, Números difusos, Regresión lineal

INTRODUCTION

El análisis de regresión constituye una de las herramientas matemáticas y estadísticas más utilizadas para estudiar las relaciones existentes entre variables. Su propósito fundamental consiste en representar mediante una expresión matemática el comportamiento de una variable dependiente en función de una o varias variables independientes. En la regresión lineal tradicional, los parámetros se estiman a partir de observaciones numéricas y normalmente se supone que la incertidumbre presente en los datos puede explicarse mediante componentes probabilísticos. Sin embargo, numerosas situaciones reales incluyen información caracterizada por vaguedad, ambigüedad e imprecisión que no siempre puede representarse adecuadamente mediante probabilidades. La lógica difusa constituye una alternativa para trabajar con este tipo de incertidumbre. A diferencia de la lógica clásica, donde un elemento pertenece o no pertenece a un conjunto, los conjuntos difusos permiten asignar diferentes grados de pertenencia entre cero y uno. Esta característica permite representar expresiones como “aproximadamente alto”, “moderadamente bajo”, “cercano a diez” o “alrededor de determinado valor”. Por esta razón, la lógica difusa ha encontrado aplicaciones en ingeniería, control automático, economía, estadística, inteligencia artificial y sistemas de apoyo a decisiones. El artículo de Gutiérrez y Ferreira aplica precisamente estos principios al análisis de regresión lineal. Los autores presentan un modelo con parámetros difusos siguiendo criterios asociados con Bo Yuan y George Klir y posteriormente comparan los resultados obtenidos con un modelo probabilístico convencional. El trabajo utiliza valores de entrada ordinarios y representa determinados resultados mediante números difusos triangulares y simétricos. La regresión difusa pretende representar no solamente una tendencia central, sino también la imprecisión que acompaña a la relación entre variables. En lugar de obtener exclusivamente una línea de regresión, puede construirse una región difusa alrededor de las estimaciones. La amplitud de esta región refleja el nivel de incertidumbre asociado con cada predicción. Este enfoque puede resultar particularmente valioso cuando las observaciones contienen información subjetiva o cuando las variables no pueden medirse con exactitud absoluta. La investigación contemporánea demuestra que la regresión difusa continúa evolucionando. Durante 2020 se publicaron modelos basados en aproximación bayesiana, regresión aditiva difusa, operadores de producto difuso, números difusos gaussianos, transformada F y métodos robustos para valores atípicos. Estos trabajos muestran que la regresión difusa constituye un área activa de investigación matemática y computacional. La comparación entre modelos probabilísticos y difusos no implica necesariamente determinar que uno sustituya completamente al otro. Ambos enfoques representan diferentes tipos de incertidumbre. La probabilidad es apropiada cuando la variabilidad puede explicarse mediante fenómenos aleatorios, mientras que los conjuntos difusos resultan especialmente útiles cuando la incertidumbre procede de conceptos vagos, evaluaciones lingüísticas o límites imprecisos.

