Data Science for weather forecast with Phyton MMCCD 2020

Alfonso Barajas Cervantes

Data Science, IIMAS

Keywords: Weather, Polynomial, Approximation, Least Squares Method.

Abstract

Una de las mayores ventajas de aproximar información discreta o funciones complejas con funciones analíticas sencillas, radica en su mayor facilidad de evaluación y manipulación. En este caso se utiliza el método de la aproximación *polinomial* de grado n. Esto con el fin de facilitar la obtención de una fórmula polinomial que mejor se aproxime a pronosticar la temperatura dado el mes de uno o varios estados de la República Mexicana. Además se realizará un pequeño análisis por zona geográfica.

1 Introducción

En el caso de ajuste de ajuste de polinomios o de aproximación polinómica, tenemos que dado un conjunto de datos $(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)$ nos proponemos a encontrar un polinomio $p(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$, con $n < m$ de manera que esté más cerca de los puntos reales. Es decir, que la distancia entre los pronosticados usando el polinomio y el dato real sea la mínima. Es decir, que

$$\sum_{i=1}^{m} |p(x_i) - y_i|^2$$

sea mínima. Además se hará una aproximación de grado $n = 3$, donde se podrá notar una aproximación con ayuda de la siguiente resolución de sistema de matrices.

$$\begin{aligned} AX &= B \\ A^T(AX) &= A^T B \\ (LL^T)X &= A^T B \\ \rightarrow LY &= A^T B \\ L^T X &= Y \end{aligned}$$

2 Segmentación y Delimitación de Tiempo

Sucede que al momento de decidir los intervalos de tiempo para realizar el análisis fue ciertamente algo complicado de elegir. Esto en el sentido para ver nuestro intervalo de tiempo, ya que al observar detalladamente las muestras de cada año, parecen ser periódicas. Por lo tanto, tomaremos al Estado de Aguascalientes (AGS) con los siguientes intervalos de tiempo: Dado que tenemos que las fechas de las estaciones de año en México son las siguientes:

- Primavera: 21 de marzo a 20 de junio.
- Verano: 21 de junio a 21 de septiembre.
- Otoño: 22 de septiembre a 20 de diciembre
- Invierno: 21 de diciembre a 20 de marzo

De esta manera, lo que se decidio a realizar dado que los datos de la temperatura están por meses, esto significa que nuestra segmentación por estaciones del año son las siguientes:

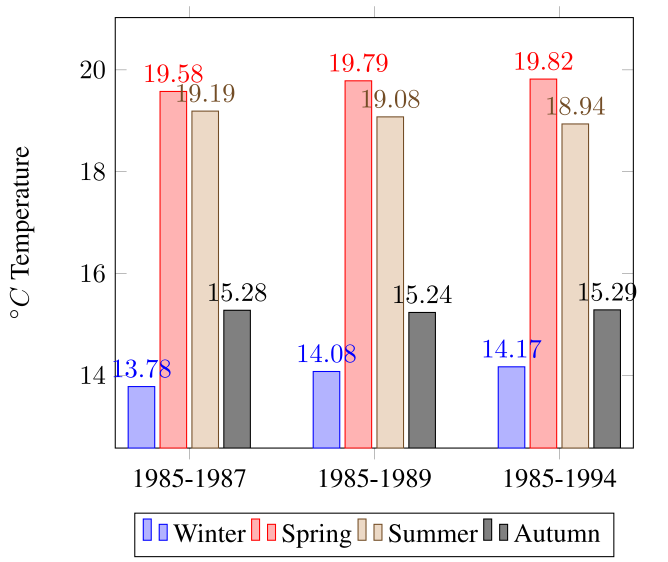
- Primavera: abril a junio
- Verano: julio a septiembre
- Otoño: octubre a diciembre
- Invierno: enero a marzo

Por lo tanto, se realizarán análisis primero por cada estación en 3 años, 5 años y 10 años para poder encontrar así cual polinomio es el que mejor puede llegar a predecir a los siguientes y anteriores años en sus respectivas categorías. De manera tal, que al final se realizarán las respectivas comparaciones con las funciones predictoras.

3 Aproximación polinomial y resultados

3.1 Segmención: Estacioón de Año

Se obtuvieron los siguientes resultados al realizar un análisis de las temperaturas de Aguascalientes con respecto a sus temperaturas respectivamente por estación de año en $1985 - 1987, 1985 - 1989, 1985 - 1994$. Por lo tanto, primero notamos la temperatura media en cada estación de año.



