ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
FACULTAD DE CIENCIA NATURALES Y MATEMÁTICAS
ÁLGEBRA LINEAL • MATG1003

PROYECTO

Término II · 2017 – 2018

APROXIMACIÓN POR MÍNIMOS CUADRADOS

En las ciencias biológicas es esencial estudiar las relaciones entre distintos parámetros con el objetivo de tener indicadores con valor predictivo, descubrir conexiones de causa-efecto, o conocer el estado de un sistema según una variable relacionada al mismo. Esto conlleva una apreciable cantidad de trabajo a realizar. Las mediciones y observaciones experimentales raramente se comportan de manera exacta según una fórmula, y uno de los recursos que queda al experimentador es lograr hallar la fórmula que mejor se aproxime al comportamiento dado por las mediciones.

Por ejemplo, en 2015 se publicó un estudio sobre el efecto tóxico que tiene la concentración de nitratos en los cultivos del camarón de cola blanca (*Litopenaeus Vannamei*), sobre el crecimiento de los mismos. Se determinó que la tasa de crecimiento semanal G (en gr/semana) decrece a medida que se incrementa la concentración de nitratos N (en mg/L), según la expresión:

$$\mathbf{G} = 0{,}874 - 0{,}0007 \times \mathbf{N}$$

La Figura 1 ilustra los resultados de ese estudio:

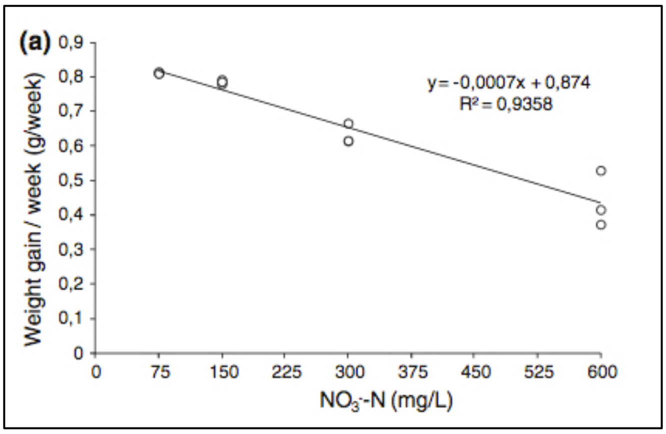

Figura 1: Decrecimiento de la ganancia de peso semana según la concentración de nitratos en las piscinas de cultivo de camarón de cola blanca, *L. Vannamei*.

En Figura 1, los pequeños círculos indican los datos reales medidos, mientras que la líne recta corresponde a la función lineal que mejor se aproxima a los datos medidos, no necesariamente que contenga dichos puntos.

REGRESIÓN LINEAL

Suponga que se dispone de dos secuencias de mediciones **x** e **y**:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

con correspondencia entre mediciones (a la i-ésima medida x_i corresponde la medición y_i). Suponga que hay evidencia experimental de una relación lineal entre los vectores dados, de modo que se desea encontrar una expresión de tipo $y = mx + b$ (lineal), cuya gráfica represente mejor el conjunto de puntos (x_i, y_i) de las mediciones tomadas.

Si la relación fuera exacta, y las mediciones perfectas y sin errores, se cumpliría que:

$$y_1 = b + mx_1,\quad y_2 = b + mx_2,\quad \dots\quad y_n = b + mx_n$$

pero en la vida real esto nunca se cumple, por lo cual, al evaluar los x_i en la recta $y = mx + b$ se producen como resultado $\hat{y}_i$ que no son iguales a los y_i:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{bmatrix} b \\ m \end{bmatrix} \quad \text{o} \quad \hat{\mathbf{y}} = A\mathbf{u}.$$

¿Cuáles son los valores de los parámetros m y b que producen la menor distancia entre los y_i y los $\hat{y}_i$? Es decir, que minimizan la distancia entre $\mathbf{y}$ y $\hat{\mathbf{y}}$, $\|\mathbf{y} - \hat{\mathbf{y}}\|$.