Fundamentos de la lógica difusa y los conjuntos difusos

La lógica difusa se fundamenta en la teoría de conjuntos difusos, donde la pertenencia de un elemento a un determinado conjunto se expresa mediante una función cuyo valor se encuentra normalmente entre cero y uno. En la teoría clásica de conjuntos, un valor pertenece completamente al conjunto o queda completamente fuera de él. En cambio, un conjunto difuso admite grados intermedios de pertenencia. Esta característica permite representar fenómenos donde las fronteras no son exactas. Por ejemplo, si se define el concepto “temperatura alta”, no existe necesariamente un punto universal a partir del cual una temperatura deba considerarse alta. Una temperatura de 28 °C puede poseer determinado grado de pertenencia al concepto, mientras que una temperatura de 35 °C puede tener un grado superior. La función de pertenencia describe matemáticamente este comportamiento. Entre las funciones más utilizadas se encuentran las triangulares, trapezoidales y gaussianas. Los números difusos triangulares son especialmente populares debido a la facilidad con la que pueden representarse y manipularse mediante operaciones matemáticas. Un número difuso triangular suele caracterizarse por un valor central y límites asociados con su dispersión. El valor central representa el punto con mayor grado de pertenencia, mientras que los extremos indican el intervalo donde la pertenencia disminuye progresivamente. Esta estructura permite expresar valores aproximados de manera matemática. La lógica difusa resulta especialmente útil cuando una observación no puede expresarse con exactitud. Por ejemplo, un experto puede indicar que una determinada variable se encuentra “aproximadamente en 50 unidades”. Convertir esta información directamente en un único número exacto eliminaría una parte importante de su significado. Un número difuso permite conservar simultáneamente el valor central y la incertidumbre que lo rodea. La utilización de conjuntos difusos en regresión amplía, por tanto, el concepto tradicional de estimación. En lugar de buscar únicamente parámetros exactos que minimicen errores, el modelo puede representar parámetros mediante números difusos capaces de incorporar imprecisión estructural.

2. Regresión lineal probabilística y sus características

La regresión lineal probabilística intenta describir la relación entre una variable dependiente y una o más variables independientes. En su forma más sencilla, puede representarse mediante la expresión:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

donde Y representa la variable dependiente, X corresponde a la variable independiente, β_0 es el intercepto, β_1 representa la pendiente y ε corresponde al componente de error. El objetivo principal consiste en estimar los parámetros β_0 y β_1 de tal forma que la línea obtenida represente adecuadamente las observaciones disponibles. El método de mínimos cuadrados constituye uno de los procedimientos más utilizados. Su principio consiste en seleccionar los parámetros que minimizan la suma de los cuadrados de las diferencias entre valores observados y valores estimados. La regresión convencional proporciona una estimación central y diferentes mecanismos para estudiar su incertidumbre, como errores estándar, intervalos de confianza y predicción. Estos procedimientos son especialmente adecuados cuando la variabilidad observada puede representarse mediante supuestos probabilísticos. Sin embargo, algunas aplicaciones presentan incertidumbre que no proviene únicamente del comportamiento aleatorio. Evaluaciones humanas, escalas lingüísticas y datos aproximados pueden contener ambigüedad conceptual. En estas circunstancias, imponer valores exactos puede producir una representación artificialmente precisa. La regresión difusa surge como una alternativa destinada a incorporar dicha imprecisión dentro de los propios parámetros y resultados del modelo. Esta diferencia conceptual constituye uno de los principales aspectos analizados por Gutiérrez y Ferreira.

Table 1 Comparación conceptual entre regresión probabilística y regresión difusa

Característica	Regresión probabilística	Regresión difusa
Tipo principal de incertidumbre	Aleatoriedad	Vaguedad e imprecisión
Representación de parámetros	Valores reales estimados	Números o coeficientes difusos
Resultado principal	Línea o función estimada	Función acompañada de una región difusa
Tratamiento del error	Distribuciones probabilísticas y residuos	Grados de pertenencia y amplitud difusa
Datos lingüísticos	Requieren transformación	Pueden representarse naturalmente
Interpretación	Probabilidad y variabilidad estadística	Posibilidad y grado de pertenencia
Aplicación apropiada	Datos numéricos con supuestos estadísticos	Datos imprecisos, aproximados o lingüísticos

