U N I V E R S I D A D D E
G U A D A L A J A R A

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍA
DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

Seminario de Solución de Problemas de Sistemas Basados en Conocimiento

Práctica No. 6

Regresión Polinomial Múltiple

Nombre: Hurtado González Edgar Arturo
Código: 212597894

Introducción

La regresión polinomial múltiple es una extensión de la regresión lineal múltiple que permite modelar relaciones no lineales entre variables. Mientras que en la regresión lineal se asume una relación lineal entre las variables independientes y la variable dependiente, en la regresión polinomial múltiple se introducen términos polinomiales para capturar relaciones más complejas.

En la regresión polinomial múltiple, se utilizan términos polinomiales de las variables independientes en lugar de simplemente incluir las variables originales. Por ejemplo, si se tiene una variable $\boldsymbol{x}$ como predictora en lugar de simplemente utilizar $\boldsymbol{x}$, se pueden incluir $\boldsymbol{x}^2$, $\boldsymbol{x}^3$, $\boldsymbol{x}^4$ y así sucesivamente, lo que permite capturar curvaturas o formas no lineales en los datos.

La fórmula general para un modelo de regresión polinomial múltiple con una variable independiente se vería así: $\boldsymbol{y} = \boldsymbol{b_0} + \boldsymbol{b_1 x} + \boldsymbol{b_2 x^2} + \cdots + \boldsymbol{b_n x^n} + \boldsymbol{\epsilon}$

Donde:

- $\boldsymbol{y}$ es la variable dependiente.
- $\boldsymbol{x}$ es la variable independiente.
- $\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \dots, \boldsymbol{\beta_n}$ son los coeficientes a estimar.
- $\boldsymbol{n}$ es el grado del polinomio.
- $\boldsymbol{\epsilon}$ es el término de error.

Este modelo puede ajustarse utilizando técnicas de regresión, como el método de mínimos cuadrados, para encontrar los valores óptimos de los coeficientes que minimizan la diferencia entre los valores observados y los predichos por el modelo.

Es importante tener en cuenta que la elección del grado del polinomio es crucial en la regresión polinomial, ya que un grado muy alto puede llevar a un sobreajuste (overfitting) del modelo a los datos de entrenamiento, lo que significa que el modelo se adapta demasiado a esos datos específicos y no generaliza bien a nuevos datos. Por otro lado, un grado muy bajo puede no capturar la complejidad de los datos.

La validación cruzada y otras técnicas de evaluación del rendimiento del modelo son fundamentales para seleccionar el grado óptimo del polinomio y evitar problemas de sobreajuste.

La regresión polinomial múltiple tiene varias características y consideraciones interesantes:

1. *Flexibilidad en la modelización: Al* permitir términos polinomiales de diferentes grados, la regresión polinomial múltiple puede modelar una amplia gama de relaciones no lineales entre las variables independientes y la variable dependiente. Esto es útil cuando la relación real en los datos no es lineal.

2. *Selección del grado del polinomio:* Encontrar el grado adecuado del polinomio es crucial. Se pueden usar técnicas como la validación cruzada o métodos de selección de modelos (como el criterio de información de Akaike o el criterio de información bayesiano) para determinar el grado óptimo que equilibre el ajuste del modelo y su capacidad de generalización.

3. *Riesgo de sobreajuste:* A medida que aumenta el grado del polinomio, el modelo puede volverse demasiado complejo y ajustarse demasiado a los datos de entrenamiento, lo que conlleva un riesgo de sobreajuste. Esto significa que el modelo puede funcionar bien con los datos de entrenamiento, pero tendrá un rendimiento pobre con datos nuevos no vistos.

4. *Interpretación de coeficientes:* En la regresión polinomial, la interpretación de los coeficientes puede ser más compleja que en la regresión lineal, especialmente con términos polinomiales de grado superior. Por ejemplo, interpretar el efecto de un predictor $\boldsymbol{x}$ en la variable dependiente puede requerir considerar cómo interactúan los diferentes términos polinomiales.

5. *Transformaciones de variables:* En ocasiones, transformar las variables puede ser útil para mejorar el ajuste del modelo. Por ejemplo, si los datos muestran una relación exponencial, tomar el logaritmo de las variables puede convertir la relación en una más lineal, facilitando su modelización mediante una regresión polinomial.

6. *Aplicaciones prácticas:* La regresión polinomial múltiple se utiliza en una amplia gama de campos, desde ciencias físicas y biológicas hasta finanzas y análisis de mercado. Se emplea para modelar relaciones no lineales en situaciones donde la regresión lineal no sería adecuada.

Desarrollo

Este código realiza una regresión polinomial usando el método del gradiente descendente para predecir el peso de diferentes especies de peces en función de diversas características.

1. *Preparación de datos:*
   - Lee datos de un archivo CSV ('Fish.csv') usando Pandas.
   - Mapea las especies a valores numéricos para poder trabajar con ellas.
   - Divide los datos en características (X) y la variable objetivo, el peso (Y).

2. *Funciones para el modelo de regresión polinomial por gradiente descendente:*
   - polynomial_features: Crea características polinomiales a partir de las características originales hasta un grado dado.
   - gradient_descent: Implementa el gradiente descendente para ajustar los coeficientes del modelo.
   - polynomial_Rfit: Realiza la regresión polinomial ajustando los coeficientes mediante el gradiente descendente.
   - polynomial_Rpredict: Realiza predicciones utilizando las características polinomiales y los coeficientes del modelo.

3. *División de datos:*
   - Divide los datos en conjuntos de entrenamiento, validación y prueba usando train_test_split.

