Tarea 2 Minimos cuadrados

January 14, 2025

Integrantes: Juan David Clavijo Fernandez, Miguel Angel Fonseca Aldana, Richard Muñoz Henao

1 1. Teoría

1.0.1 Introducción

El **método de mínimos cuadrados** fue descubierto, y publicado por primera vez, por Adrien-Marie Legendre en 1805, aunque hay razones para creer que Gauss, quien lo publicó en 1809, fue la primera persona en descubrirlo (de acuerdo con Wikipedia, Gauss planteó los fundamentos del método en 1795).

El método de mínimos cuadrados se considera la primera aplicación de lo que posteriormente se denominó **análisis de regresión**. Legendre y Gauss aplicaron el método para determinar, a partir de observaciones astronómicas, las órbitas de los cuerpos alrededor del Sol (principalmente cometas, pero también otros cuerpos menores, como el recién descubierto planeta enano **Ceres**).

El análisis de regresión es a su vez un caso particular de un proceso conocido como **ajuste de curvas**. Dicho proceso consiste en encontrar una curva que contenga una serie de puntos y que posiblemente cumpla una serie de restricciones adicionales. Dos casos particulares del ajuste de curvas son la **interpolación** (cuando se espera un ajuste exacto a determinadas restricciones) y el análisis de regresión (cuando se permite una aproximación).

De acuerdo con Hoffman, para un conjunto pequeño y "suave" de datos es preferible usar métodos de interpolación. Ya que la tarea no trata sobre esto, no vamos a profundizar mucho al respecto. En cambio, para un conjunto grande y "áspero" de datos, es preferible usar métodos de ajuste aproximado. Profundicemos un poco sobre esto: dado un conjunto de datos con tales características, aún es posible utilizar interpoloación para obtener un ajuste exacto. Si es así, entonces ¿por qué llevar a cabo un ajuste aproximado? Existen varias razones para ello. Algunas de ellas son:

– Aunque hay métodos teóricos que aseguran que la interpolación es posible, en la práctica podríamos encontrar varios problemas que impedirían hallarla (por ejemplo: el coste computacional para hallar la solución podría ser muy alto). La solución en este caso es aceptar una aproximación.

– Quizá prefiramos el efecto de promediar datos cuestionables en una muestra, en lugar de distorsionar la curva para que se ajuste a ellos de forma exacta.

– Los polinomios de grado superior pueden oscilar mucho. Si hacemos pasar una curva por los puntos A y B, esperaríamos que la curva pase también cerca del punto medio entre A y B. Esto puede no suceder con curvas polinómicas de grados altos, ya que pueden tener valores de magnitud positiva o negativa muy grandes. Sobre esto, hay un ejemplo interesante en el Burden. Considere el problema

de calcular los valores de una función en puntos no tabulados, dados los datos experimentales en la siguiente tabla:

A continuación se muestra una gráfica con los valores de la tabla anterior:

A partir de esta gráfica, parece que la relación real entre x y y es lineal. La razón probable para que ninguna línea se ajuste con precisión a los datos son los errores en estos últimos. Por lo que es poco razonable solicitar una interpolación que concuerde exactamente con los datos. Si aún así nos inclinamos por la interpolación, nos encontraremos con oscilaciones que antes no estaban presentes. Por ejemplo, la gráfica del polinomio de interpolación de noveno grado para los datos de la tabla en cuestión se muestra a continuación:

Así pues, de forma muy general, podemos describir el método de mínimos cuadrados como un procedimiento para encontrar la curva que mejor se ajusta a un conjunto dado de puntos mediante el mecanismo que procura minimizar las diferencias entre las ordenadas de los puntos de la curva de la función elegida y los correspondientes valores en los datos. Dicho de una forma un poco más formal, el método de mínimos cuadrados es una técnica en la que, dados un conjunto de pares ordenados (variable independiente, variable dependiente) y una familia de funciones, se intenta encontrar la función continua, dentro de dicha familia, que mejor se aproxime a los datos (un "mejor ajuste"), de acuerdo con el criterio de **mínimo error cuadrático**. Vamos a profundizar en varios aspectos del método a medida que lo describimos desde las perspectivas algebraica, geométrica y estadística.