Regresión lineal con parámetros difusos

En un modelo de regresión difusa, algunos o todos los coeficientes pueden representarse mediante números difusos. Conceptualmente, el modelo puede escribirse como:

$$\tilde{Y} = \tilde{C}_0 + \tilde{C}_1 X_1 + \tilde{C}_2 X_2 + \dots + \tilde{C}_n X_n$$

donde los coeficientes $\tilde{C}$ representan parámetros difusos, mientras las variables independientes pueden ser valores ordinarios o también variables difusas dependiendo del enfoque utilizado. En el trabajo de Gutiérrez y Ferreira, la regresión se desarrolla utilizando parámetros difusos y números triangulares simétricos. Los autores aplican criterios asociados con Klir y Yuan para obtener los coeficientes y posteriormente construyen las correspondientes observaciones difusas. Un coeficiente difuso posee normalmente un componente central y un término de dispersión. El componente central desempeña un papel semejante al coeficiente de regresión convencional, mientras que la dispersión indica la cantidad de imprecisión asociada con ese parámetro. Cuando los valores de las variables independientes cambian, esta dispersión se propaga hacia los valores estimados de Y . Por ello, el resultado gráfico no se reduce necesariamente a una única línea, sino que puede visualizarse como una banda cuya amplitud representa diferentes grados posibles de respuesta. El desarrollo de estos modelos puede realizarse mediante diferentes enfoques matemáticos. Algunos métodos buscan asegurar que las observaciones estén contenidas dentro de regiones difusas, mientras que otros minimizan distancias entre datos observados y estimaciones. Investigaciones publicadas en 2020 continuaron desarrollando variantes que combinan métodos posibilísticos, mínimos cuadrados y técnicas bayesianas. Esta flexibilidad hace posible diseñar el modelo en función de la naturaleza específica de los datos disponibles.

Números difusos triangulares y representación de la incertidumbre

Los números triangulares constituyen una de las representaciones más utilizadas dentro de la lógica difusa. Su popularidad se debe a que permiten describir de forma sencilla un valor aproximado utilizando tres elementos: límite inferior, valor central y límite superior. Por ejemplo, un valor representado como aproximadamente 50 podría expresarse mediante un número difuso cuyos límites fueran 45 y 55 y cuyo punto central fuera 50. El grado de pertenencia aumenta progresivamente desde 45 hasta alcanzar su máximo en 50 y posteriormente disminuye hasta llegar nuevamente a cero en 55. Cuando el número es simétrico, la distancia entre el centro y ambos extremos es igual. Esta estructura facilita considerablemente las operaciones utilizadas dentro de modelos de regresión. La amplitud del número difuso desempeña un papel fundamental. Una amplitud pequeña indica que existe relativamente poca imprecisión, mientras que una amplitud grande representa mayor incertidumbre. En una regresión difusa, cada predicción puede interpretarse conjuntamente mediante su centro y su amplitud. Sin embargo, los números triangulares no constituyen la única alternativa. Lim y Yoon propusieron en 2020 una regresión lineal basada en números difusos gaussianos, mostrando que determinadas aplicaciones pueden requerir funciones de pertenencia más flexibles que las formas triangulares o trapezoidales. La elección del tipo de número difuso debe depender de las características del problema. Si la información puede describirse de manera razonable mediante un centro y dos límites lineales, una representación triangular puede ser suficiente. Cuando existe una transición más suave alrededor del valor central, una función gaussiana puede resultar más apropiada.

Comparación entre regresión difusa y regresión probabilística

La comparación entre ambos métodos permite comprender que representan diferentes concepciones de incertidumbre. En la regresión probabilística, las diferencias entre observaciones y predicciones suelen considerarse errores aleatorios. El análisis se centra en estudiar distribuciones, varianzas y probabilidades. En la regresión difusa, la incertidumbre puede interpretarse como una propiedad inherente a la descripción de los datos. Una observación puede ser imprecisa no porque haya ocurrido un error aleatorio, sino porque la variable observada no admite un límite completamente exacto. El artículo original concluye que la regresión lineal difusa desarrollada no resulta tan precisa como la regresión probabilística desde el punto de vista del ajuste convencional, pero ofrece la ventaja de adaptarse mejor al lenguaje y a situaciones caracterizadas por imprecisión. Esta conclusión es especialmente importante porque la utilidad de un modelo no debe juzgarse únicamente mediante el grado de proximidad a una línea. Un modelo también debe representar adecuadamente la naturaleza de la información que pretende estudiar. Investigaciones posteriores y contemporáneas han desarrollado procedimientos que buscan mejorar simultáneamente el ajuste y la representación de incertidumbre. Por ejemplo, Chen y Nien propusieron un nuevo enfoque para formular modelos de regresión difusa y reportaron mejoras utilizando medidas de distancia, similitud y credibilidad. Otros trabajos han desarrollado métodos para manejar valores atípicos mediante regresión difusa robusta, demostrando que la metodología puede integrarse con técnicas modernas de aprendizaje automático.

