F210 PROGRAMACION APLICADA A FINANZAS

Interpolación polinómica a un conjunto de datos

Regresion lineal

1. Desarrolle una subclase de poly `linreg` que sirva para encontrar la "mejor" recta que interpola linealmente una colección de N puntos (x_i, y_i), $1 \leq i \leq N$, correspondiente a un diagrama de dispersión.

   Cada instancia de clase de `linreg` debe contener – además de los atributos de poly– los siguientes atributos específicos:

   (a) Una lista `datos` con los pares ordenados (x_i, y_i) que se interpolan

   (b) Un escalar `beta_MCO` que indica el valor de la pendiente que se obtiene de resolver el problema mínimos cuadrados ordinarios mediante la expresión

   $$\beta_{MCO} = \frac{\sum_{i=1}^{N} \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum_{i=1}^{N} \left(x_i - \overline{X}\right)^2}$$

   donde $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ y $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$

   (c) Un escalar `alpha` que indica la pendiente de la recta interpolante, y que se computa mediante la expresión

   $$\alpha = \overline{Y} - \beta_{MCO}\overline{X}$$

   (d) Un `__init__` method que reciba una lista de tuplas de la forma (x_i, y_i) con x_i la variable que llamamos independiente o exógena e y_i la variable dependiente o endógena, y con ella inicialice los atributos de la instancia de clase.

   (e) Un método `__str__` que permita imprimir la ecuación de la recta interpolante.

   (f) Un método `regplot(self)` que haga el diagrama de dispersión de los pares ordenados (x_i, y_i), grafique superpuesta la recta interpolante óptima, junto con la ecuación de dicha recta. (importe y use la librería matplotlib o seaborn)

   (g) Un método `NR_reg(self)` que a partir de `datos` y la condición encontrada en clase que la recta óptima debe pasar por el punto $\left(\overline{X}, \overline{Y}\right)$ encuentre el valor de b que minimice la función objetivo $L(b)$ consistente en la suma de los cuadrados de los errores de interpolación

   $$\begin{aligned} L(b) &= \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \left(y_i - (a + bx_i)\right)^2 \\ &= \sum_{i=1}^{N} \left(y_i - \left(\overline{Y} - b\overline{X} + bx_i\right)\right)^2 \\ &= \sum_{i=1}^{N} \left(\left(y_i - \overline{Y}\right) - b\left(x_i - \overline{X}\right)\right)^2 \end{aligned}$$

Use la regla de derivación numérica centrada que vimos en clase para buscar un cero de $L'(b)$ mediante el algoritmo de Newton Raphson que ya hemos desarrollado en clase. Compare el valor de b resultante con el β_{MCO} del punto b.

2. Muestre matemáticamente que para cualquier conjunto de datos (x_i, y_i) se verifica que toda recta de la forma $y = a + bx$ que pase por el punto promedio $\left(\overline{X}, \overline{Y}\right)$ satisface que la suma de los errores de interpolación e_i generados por ella, con $e_i = y_i - \widehat{y}_i = y_i - (a + bx_i)$, debe ser igual a cero.

3. Muestre matemáticamente que la pendiente óptima que resuelve el problema de interpolación

$$\beta_{MCO} = \frac{\displaystyle\sum_{i=1}^{N} \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\displaystyle\sum_{i=1}^{N} \left(x_i - \overline{X}\right)^2}$$

se puede obtener también como el promedio ponderado

$$\beta_{MCO} = \sum_{i=1}^{N} \beta_i \, \omega_i$$

de las pendientes β_i de las rectas que forman cada dato muestral (x_i, y_i) con el punto medio $\left(\overline{X}, \overline{Y}\right)$, usando como ponderador ω_i un escalar proporcional al cuadrado de la distancia entre x_i y su promedio, $\left(x_i - \overline{X}\right)^2$. Llamaremos a ω_i *el peso de la pendiente* β_i *en* β_{MCO}.

4. Lea la documentacion del sitio `https://pypi.org/project/yfinance/` sobre la librería yfinance para descargar datos financieros de yahoo finance. (Note el enfoque de programación orientada a objetos, aproveche a explorar que hacen las clases Ticker y Tickers) Descargue usando la librería los precios ajustados de cierre mensuales $P_{i,t}$ para el periodo 2000-2024 de las siguientes compañías: Coca Cola, Microsoft, Chevron, Johnson & Johnson y Ford Motors. Asimismo, descargue también los precios de cierre ajustados del índice SP500 (ticker "^GSPC") para igual periodo. En la notacion $P_{i,t}$ i representa a la empresa y t al tiempo.

   Compute para cada activo los *retornos precio* mensuales $R_{i,t}$ como la variacion porcentual del precio ajustado

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}$$

   Use matplotib o seaborn y realice los siguientes gráficos:

   (a) Normalice en 100 el precio de cierre ajustado para los distintos activos al 2000-1-1, y haga una grafica conjunta de sus distintos precios de cierre mensuales a lo largo del tiempo. Indique de la gráfica que rendimiento tuvieron los distintos papeles y si le ganaron o no al índice S&P en dicho periodo

(b) Haga gráficos individuales para los retornos de cada activo, junto con el índice S&P como función del tiempo. Vea si (a simple vista) la volatilidad de los retornos es mayor o menor que la del índice S&P

(c) Haga un gráfico de dispersión para cada uno de los activos, con los retornos mensuales del índice S&P en el eje horizontal y los del activo en el eje vertical. Encuentre la recta de regresión lineal que mejor aproxima a los puntos para cada activo, asumiendo que el modelo entre los retornos del activo y del índice es de la forma

$$R_{i,t} = a_i + b_i R_{S\&P,t}$$

En base a los valores de b_i encontrados (la beta) para cada activo, indique cuáles son más volátiles que el índice S&P y cuáles son menos volátiles. Explique intuitivamente su respuesta. Compare sus conclusiones con las del punto b.

(d) En base al problema 3, explique intuitivamente porqué datos de periodos de crisis como la pandemia COVID, o la crisis subprime tendrán una mayor influencia/peso que otros datos históricos a la hora de determinar el beta β_{MCO} del activo. Ordene los retornos del S&P en orden decreciente segun el peso ω_i de cada dato, e identifique para qué periodos se observan los pesos mayores.

(e) Subdivida la muestra en periodos de 5 años, y mida para cada activo la β_{MCO} correspondiente a cada periodo. Indique si la beta se mantiene relativamente constante o no en el tiempo.