1.0.2 Perspectiva algebraica

El resultado de aplicar el método de mínimos cuadrados es la curva que "mejor" se ajusta a un conjunto dado de puntos. No hay una única forma de definir el "mejor ajuste". Consideremos el conjunto discreto de puntos $\{(x_i, Y_i(x_i))\}$ y un polinomio f que se ajusta de forma aproximada al conjunto de puntos en cuestión:

Las desviación de los puntos a la gráfica de la función se deben minimizar de alguna manera. Según Hoffman, hay varias formas de definir dicha desviación. Por ejemplo, si los valores de la variable independiente x_i se consideran exactos, la desviación es asignada a la variable dependiente Y_i, de modo pues que la desviación e_i es la distancia (dirigida) vertical entre Y_i y $y_i = f(x_i)$. Es decir:

$e_i = Y_i - y_i$.

También podemos considerar distancias horizonatales e incluso distancias perpendiculares (veáse la figura). Por lo general, se considera que la desviación está dada por la expresión descrita anteriormente para e_i. Así pues, asumimos en adelante que $e_i = Y_i = y_i$.

Ya tenemos claro cuáles son las desviaciones que queremos minimizar. Ahora debemos definir la "mejor" forma de hacerlo. Algunos criterios a tener en cuenta son:

– **Mimimizar la suma de las desviaciones**. Este enfoque tiene una desventaja inmediata: las desviaciones son distancias dirigidas, por lo que las desviaciones negativas se podrían compensar con las desviaciones positivas.

– **Minimizar la suma de los valores absolutos de las desviaciones**. Este enfoque presenta varias desventajas de naturaleza estadística y analítica.

– **Minimizar la distancia de la máxima desviación**. Este enfoque también presenta un desventaja inmediata: se le asigna demasiado "peso" a un solo dato.

El criterio del método de mínimos cuadrados consiste en minimizar el **error cuadrático**, que es simplemente la suma de los e_i^2. Éste criterio es un buen candidato para ser el "mejor" ajuste que queremos llevar a cabo, y hay varias razones que así lo demuestran. Por ejemplo, los e_i^2 son no negativos, lo que previene la "compensación" entre cantidades negativas y positivas. Además, los cuadrados "penalizan" las desviaciones grandes (las desviaciones grandes se harán más grandes mientras que las desviaciones pequeñas se harán más pequeñas), lo cual es de esperar en el ajuste de un modelo (se penaliza la desviación de varias datos, y no necesariamente la de uno solo).

Así pues, podemos definir el método de mínimos cuadrados de la siguiente manera: dado un conjunto de N puntos $\{(x_1, Y_1(x_1)), ..., (x_N, Y_N(x_N))\}$ y una familia de funciones F, el método de mínimos cuadrados encuentra la función $f \in F$ (que llamamos **función de aproximación**) que minimiza $\sum_{i=1}^{N} e_i^2$. Por lo general, se considera la familia de polinomios de un cierto grado n (entre otras cosas, por sus importantes propiedades analíticas y la facilidad para trabajar con ellos).

Sea F la familia de polinomios de grado 1. Los polinomios de dicha familia son de la forma $y = a+bx$, cuyas gráficas son líneas rectas. Consideremos un conjunto de N puntos $\{(x_1, Y_1), ..., (x_N, Y_N)\}$. La función de aproximación es:

$$y = ax + b$$

De momento, a y b son incógnitas. Ahora, evaluemos la función de aproximación en los x_i:

$$y_i = a + bx_i$$

Definimos las desviaciones de los datos de partida a la función de aproximación como es usual:

$$e_i = Y_i - y_i$$

A continuación, consideramos el **error cuadrático**:

$$S(a, b) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (Y_i - a - bx_i)^2$$

El error cuadrático alcanza un mínimo cuando $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = 0$. Es decir, cuando:

– $2\sum_{i=1}^{N}(Y_i - a - bx_i)(-1) = 0$.

– $2\sum_{i=1}^{N}(Y_i - a - bx_i)(-x_i) = 0$.

LLevando a cabo algunas operaciones adicionales, obtenemos las **ecuaciones normales** del ajuste por mínimos cuadrados:

1. $aN + b\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} Y_i$.

2. $a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} Y_i x_i$.

Finalmente, para hallar a y b, los parámetros que caracterizan a la función de aproximación, utilizamos cualquier método para resolver sistemas de ecuaciones lineales.

Es bastante común, e increíblemente útil, obtener la función de aproximación de la familia de polinomios de grado 1. Sin embargo, también podríamos considerar familias de polinomios de grado superior. El proceso para derivar las ecuaciones normales en el caso general de un polinomio de grado n como función de aproximación es bastante similiar, y preferimos omitirlo para mantener las cosas simples (y porque, en este punto, creemos que los principios del método están bastante claros).