Interpretación gráfica y aplicaciones del modelo

Una de las formas más claras de comprender la regresión difusa consiste en representarla gráficamente. En la regresión tradicional se dibuja una línea central que busca aproximarse a los puntos observados. En la regresión difusa puede representarse una banda alrededor de esa línea. La banda muestra la región asociada con la incertidumbre del modelo. Los límites superior e inferior pueden cambiar dependiendo de los valores de X y de la dispersión correspondiente a los parámetros difusos. La interpretación visual demuestra una diferencia fundamental. Si una observación se encuentra alejada de la línea tradicional, normalmente se interpreta mediante un residuo. En una representación difusa,

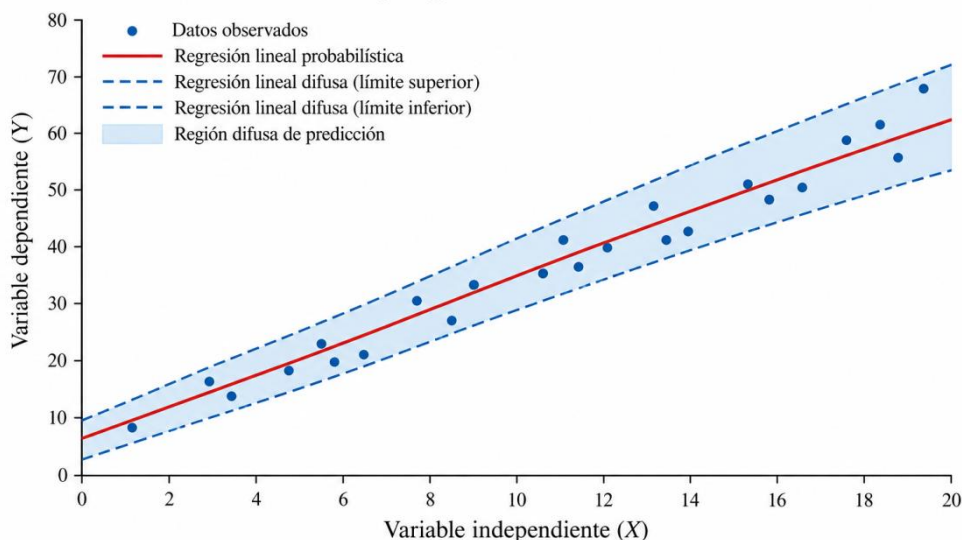
determinadas observaciones pueden encontrarse dentro de una región de posibilidad asociada con la amplitud del modelo. Esta característica puede ser relevante en problemas de ingeniería donde algunos parámetros proceden de estimaciones de expertos, en economía cuando se utilizan conceptos lingüísticos y en sistemas de decisión donde las variables se expresan mediante categorías como bajo, medio o alto. La regresión difusa también puede combinarse con otras metodologías. En 2020 se propusieron enfoques híbridos basados en F-transform, regresión difusa bayesiana, modelos aditivos y sistemas de predicción temporal.