Ahora bien, se mostrarán los valores respectivos de los coeficientes de la aproximación polinomial. El cual, recordemos lo hicimos por segmentación de acuerdo a las estaciones del año.

Winter				
Years/Coeff.	a (X^3)	b (X^2)	c (X)	d (ind)
1985-1987	0.083	-1.027	3.240	12.150
1985-1989	0.004	-0.066	0.316	13.484
1985-1994	0.001	-0.0260	0.297	13.225

Spring				
Years/Coeff.	a (X^3)	b (X^2)	c (X)	d (ind)
1985-1987	0.066	-0.815	2.617	18.095
1985-1989	0.005	-0.110	0.582	18.939
1985-1994	0.0004	-0.020	0.261	18.984

Summer				
Years/Coeff.	a (X^3)	b (X^2)	c (X)	d (ind)
1985-1987	0.002	–0.025	0.319	18.216
1985-1989	-0.001	-0.016	0.350	18.177
1985-1994	0.001	-0.0228	0.243	18.503

Autumn				
Years/Coeff.	a (X^3)	b (X^2)	c (X)	d (ind)
1985-1987	-0.050	0.653	-2.406	17.339
1985-1989	-0.0135	0.272	-1.482	17.078
1985-1994	0.0001	0.0019	-0.121	15.928

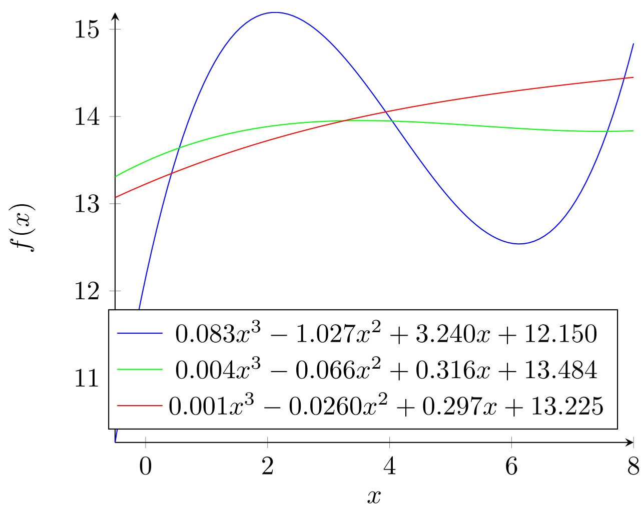
Podemos notar, que en un año de los datos pero segmentados en cuatro particiones, con respecto a las estaciones del año, se tiene que no llega a ser tan precisa como las que le siguen, y esto se debe a la cantidad de datos que se introdujeron en para conocer su aproximación polinomial.

Además se puede llegar a notar, que el índice del término cúbico es casi nulo. indicando que casi no es "necesario" para el polinomio. Agregando que el termino independiente se acerca demasiado a su valor promedio de Temperatura respectivo.

