Subject: Probability and Statistics

Class XII: Regresiones no lienales, y Logísticas

UdeA

Bioengineering

Francisco José Campuzano Cardona

Bioengineerer, MSc in Engineering

UdeA

Variables Categóricas

En la regresión lineal pueden usarse variables categóricas, pero estas deben ser codificadas apropiadamente en números.

Por ejemplo, si la variable es binaria, la codificación más apropiada es 1 y 0.

Si las categorías de la variable reflejan algún orden de peso o importancia, entonces debería usarse una variable ordinal: 1, 2, 3, 4, 5,

Si la variable tiene más de 2 niveles, entonces se podría tratar cada nivel como una variable binaria independiente.

Existen múltiples formas de codificar las variables categóricas.

Regresión Lineal – Otros conceptos

UdeA

Los modelos de Regresión siguen el principio de Parsimonia

Esto quiere decir que se debe buscar el modelo más simple posible. Complejizar un modelo generalmente no es mejor.
Por ejemplo, la inclusión de variables descriptoras siempre mejorará el RMSE y el R2, pero esto no significa necesariamente que el modelo tiene una mejor especificación, o que es justificable.

Existen estrategias de penalización para la inclusión de nuevas variables, que permiten guiar la construcción del modelo. Por ejemplo Akaike's Information Criteria (AIC), (se busca minimizarlo).

$$AIC = 2p + n\log(RSS/n)$$

p: numero de descriptores
n: número de observaciones
RSS: Cuadrado de la sumatoria de los residuos

UdeA

Los modelos de Regresión en general no son extrapolables

Un modelo de regresión no puede ser generalizado a cualquier valor de las variables descriptoras, en general, solo debería considerarse válido para el rango de las variables descriptoras que fue usado por el entrenamiento.

Por ejemplo, si un variable X tenía valores entre 10 y 20, y se usó para describir una variable Y, no deberíamos usar el modelo construido para saber el valor de Y cuando X = 50

UdeA

Variables predictoras Correlacionadas

Se puede dar el caso que haya correlación entre las variables predictoras, esto puede causar que la interpretación del peso y signo de los términos del modelo, sea más complejo.

Multicolinearidad

Un caso extremo de variables correlacionadas, es que sean combinaciones lineales perfectas, lo cual podría verse en un análisis explotarlo al analizar la correlación de las variables. Si estas variables tienen correlación perfecta, no deberían ser incluidas ambas en el modelo. Esto causará que el modelo no converja.

UdeA

Regresión Polinómica

La relación entre las variables predictoras y la variable a ser descrita, no necesariamente tiene que ser lineal, podrían guardar un relación de orden superior, y en consecuencia el modelo es polinomial.

$$Y = b_0 + b_1 X + b_2 X^2$$

Se puede generalizar a más de un regresor.

UdeA

Regresión por Splines

Para describir curvaturas entre los datos regresores y los descritos, no siempre la mejor opción es aumentar el orden del polinomio, sino describir la relación a tramos. Entonces un *Spline* es una serie de polinomios continuos a tramos. Estos polinomios están unidos de forma suave por puntos fijos de la variable predictora, conocidos como *knots.*

Regresión Polinómica

UdeA

Ejemplo:

Vamos a intentar encontrar un modelo que describa el comportamiento entre el ángulo de difracción y la intensidad de la difracción, en un ensayo de DRX de una muestra de un material lignocelulósico

UdeA

Interacciones y efectos principales

Los efectos de las variables independientes se conocen como efectos principales.

Al igual que en los modelos ANOVA, las variables predictoras pueden tener interacciones entre ellas, lo cual implica efectos adicionales en el modelo.

Veamos como se construyen estos otros efectos con sklearn.

Regresión Lineal – Otros conceptos

Los coeficientes cuando se incluyen las interacciones, quedan en el siguiente orden:

$[X_1, X_2, X_3, \ldots, X_n, X_1*X_2, X_1*X_3, \ldots, X_1*X_n, X_2*X_3, \ldots, X_2*X_n, \ldots X_{n-1}*X_n]$

Subject: Probability and Statistics

UdeA

Bioengineering

¡Thanks!

Francisco José Campuzano Cardona

Bioengineering. MSc in Engineering