Limitaciones, aplicaciones futuras y desarrollo de nuevos modelos

A pesar de sus ventajas, la regresión difusa presenta varios desafíos. El primero corresponde a la selección de funciones de pertenencia. Una elección inadecuada puede introducir una cantidad excesiva o insuficiente de incertidumbre. Otro desafío consiste en determinar la amplitud apropiada de los parámetros. Si los intervalos difusos son demasiado amplios, prácticamente todas las observaciones pueden quedar contenidas dentro del modelo, pero la capacidad explicativa disminuye. Si son demasiado estrechos, el modelo puede dejar de representar adecuadamente la vaguedad de los datos. También debe considerarse la interpretación. Mientras los resultados de una regresión clásica pueden expresarse mediante probabilidades e intervalos estadísticos ampliamente conocidos, los grados de pertenencia y posibilidad requieren una interpretación diferente. Es importante evitar confundir probabilidad con pertenencia difusa. No obstante, la investigación actual continúa desarrollando métodos destinados a mejorar estos aspectos. En 2020 se publicaron modelos de regresión aditiva para respuestas difusas, metodologías completamente difusas basadas en programación por metas y procedimientos para incorporar información lingüística probabilística.

La integración con aprendizaje automático constituye otra línea importante. Las técnicas difusas pueden utilizarse para mejorar la interpretación de sistemas donde los datos poseen ruido, valores atípicos o incertidumbre difícil de representar mediante métodos convencionales. En aplicaciones industriales, los modelos pueden emplearse para estimar comportamientos de procesos cuyas variables se obtienen mediante mediciones imprecisas. En administración y economía pueden aplicarse cuando algunas variables proceden de evaluaciones subjetivas. En ciencias ambientales pueden representar condiciones descritas mediante categorías lingüísticas. También pueden utilizarse en sistemas inteligentes donde la información proviene simultáneamente de sensores y expertos humanos. Esta combinación resulta particularmente apropiada porque la lógica difusa proporciona un lenguaje matemático para incorporar conocimiento cualitativo dentro de procedimientos cuantitativos. Por consiguiente, el trabajo desarrollado por Gutiérrez y Ferreira representa una demostración introductoria de cómo la regresión tradicional puede ampliarse mediante teoría de conjuntos difusos. La investigación posterior demuestra que existe un amplio campo para desarrollar modelos híbridos capaces de combinar probabilidades, posibilidad, aprendizaje automático y diferentes representaciones de incertidumbre.

Comparación conceptual entre regresión probabilística y regresión difusa



Representación ilustrativa; no corresponde a los datos exactos del estudio original.

Resumen

La regresión lineal constituye una herramienta fundamental para estudiar relaciones entre variables y elaborar predicciones. Sin embargo, el enfoque probabilístico tradicional está diseñado principalmente para trabajar con