Terminamos mencionando un par de cuestiones adicionales:

– Las ideas que descrbimos antes y los procedimientos que llevamos a cabo se pueden aplicar de forma análoga al caso de funciones de varias variables y polinomios multivariables.

– Hasta el momento hemos hablado solamente de utilizar polinomios (funciones lineales) como funciones de aproximación. Sin embargo, la naturaleza de algunos problemas sugiere el uso de funciones de aproximación no lineales. Al aplicar los principios del método de mínimos cuadrados para derivar las ecuaciones normales, eventualmente tendremos la necesidad de resolver un sistema de ecuaciones no lineales, con la dificultad que ello conlleva. Una posible solución es aproximar la solución del problema mediante la solución de un problema lineal. Sin embargo, de acuerdo con Burden, esta solución está lejos de ser la mejor. Y aunque fuera una solución aceptable, es simplemente imposible aproximar todos los problemas no lineales mediante problemas lineales. La mejor forma de abordar el problema es con métodos para resolver sistemas de ecuaciones no lineales.

1.0.3 Perspectiva geométrica

¿Qué es una proyección?: Consideremos un vector en $\vec{v} \in \mathbb{R}^2$. Dicho vector determina una recta l, que será el conjunto de vectores en $\mathbb{R}^2$ que son múltiplos escalares de $\vec{v}$. A continuación, consideremos un vector $\vec{x} \in \mathbb{R}^2$. En la siguiente representación, $\overline{x}$ no es un múltiplo escalar de $\vec{x}$:

Queremos encontrar el punto $\overline{x}$ sobre l que se encuentre más cerca de $\vec{x}$ (en este contexto medimos distancias con la métrica estándar). Como el punto en cuestión $\overline{x}$ se encuentra sobre l, y lo podemos localizar mediante un vector, resulta que $\overline{x} = c\vec{v}$, para algún $c \in \mathbb{R}$.

Podemos medir la distancia entre $\overline{x}$ y $\vec{x}$ como $||\overline{x} - \vec{x}|| = ||c\vec{v} - \vec{x}||$. Como $\vec{v}$ y $\vec{x}$ son conocidos, solo falta determinar c para que la distancia $||c\vec{v} - \vec{x}||$ sea mínima. Y como la función raíz cuadrada "respeta" desigualdades ($a \leq b$, si y sólo si, $\sqrt{a} \leq \sqrt{b}$, donde a y b son reales no negativos), la

cuestión es equivalente a minimizar $||c\vec{v} - \vec{x}||^2$. Ahora, observe que $||c\vec{v} - \vec{x}||^2$ es una función de c. Por lo tanto, dicha cantidad será mínima cuando $\frac{d}{dc}||c\vec{v} - \vec{x}||^2 = 0$. Es decir, cuando:

$$\sum_{i=1}^{2} v_i(cv_i - x_i) = 0$$

Pero:

$$\begin{aligned}\sum_{i=1}^{2} v_i(cv_i - x_i) &= \sum_{i=1}^{2}(cv_i^2 - x_i v_i) \\ &= c\sum_{i=1}^{2} v_i^2 - \sum_{i=1}^{2} x_i v_i \\ &= c(\vec{v}\cdot\vec{v}) - (\vec{v}\cdot\vec{x}) \\ &= c\vec{v}^T\vec{v} - \vec{v}^T\vec{x}\end{aligned}$$

Como $\overline{x} = c\vec{v}$, tenemos entonces que $||c\vec{v} - \vec{x}||$ es mínimo cuando $\vec{v}^T(\overline{x} - \vec{x}) = \vec{v}\cdot(\overline{x} - \vec{x}) = 0$; ésto es, cuando $\overline{x} - \vec{x}$ es perpendicular a $\vec{v}$. Encontremos una expresión explícita para c:

$$\begin{aligned}c\vec{v}^T\vec{v} - \vec{v}^T\vec{x} &= 0 \\ c\vec{v}^T\vec{v} &= \vec{v}^T\vec{x} \\ c &= (\vec{v}^T\vec{v})^{-1}\vec{v}^T\vec{x}\end{aligned}$$

Ahora el vector $\overline{x}$ es conocido. Dicho vector es la **proyección** de $\vec{x}$ sobre $\vec{v}$ (o, mejor dicho, la recta que determina $\vec{v}$). Observe que:

$$\begin{aligned}\overline{x} &= \vec{v}c \\ &= \vec{v}(\vec{v}^T\vec{v})^{-1}\vec{v}^T\vec{x} \\ &= P\vec{x}\end{aligned}$$

donde $P = \vec{v}(\vec{v}^T\vec{v})^{-1}\vec{v}^T$ es una matriz. Hasta ahora hemos llevado a cabo la explicación en $\mathbb{R}^2$. Siguiendo la misma idea y un procedimiento similar llegamos a la misma expresión para P en el caso general de $\mathbb{R}^n$. Dicha matriz es importante porque nos permite hallar la proyección de cualquier vector $\vec{x}$ sobre $\vec{v}$.