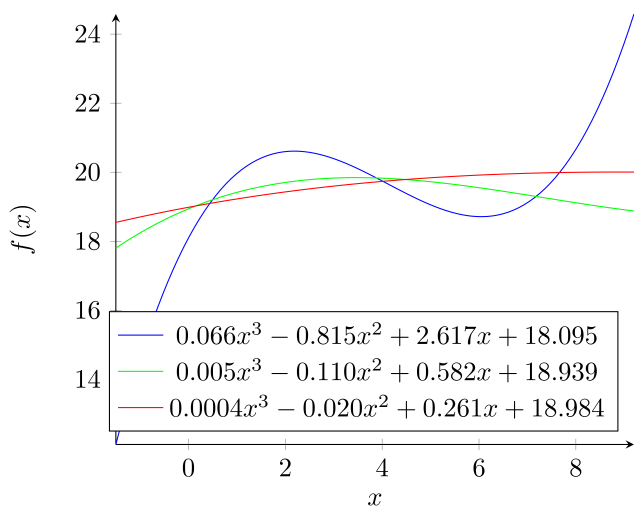
3.2 Plots

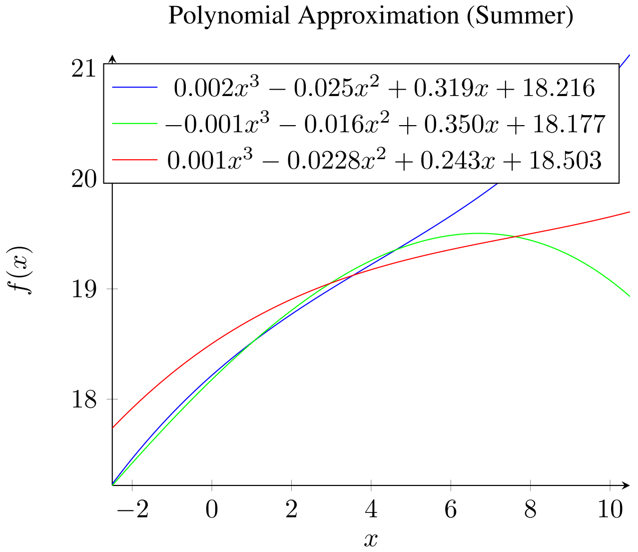
A continuación se muestran los resultados en términos de gráficas comparando cada una entre sí con sus respectivas épocas del año. La siguiente gráfica muestra la aproximación con respecto a la estación de Invierno hasta la de Otoño.

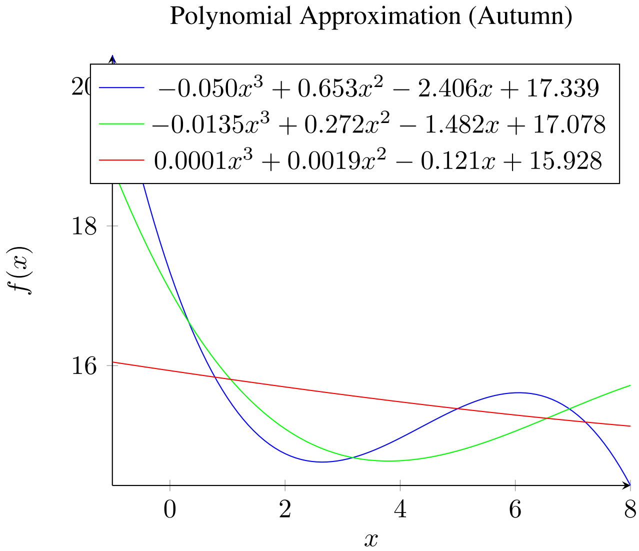
Polynomial Approximation (Winter)



Polynomial Approximation (Spring)







3.3 Interpretación

A continuación exprondremos lo que se puede observar en las gráficas propocionadas anteriormente, que justo era lo que se esperaba al momento de elegir uno de los estados con temperaturas más cálidas de México, Aguascalientes.

Lo que se puede observar en las tendencias desde el comienzo del año en Aguascalientes, se tiene que tomar en cuenta que desde Enero a Marzo, tiene una temperatura entre los 12 °C y los 15°C. Pero al incrementar los meses, es decir, conforme pasan los meses, va incrementando para que justo cuando sea Abril, en primavera nos encontramos, su temperatura subió hasta los 18 °C - 20 °C. Lo cual se mantiene en el verano. Sin embargo, al pasar al otoño, se reduce drásticamente la temperatura hasta llegar a los 14°C - 16°C

4 Calculo de Mínimos Cudrados

A continuación, se muestran los resultados de calcular los mínimos cuadrados de la función predictora con respecto a los datos colocados.

Winter	
Years/Coeff.	$\sum_{i=1}^{m}(f(x_i) - y_i)^2$
1985-1987	13.036
1985-1989	29.92
1985-1994	72.059

Spring	
Years/Coeff.	$\sum_{i=1}^{m}(f(x_i) - y_i)^2$
1985-1987	5.909
1985-1989	18.596
1985-1994	50.956

Summer	
Years/Coeff.	$\sum_{i=1}^{m}(f(x_i) - y_i)^2$
1985-1987	0.505
1985-1989	3.325
1985-1994	11.885

Autumn	
Years/Coeff.	$\sum_{i=1}^{m}(f(x_i) - y_i)^2$
1985-1987	14.875
1985-1989	29.574
1985-1994	73.961

4.1 Interpretación

Lo que se obtuvo en este análisis de temperaturas fue realmente sorprendente. Al contrario de lo que uno esperaría encontrarse, mientras más datos se coloquen, peor predice. Esto se puede notar al tomar en cuenta la función cuya distancia a cada punto predicho sea la mínima. Sin embargo, se tiene que los mejores fueron los que obtuvieron muy pocos datos de alrededor 3 años.