La expresión $\hat{\mathbf{y}} = A\mathbf{u}$ corresponde a un sistema de ecuaciones lineales, con $A_{n\times 2}$, y donde $\hat{\mathbf{y}}$ es un vector de $\mathbb{R}^n$ que pertenece a la imagen de A. Así también, $\mathrm{Im}(A)$ es un subespacio de dimensión menor o igual que 2, pues A solo tiene dos columnas.

Según el teorema de la Aproximación de la Norma, la distancia $\|\mathbf{y} - \hat{\mathbf{y}}\|$ es mínima cuando $\mathbf{y} - \hat{\mathbf{y}}$ es perpendicular a $\mathrm{Im}(A)$.

Tarea 1: Enuncie y demuestre el Teorema de la Aproximación de la Norma.

Sea $\bar{\mathbf{u}}$ el vector minimiza $\|\mathbf{y} - \hat{\mathbf{y}}\|$, es decir, se cumple que:

$$A\mathbf{u} \perp (\mathbf{y} - A\bar{\mathbf{u}}),$$

o:

$$\langle A\mathbf{u}, (\mathbf{y} - A\bar{\mathbf{u}}) \rangle = 0$$

Aplicando las propiedades del producto interno:

$$\langle A\mathbf{u}, \mathbf{y} \rangle - \langle A\mathbf{u}, A\bar{\mathbf{u}} \rangle = 0$$

$$\mathbf{u}^T (A^T \mathbf{y} - A^T A\bar{\mathbf{u}}) = 0$$

lo cual se cumple para cualquier $\mathbf{u}$, solo si $A^T \mathbf{y} = A^T A\bar{\mathbf{u}}$. Y el vector $\bar{\mathbf{u}}$ es:

$$\bar{\mathbf{u}} = (A^T A)^{-1} A^T \mathbf{y}$$

Tarea 2: Justificar por qué se afirma que necesariamente $A^T \mathbf{y} = A^T A\bar{\mathbf{u}}$.

Teorema 1: La matriz $A^T A$ es invertible si los n puntos de datos (x_i, y_i) no son colineales.

Tarea 3: Demostrar el Teorema 1.

ꕤ

Tarea 4: Deducir la formulación para la regresión cuadrática, cuando se quiere relacionar dos secuencias de mediciones $\mathbf{x}$ e $\mathbf{y}$ mediante la expresión $y_i = c + bx_i + ax_i^2$. ¿Qué forma tienen en este caso las matrices $\mathbf{A}$ y $\mathbf{u}$ en la ecuación $\hat{\mathbf{y}} = A\mathbf{u}$? ¿Cuál es la expresión para el vector $\bar{\mathbf{u}}$ que minimiza la distancia entre $\mathbf{y}$ e $\hat{\mathbf{y}}$?

Tarea 5: Estudios sobre nutrición mineral de las plantas de arroz han determinado que la concentración de sulfuros en el terreno de cultivo afecta la altura de la planta (en cm), y el rendimiento de arroz producido (en gramos por planta). Si la relación entre Rendimiento y concentración de fertilizante (en mg/Kg) es de tipo cuadrático, encuéntrese la función de rendimiento, basado en los datos de la tabla siguiente:

Tasa de Fertilizante sulfuroso [mg/Kg]	Altura de la planta [cm]	Rendimiento g/planta
0	99.5	7.57
10	112.5	11.83
20	112.75	10.78
30	114.25	12.05
40	106.75	12.00
80	109.75	9.96

La relación entre la altura de la planta y la concentración de fertilizante también es cuadrática, halle la función de altura según los datos de la tabla dada.

ꕤ

MONITOREO DE CRECIMIENTO CONTROLADO DE BACTERIAS

En un laboratorio de análisis de contaminación y control de calidad alimentaria se estudia la tasa de crecimiento de la bacteria *Escherichia Coli* como una función de la concentración de glucosa.