En $\mathbb{R}^2$ y $\mathbb{R}^3$ la noción de proyección es bastante intuitiva. Pero definir dicha noción en general puede ser un poco complicado. Por suerte, el enfoque que acabamos de describir proporciona un punto de vista abstracto, pero bastante natural también, que nos permite definir la noción de proyección fácilmente: la proyección de $\vec{x}$ sobre $\vec{v}$ es el vector $\overline{x}$ paralelo a $\vec{v}$ (que se encuentra sobre la recta determinada por $\vec{v}$) tal que la distancia $||\overline{x} - \vec{x}||$ es mínima (es decir que si consideramos otro vector $\vec{u}$ paralelo a $\vec{v}$, encontraremos que $||\overline{x} - \vec{x}|| \leq ||\vec{u} - \vec{x}||$).

Relación con el método de mínimos cuadrados: Consideramos un conjunto de N datos $\{(x_1, Y_1), ..., (X_N, Y_N)\}$, junto con la familia F de polinomios de grado 1. Vamos a aplicar el método de mínimos cuadrados para hallar una función de aproximación a los datos en cuestión. Por simplicidad, supongamos que la función de aproximación es de la forma $y = ax$ (la recta

pasa por el origen). Consideremos $\vec{x} = (x_1, ..., x_N)$, $\vec{Y} = (Y_1, ..., Y_N)$ y $\vec{y} = (y_1, ..., y_n)$, donde $y_i = ax_i$. Teniendo en cuenta esta notación, resulta un vector de desviación: $e = (e_1, ..., e_n) = (Y_1 - y_1, ..., Y_n - y_n) = \vec{Y} - \vec{y}$. De acuerdo con lo anterior, al aplicar el método de mínimos cuadrados encontraremos un vector $\vec{y}$ paralelo a $\vec{x}$ tal que la distancia $||\vec{Y} - \vec{y}||$ es mínima. Todo esto ya debería ser famliar: en otras palabras, estamos proyectando $\vec{Y}$ sobre $\vec{x}$.

Imágenes tomadas de: https://medium.com/@vladimirmikulik/why-linear-regression-is-a-projection-407d89fd9e3a

1.0.4 Perspectiva estadística

Como se mencionó en la introducción, el método de mínimos cuadrados es un tipo de **análisis de regresión**. En pocas palabras, un análisis de regresión es un proceso mediante el cual se espera entender cómo una variable depende de otra variable.

En 1801, Gauss utilizó el método para predecir la órbita de Ceres a partir de los datos de la misma tomados por Giuseppe Piazzi a lo largo de cuarenta días. La predicción era necesaria porque, tras los mencionados cuarenta días de observación, Ceres desapareció detrás del brillo del sol, lo que impidió seguir rastreando su posición. Los métodos de la época requerían una mayor cantidad de datos para calcular la órbita de cualquier cuerpo celeste, por lo que fue imposible para la mayoría de Astronomos predecir con éxito la posición de Ceres en el firmamento. Fue Gauss, con el método de mínimos cuadrados, quien logró finalmente estimar la órbita de Ceres a partir de los pocos datos tomados, y con ésto predecir la posición en la que Ceres sería visible de nuevo.

El ejemplo anterior ilustra a la perfección la necesidad del análisis de regresión: ¿Por qué necesitamos saber cómo depende una variable de otra? Porque entender dicha relación nos permite hacer **predicciones** y **previsiones**.

En términos simples, el resultado de una regresión es una curva (el gráfico de una función llamada **función de regresión**) que se ajusta a los puntos de un conjunto de datos, mostrando la relación entre dos variables (aunque el análisis de regresión incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes o **predictoras**).

Muchas técnicas han sido desarrolladas para llevar a cabo el análisis de regresión. Una de ellas es, como se viene diciendo, el método de mínimos cuadrados. Dentro del análisis de regresión, el método de mínimos cuadrados se considera un método **paramétrico**, ya que la función de regresión se define en términos de un número finito de parámetros desconocidos que se estiman a partir de los datos. También existen métodos **no paramétricos**, llamados así porque permiten que la función de regresión consista en un conjunto específico de funciones, que puede ser de dimensión infinita. Dentro de los métodos no paramétricos encontramos un análogo del método de mínimos cuadrados (Burden distingue entre **mínimos cuadrados discretos**, lo que para nosotros es simplemente el método de mínimos cuadrados, y **mínimos cuadrados**, el análogo no paramétrico del método anterior).