Ahora bien, también se debe de tomar en cuenta que hubo unos que inclusive al considerar bastantes datos, fueron bastante buenos, y esto se debe a que no hay mucha variabilidad en cuando a las temperaturas en los respectivos estaciones de tiempo. COmo se mencionaba anteriormene, en el Invierno hay gran variabilidad debido a que se encuentra a una temperatura de 12°C, pero llega a subir a los 17°C o inclusive 18°C. QUe seria la temeperatura que se reporta para la primavera. Mientras que los modelos para la predicción de temperaturas para los estados del tiempo en Primavera y Verano, resultan ser los mejores debido a que en sus respectivas temperaturas no hay gran variabilidad. Y lo mismo ocurre con el modelo de otoño, que se encontraba a los 17°C, pero se reduce hasta los 12 °C que seria la temperatura con la que inicia el Invierno.

5 Predicciones

A continuación se mostrarán las predicciones a intervalos posteriores del tiempo elegido. Y se mostrara a manera de tabla. Esta comparación será con el modelo que mejor predice de acuerdo a los datos, con respecto a la que se tomo mayor tiempo para ver su comparación.

Invierno						
Año	85-87		85-89		85-94	
Pred.	1988		1990		1995	
Mes	$f_1(x_i)$	y_1	$f_2(x_i)$	y_2	$f_3(x_i)$	y_3
ENE	12.5	12.1	16.5	13.1	16.1	13.0
FEB	13.0	15.2	17.4	13.1	16.6	15.4
MAR	14.8	15.84	18.8	16.4	17.1	17.2

Primavera						
Año	85-87		85-89		85-94	
Pred.	1988		1990		1995	
Mes	$f_1(x_i)$	y_1	$f_2(x_i)$	y_2	$f_3(x_i)$	y_3
ABR	23.7	18.6	22.3	18.8	20.5	19.3
MAY	28.6	21.6	23.6	21.5	20.7	21.9
JUN	35.9	20.6	25.3	21	21.0	21.5

Verano						
Año	85-87		85-89		85-94	
Pred.	1988		1990		1995	
Mes	$f_1(x_i)$	y_1	$f_2(x_i)$	y_2	$f_3(x_i)$	y_3
JUL	20.4	19.3	17.7	18.8	19.9	19.5
AGO	20.8	19.0	17.1	18.4	20.3	19.2
SEP	21.1	18.3	16.4	17.9	20.7	18.5

Otoño						
Año	85-87		85-89		85-94	
Pred.	1988		1990		1995	
Mes	$f_1(x_i)$	y_1	$f_2(x_i)$	y_2	$f_3(x_i)$	y_3
OCT	11.9	17.2	10.8	16.9	16.5	16.9
NOV	8.8	15.9	8.1	14.8	16.7	15.6
DIC	3	13.7	4.6	13.1	17.0	13.3

5.1 Interpretación

Ahora, a pesar de que se eligiese siempre la que tenga la menor distancia entre sus temperaturas predichas con el dato real, tenemos que aquella función que fue alimentada con set menor de entrenamiento, resulto ser que al predecir, se alejó bastante al de la realidad, a comparación de aquellos con los que fue alimentado por la temperatura de 10 años. LLegando así, a predecir mejor la función alimentada con mayor información de temperatura, cuya predicción si cuenta con cierta precisión.

6 Conclusiones

Por medio de la programación y de las matemáticas, usando los teoremas establecidos como el de Cuadrados Mínimos, así como el de Cholesky, se pudieron encontrar grandes resultados y aproximaciones bastante buenas usando datos del clima de un estado de la república. Hubo un problema al elegir la segmentación del tiempo, debido a que el estado que se eligió cambia drásticamente a lo largo del año, al ser uno de los cálidos y al tener alto grado de continentalidad, se dificultaba tratar con certeza en cada mes y ver su predicción. Por lo que se decidió segmentarlo con base a las Estaciones del año. El lapso para generar las funciones fueron de 3, 5 y 10 años.

A pesar de que una función polinomial tenga menor mínimos cuadrados con respecto a otras funciones que se generaron a partir de data sets, mayores que el anterior, no indica que sea mejor o con mayor precisión su aproximación. Es decir, que para aquella función con la que se llenó con menor conjunto de datos, sólo nos permitía predicir hasta máximo 3 meses. Mientras que los otros dos, podían con 4 o inclusive 5 meses. Pero la de mayor precisión fue aquella función con la que se generó con una cantidad mayor de datos, es decir, la de 10 años.

7 Referencias

Conjunto de Datos: Temperatura promedio Excel, publicador Jorge Bustamante, CONAGUA, https://datos.gob.mx/busca/dataset/temperatura-promedio-excel

Código creado y usado en proyecto

Se encontrará disponible en repositorio de GitHub https://github.com/AlfonsBC