incertidumbre asociada con variabilidad aleatoria y puede presentar limitaciones cuando los datos contienen vaguedad o imprecisión lingüística. La lógica difusa proporciona una alternativa mediante la teoría de conjuntos difusos. En estos conjuntos, la pertenencia no se limita a valores binarios, sino que puede expresarse mediante diferentes grados. Este principio permite representar conceptos que poseen límites imprecisos. El modelo estudiado por Gutiérrez y Ferreira aplica lógica difusa a una regresión lineal utilizando parámetros representados mediante números difusos. La propuesta se basa en criterios asociados con Bo Yuan y George Klir y compara sus resultados con los obtenidos mediante regresión probabilística. Los números difusos triangulares desempeñan un papel importante dentro del modelo. Estos números permiten representar simultáneamente un valor central y una determinada dispersión alrededor de él. Como resultado, las estimaciones pueden mostrar no solamente una tendencia central, sino también la imprecisión asociada con la predicción. Una diferencia fundamental entre ambos enfoques se encuentra en la interpretación de la incertidumbre. La regresión probabilística considera principalmente variabilidad aleatoria y utiliza herramientas como distribuciones, residuos e intervalos de confianza. La regresión difusa estudia principalmente vaguedad e imprecisión mediante grados de pertenencia y regiones de posibilidad. El trabajo original señala que el modelo difuso puede resultar menos preciso que el modelo probabilístico cuando la comparación se realiza exclusivamente desde el punto de vista del ajuste tradicional. No obstante, ofrece una ventaja importante al representar información asociada con lenguaje impreciso y situaciones donde los datos no pueden describirse mediante valores completamente exactos. La representación gráfica facilita la comprensión de esta diferencia. Una regresión probabilística puede visualizarse mediante una línea central acompañada eventualmente de intervalos estadísticos. Una regresión difusa puede visualizarse mediante una región delimitada por diferentes niveles de posibilidad alrededor del comportamiento central. Los avances publicados durante 2020 muestran que la regresión difusa continúa evolucionando. Se desarrollaron modelos utilizando aproximación bayesiana, números gaussianos, regresión aditiva, métodos robustos para valores atípicos, F-transform y programación difusa por metas. Estas metodologías buscan resolver diferentes limitaciones asociadas con la selección de funciones de pertenencia, estimación de parámetros, manejo de valores extremos y reducción de dispersión innecesaria. La principal fortaleza de la regresión difusa reside en su capacidad para modelar sistemas donde la incertidumbre no puede interpretarse únicamente como aleatoriedad. Esto resulta especialmente importante cuando intervienen opiniones de expertos, variables lingüísticas, mediciones aproximadas y conceptos cuyos límites no son completamente definidos. En conclusión, la regresión lineal basada en lógica difusa no debe considerarse simplemente como una sustitución de la regresión probabilística. Ambas metodologías responden a diferentes tipos de incertidumbre y pueden complementarse dependiendo de las características del problema. La lógica difusa amplía las posibilidades de análisis estadístico y matemático al permitir que la vaguedad y la imprecisión se incorporen explícitamente dentro de los modelos de regresión.

Referencias

- Chakravarty, S., Demirhan, H., & Baser, F. (2020). Fuzzy regression functions with a noise cluster and the impact of outliers on mainstream machine learning methods in the regression setting. *Applied Soft Computing*, 96, 106535.
- Chen, L. H., & Nien, S. H. (2020). A new approach to formulate fuzzy regression models. *Applied Soft Computing*, 86, 105915.
- Efendi, R., Aini, Z., & Susanti, R. (2020). Model fuzzy linear regression untuk peramalan produksi kelapa sawit. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI 12)*.
- Gutiérrez, A., & Ferreira, W. (2020). Un modelo de regresión lineal aplicando lógica difusa. *Revista SEXTANTE*, 23, 48–54.
- Hesamian, G., & Akbari, M. G. (2020). A fuzzy additive regression model with exact predictors and fuzzy responses. *Applied Soft Computing*, 95, 106507.
- Jung, H. Y., Lee, W. J., & Choi, S. H. (2020). Hybrid fuzzy regression analysis using the F-transform. *Applied Sciences*, 10(19), 6726.
- Lim, S., & Yoon, J. H. (2020). Fuzzy linear regression using Gaussian fuzzy numbers. *Journal of Korean Institute of Intelligent Systems*, 30(5), 386–390.
- Mohamad Hanapi, A. L., Othman, M., Sokkalingam, R., Ramli, N., Husin, A., & Vasant, P. (2020). A novel fuzzy linear regression sliding window GARCH model for time-series forecasting. *Applied Sciences*, 10(6), 1949.
- Pérez-Cañedo, B., Rosete, A., Verdegay, J. L., & Concepción-Morales, E. R. (2020). A fuzzy goal programming approach to fully fuzzy linear regression. *Communications in Computer and Information Science*, 1238, 677–688.
- Wang, N., Reformat, M., Yao, W., Zhao, Y., & Zhang, X. (2020). Fuzzy linear regression based on approximate Bayesian computation. *Applied Soft Computing*, 97, 106763.

- Wang, T., Shi, P., & Wang, G. (2020). Solving fuzzy regression equation and its approximation for random fuzzy variable and their application. *Soft Computing*, 24, 4547–4559.
- Zack, Y. A. (2020). Fuzzy linear regression analysis, taking into account the influence of input factors. *Information Technologies*, 26(3), 137–144.