El **teorema de Gauss-Markov** demuestra que el método de mínimos cuadrados es óptimo en muchos sentidos. Ésto, junto con su simplicidad, hace del método una herramienta fundamental en el análisis de regresión. Sin embargo, el desempeño del método en la práctica depende de la forma del proceso de generación de datos, y cómo se relaciona con el método en sí. Hay toda una teoría estadística dedicada a la formulación y correcta implementación de métodos para el análisis de regresión. Nosotros preferimos parar la discusión aquí.

Algunas aplicaciones del método de mínimos cuadrados en el contexto de la ciencia de datos son:

–**Regresión lineal**: Es una de las aplicaciones más directas. Por ejemplo: predecir el precio de una vivienda en función del tamaño, ubicación, y número de habitaciones.

– **Análisis de series temporales**: Se usa para ajustar tendencias o modelos lineales a datos históricos, lo cual permite capturar patrones subyacentes y predecir valores futuros. Por ejemplo: modelar la evolución del tráfico web de una página o las ventas mensuales de una empresa.

– **Análisis de componentes principales (PCA)**: PCA utiliza el método de mínimos cuadrados para encontrar las direcciones de mayor variabilidad en los datos, reduciendo dimensionalidad mientras se conserva la mayor parte de la información. Por ejemplo: reducir las dimensiones de una base de datos con muchas variables (como datos de clientes) para simplificar el análisis.

– **Sistemas de recomendación**: El método de mínimos cuadrados se usa en factorización matricial para descomponer matrices de usuarios y productos, encontrando patrones de recomendación. Por ejemplo: Netflix usa estos modelos para predecir qué películas te podrían gustar basándose en tus calificaciones anteriores.

–**Machine Learning**: Los mínimos cuadrados están en el corazón de muchos algoritmos supervisados. Se optimizan modelos ajustando los parámetros para minimizar el error cuadrático medio. Por ejemplo: entrenamiento de regresión lineal.

2 2. Aplicaciones

2.1 Descripción del problema

Contamos con un conjunto de datos que contiene información sobre el desempeño académico de estudiantes, así como sobre sus hábitos de estudio e historial académico. El objetivo es estimar el desempeño académico de un estudiante teniendo en cuenta sus hábitos de estudio y su historial académico.

2.2 Dataset

Utilizamos el conjunto de datos “Student_Performance.csv”, obtenido de la página web de Kaggle.

2.3 Metodología

En primer lugar, creamos un mapa de calor de la matriz de correlación para visualizar de manera más clara las relaciones lineales entre las diferentes variables. Además, generamos un gráfico de dispersión por pares para observar con mayor detalle dichas relaciones.

Posteriormente, dividimos el conjunto de datos en dos partes: una para entrenar el modelo de regresión lineal y otra para evaluar su desempeño. Entrenamos el modelo utilizando los datos de entrenamiento, realizamos una prueba con los datos de prueba y, finalmente, presentamos la información obtenida junto con un gráfico que ilustra los resultados de la evaluación del modelo.

2.4 Implementacion y visualizacion

[77]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('Student_Performance.csv')