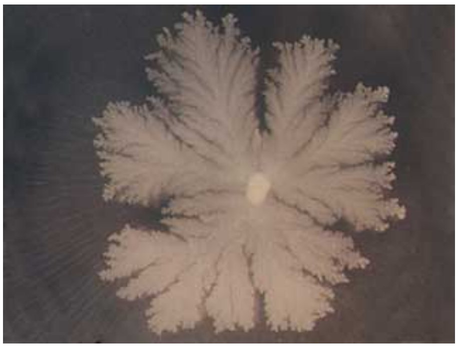

Figura 1: Cultivo de *E. Coli* bajo condiciones controladas

La tasa de crecimiento $\mathbf{V}$ (en h^{-1}) y las concentraciones de glucosa $\mathbf{s}$ (en μM) se proveen en el archivo anexo. El investigador principal hipotiza que estas dos variables están relacionadas mediante las relaciones cinética Michaelis-Menten:

$$V = V_{\max}\left(\frac{s}{Km+s}\right)$$

El diagrama de los datos no se ajusta a los modelos conocidos de regresión lineal o cuadrática por lo cual se decide graficar también los datos en el formato (1/**V**) en función de (1/**s**). Aplique los métodos de aproximación por mínimos cuadrados para obtener una expresión que permita calcular la tasa de crecimiento **V** en h^{-1}, a partir de las mediciones de concentración de glucosa **s** en μM. Provea las respuestas requeridas en Entregables.
NOTA: Los datos se proveen en el archivo adjunto:
Proyecto_Regresion_Biologia.xlsx

REFERENCIAS:

- Grossman, S. I., Flores Godoy, J.: Álgebra Lineal. McGraw-Hill Ed., México (2012).
- Furtado, P.S., Campos, B.R., Serra, F.P., Klosterhoff, M., Romano, L.A., Wasielesky Jr. W.: Effects of nitrate toxicity in the Pacific white shrimp, *Litopenaeus vannamei*, reared with biofloc technology (BFT). Aquacult Int, Springer (2015) 23:315–327. DOI 10.1007/s10499-014-9817-z
- Fageria, N.K.: Mineral Nutrition of Rice. CRC Press, Boca Ratón, FL (2014).
- Boston University: Center for Polymer Studies. Bacterial Growth, en: Exploring Patterns in Nature. Physionet.org (2012). https://physionet.org/tutorials/epn/ Recuperado el 10 de Octubre de 2017.
- Johnson, K.A., Goody, R.S.: The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper, in Biochemistry, vol. 50, nº 39, 2011, pp. 8264–8269, DOI:10.1021/bi201284u.
- Banfelder, J.: Quantitative Understanding in Biology. Department of Physiology and Biophysics of the Cornell University. http://physiology.med.cornell.edu/people/banfelder/qbio/all_posts.html Recuperado el 5 de Octubre de 2017.

Entregables:

- Se debe presentar dos secciones. La primera parte consta de la resolución de las tareas planteadas, respondiendo a las preguntas dadas.
- En la segunda sección debe constar exclusivamente el análisis de regresión con los datos del cultivo bacterial controlado, proveyendo respuestas a las siguientes cuestiones:

1. Elabore un gráfico de los datos originales provistos (**V** y **s**), y de sus recíprocos (1/**V** y 1/**s**)
2. A partir del resultado del primer ítem, justifique su decisión sobre la elección de un modelo lineal o uno cuadrático para las variables 1/**V** y 1/**s**.
3. A partir de los cálculos hechos, elabore una tabla con los valores estimados de los parámetros del vector **u** correspondiente; y calcule los parámetros V_{max} y K_m.

4. Elabore un gráfico de la función elegida para aproximar los puntos de los datos.
5. Uno de los principales problemas filosóficos relacionados es la cuestión de la correlación y la causa-efecto. La frase “correlación no implica causación” acompaña a los científicos durante toda su vida como admonición. Explique en sus propias palabras y mediante ejemplos esta frase, y distinga entre los dos casos: la correlación y la causa-efecto.
6.
7. Si leyó o consultó un libro o artículo, debe poner al final del documento una sección Bibliografía o Referencias y hacer una lista de cada trabajo consultado, incluyendo título, autor(es), Capítulo/Volumen, páginas.

NOTA: Es lícito apoyarse en la tecnología, pero si necesita utilizar un software o calculadora (Matlab®, Phyton, etc), o algún sitio de resolución de matrices online (Online Matrix Calculator de blue-bit), debe ser indicado en el documento a entregar, planteando la fórmula teórica, e indicando si se utilizó para resolver esa ecuación; y repitiendo la indicación para cada una de las ecuaciones así resueltas.