4. *Entrenamiento y evaluación del modelo:*
   - Itera sobre diferentes grados polinomiales y tasas de aprendizaje.
   - Realiza la regresión polinomial usando el conjunto de entrenamiento y evalúa el error cuadrático medio (MSE) en el conjunto de validación para cada grado polinomial.

5. *Visualización de gráficos:*
   - Utiliza Matplotlib para representar gráficos 3D de las relaciones entre las características y el peso.
   - Grafica diferentes combinaciones de características y el peso para observar la relación entre ellas y las predicciones del modelo.

6. *Resultados y visualización:*
   - Imprime el MSE para cada grado polinomial en el conjunto de validación.
   - Muestra gráficos 3D que representan la relación entre las características y el peso, además de las predicciones del modelo.

Conclusión

En conclusión, la regresión polinomial múltiple es una extensión poderosa de la regresión lineal que permite modelar relaciones no lineales entre variables. Sin embargo, su uso requiere cuidado en la selección del grado del polinomio y en la interpretación de los resultados para evitar problemas de sobreajuste y garantizar una modelización precisa y útil.

El código realiza una regresión polinomial utilizando el método del gradiente descendente para predecir el peso de los peces según diferentes características, y visualiza estas relaciones en gráficos 3D.

Resultados

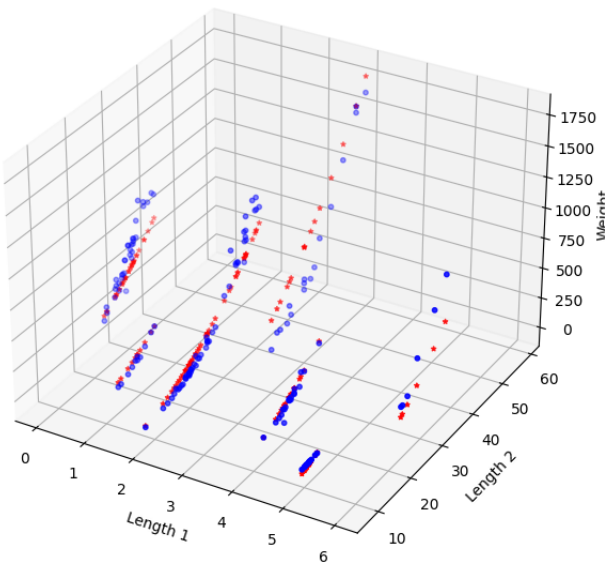
Grado 1 - MSE: 18769.926693773006

Grado 2 - MSE: 1.2213645589718399e+63

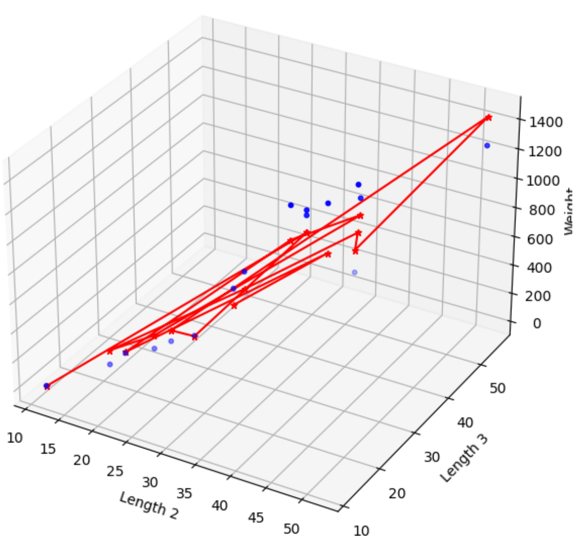
Grado 3 - MSE: 3316.3950467271084

Grado 4 - MSE: 7994.363094077428

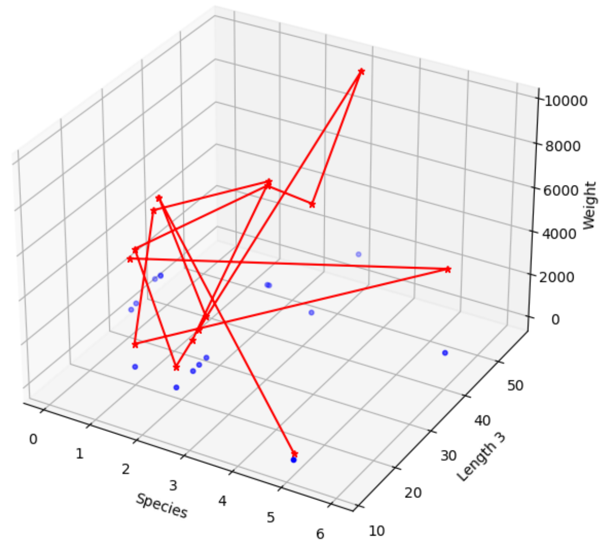
El de Menor Error es el Modelo de grado 2.



Longitudes 1 y 2 vs. Peso



Longitudes 2 y 3 vs. Peso



Especies, Longitud 3 vs. Peso

Referencias:

colaboradores de Wikipedia. (2023a, octubre 21). Regresión polinomial. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Regresi%C3%B3n_polinomial

Gonzalez, L. (2022, 7 septiembre). Regresión polinomial – teoría. 🤖 Aprende IA. https://aprendeia.com/algoritmo-regresion-polinomial-machine-learning/

Regresión polinomial. (2017, 8 mayo). WordPress.com. https://conzmr.wordpress.com/2017/04/04/regresion-polinomial/