df.head(20)
```

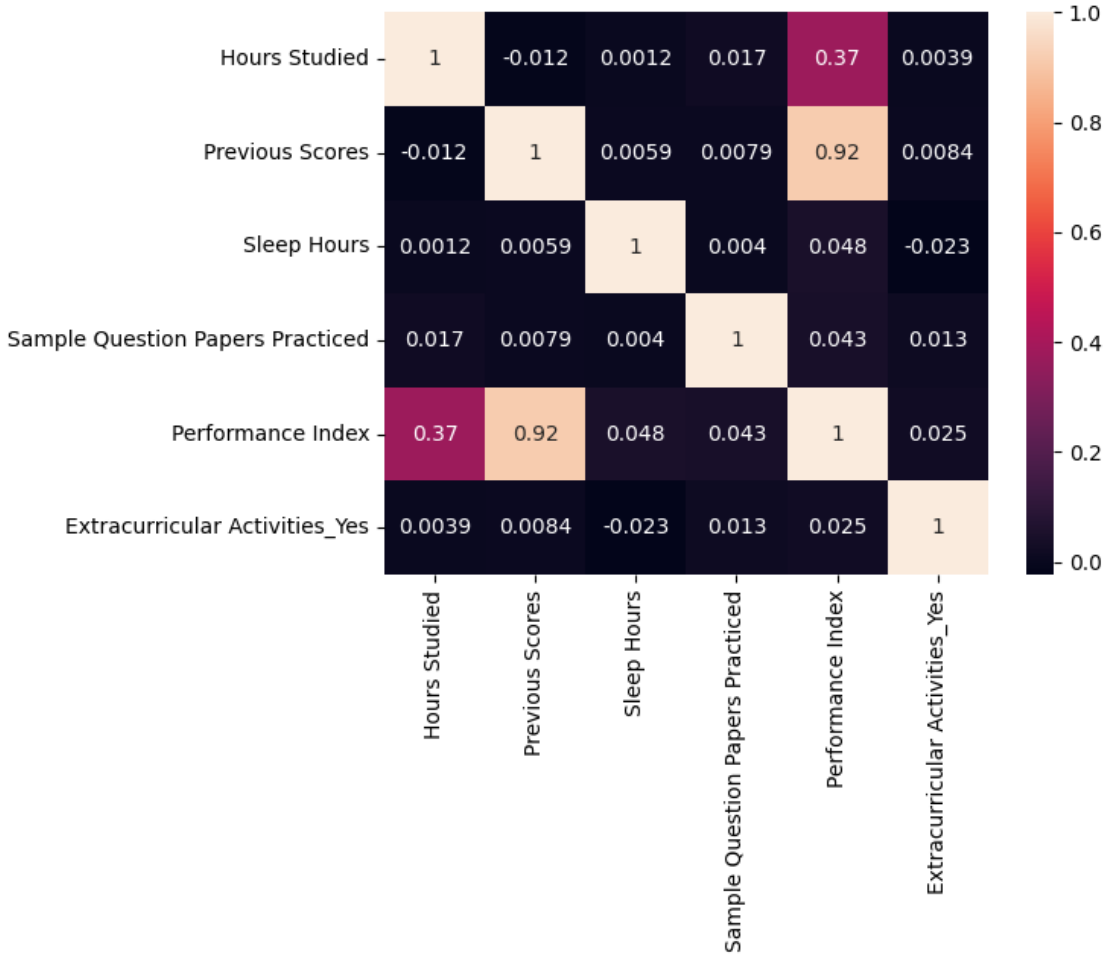
[77]:

	Hours Studied	Previous Scores	Extracurricular Activities	Sleep Hours	Sample Question Papers Practiced	Performance Index
0	7	99	Yes	9	1	91.0
1	4	82	No	4	2	65.0
2	8	51	Yes	7	2	45.0
3	5	52	Yes	5	2	36.0
4	7	75	No	8	5	66.0
5	3	78	No	9	6	61.0
6	7	73	Yes	5	6	63.0
7	8	45	Yes	4	6	42.0
8	5	77	No	8	2	61.0
9	4	89	No	4	0	69.0
10	8	91	No	4		
11	8	79	No	6		
12	3	47	No	9		
13	6	47	No	4		
14	5	79	No	7		
15	2	72	No	4		
16	8	73	Yes	8		
17	6	83	Yes	7		
18	2	54	Yes	4		
19	5	75	No	7		

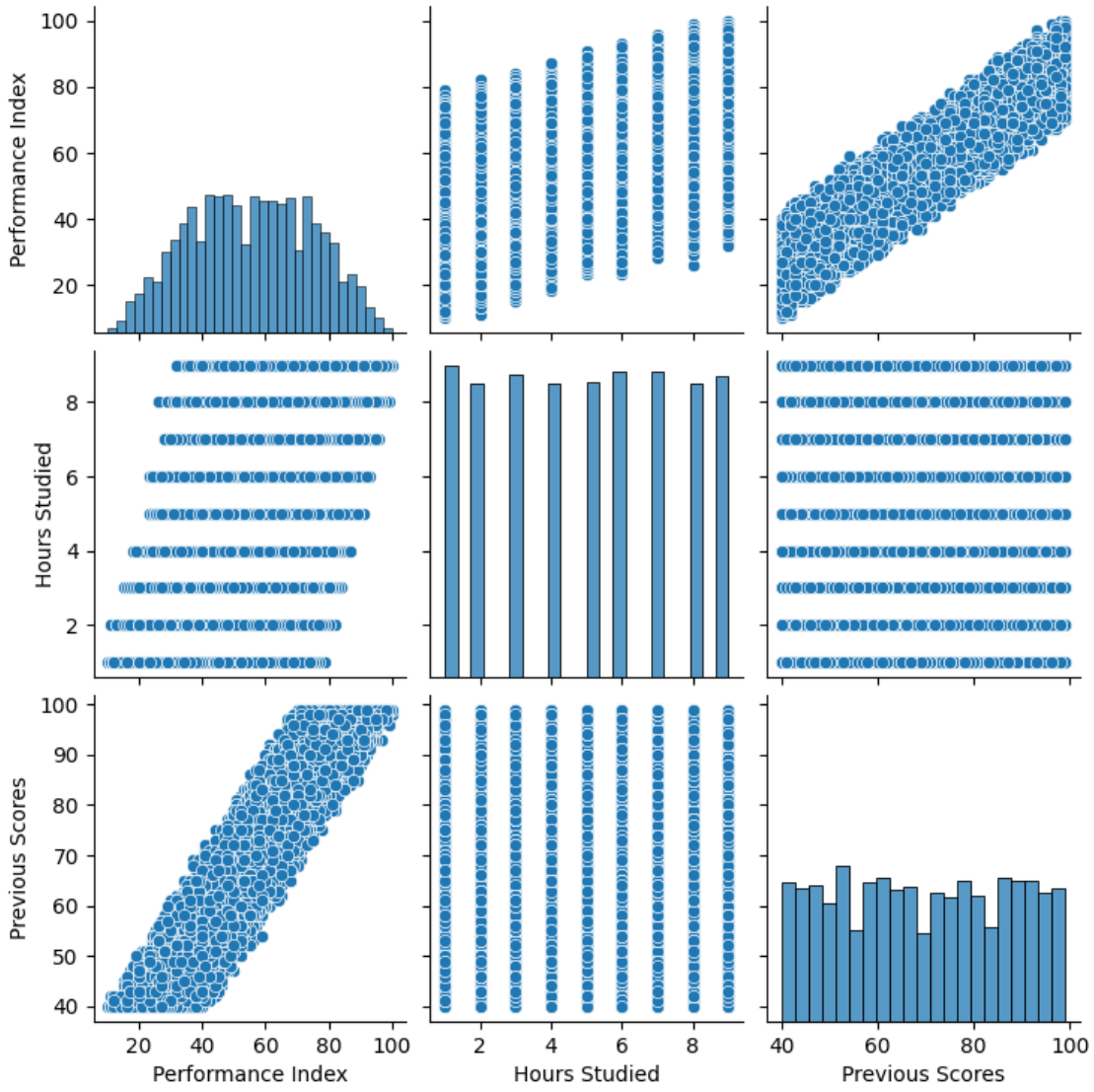
10	5	84.0
11	2	73.0
12	2	27.0
13	2	33.0
14	8	68.0
15	3	43.0
16	4	67.0
17	2	70.0
18	9	30.0
19	0	63.0

[80]:
```
df = pd.get_dummies(df, drop_first=True) # Convierte las variables categóricas del DataFrame en variables dummy (o variables indicadoras) ('Extracurricular Activities')
```

[83]:
```
df_corr = df.corr(method = 'pearson') # Calculo de la matriz de correlacion
sns.heatmap(df_corr, annot=True) # Mapa de calor de la matriz de correlacion
plt.show() #Mostrar el grafico
```



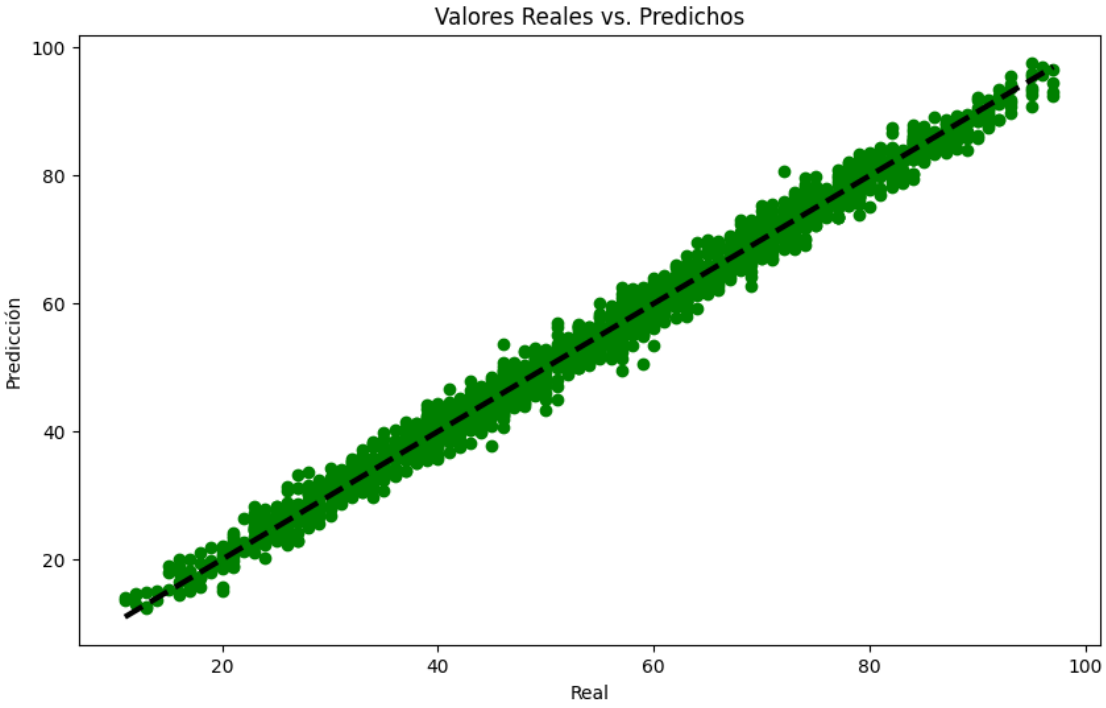
[87]:
```
sns.pairplot(data=df[['Performance Index','Hours Studied','Previous Scores']])␣
↪# Grafico por pares para visualizar la relacion entre algunas de las␣
↪caracteristicas que mostraron mayor correlacion
plt.show() # Mostrar el grafico
```



[102]:
```
print('-----------------------------------------------------')
# Definir la variable objetivo y las características
X = df.drop('Performance Index', axis=1) # Caracteristicas (Todas las columnas␣
↪excepto performance index)
y = df['Performance Index'] # Variable objetivo (Performance index)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.
 ↪8,random_state=450) # Dividir los datos en conjunto de entrenamiento y prueba
model = LinearRegression() # Crear instancia del modelo de regresion lineal
model.fit(X_train, y_train) # Entrenar instancia del modelo de regresion lineal
y_pred = model.predict(X_test) # Predecir los valores para el conjunto de prueba

# Calcular métricas de rendimiento
mse = mean_squared_error(y_test, y_pred) #Calcular el Error Cuadrático Medio
r2 = r2_score(y_test, y_pred) #Calcular el Coeficiente de Determinación
print(f"Error cuadratico medio: {mse}")
print(f"Coeficiente de determinacion: {r2}")
coeficientes = pd.DataFrame(model.coef_, X.columns, columns=['Coeficientes']) #␣
 ↪Mostrar los coeficientes del modelo
print(coeficientes)
print('-------------------------------------------------------')

# Gráfico de dispersión de los datos reales vs. los predichos
#plt.scatter(y_test, y_pred)



# Gráfico de valores reales vs. predichos
plt.figure(figsize=(10,6))
plt.scatter(y_test, y_pred, color='green')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--',␣
 ↪lw=3)
plt.xlabel('Real')
plt.ylabel('Predicción')
plt.title('Valores Reales vs. Predichos')
plt.show()
```

```
-------------------------------------------------------
Error cuadratico medio: 4.30690710021039
Coeficiente de determinacion: 0.9880309822769251
                                  Coeficientes
Hours Studied                         2.855297
Previous Scores                       1.017911
Sleep Hours                           0.482081
Sample Question Papers Practiced      0.202612
Extracurricular Activities_Yes        0.600215
-------------------------------------------------------
```



3 Análisis de los resultados

Al observar el mapa de calor (primer gráfico), podemos ver que existe una alta correlación de Pearson entre el "Performance Index" y algunas variables. Recordemos que la correlación de Pearson es una medida estadística que indica la fuerza y la dirección de la relación lineal entre dos variables. Esto sugiere que podría ser factible realizar una predicción utilizando un modelo de regresión lineal para este problema.

Para visualizar mejor esa correlación, se generó un gráfico de dispersión por pares, el cual valida la hipótesis anterior, mostrando una relación lineal entre el "Performance Index", las horas estudiadas y los resultados anteriores.

Después de realizar la prueba con el modelo de regresión lineal entrenado, observamos que el coeficiente de determinación es alto (0.98) y el error cuadrático medio es bastante bajo (4.3), lo cual indica que los datos se ajustan bien al modelo de regresión.

Esto también se confirma en el gráfico de comparación entre los valores reales y los valores predichos, donde es evidente que las predicciones son bastante acertadas en comparación con los valores reales del dataset.