

Estudio R4

Explicación del Código

Este código implementa un algoritmo de selección de características utilizando un enfoque de adición hacia atrás (backward addition) que maximiza el error de predicción. Este método es una variante del enfoque de selección hacia adelante del primer código que analizamos y del enfoque de eliminación hacia adelante del tercer código. A continuación, se detalla el funcionamiento del código paso a paso:

Normalización de los Datos

Al igual que en los códigos anteriores, se centran las variables **X** e **Y** restando sus medias respectivas. Esto es crucial para evitar sesgos en el modelo debido a la escala de las características.

Selección de Características

El algoritmo inicialmente considera todas las características disponibles. En cada iteración, intenta añadir una característica y evalúa el modelo con las características actuales. La característica cuya adición resulta en el máximo aumento del error de predicción es considerada la más importante y se añade al conjunto de características.

Modelo Polinomial

Se utiliza un modelo polinomial de orden variable para ajustar los datos y realizar predicciones. Este modelo es similar al utilizado en los códigos anteriores, pero aquí se aplica a un conjunto de características que se va incrementando en cada iteración.

Cálculo del Error

El error de predicción se calcula utilizando la media de las diferencias absolutas entre las predicciones del modelo y los valores reales de **Y**, elevadas a la potencia de **alpha**. Cuando alpha se establece en 2, lo que corresponde al error cuadrático medio y si es 1 es el error absoluto medio.

Comparación con los Códigos Anteriores

- **Enfoque de Selección:** Mientras que el primer código añadía características de forma iterativa (selección hacia adelante), el segundo código las eliminaba (selección hacia atrás). El tercer código implementaba un enfoque de eliminación hacia adelante, pero seleccionaba la característica cuya eliminación maximizaba el error de predicción. Este cuarto código también implementa un enfoque de adición, pero con una diferencia clave: en lugar de seleccionar la característica que minimiza el error de predicción cuando se añade,

selecciona la característica cuya adición maximiza el error de predicción. Finalmente, se invierte la secuencia de características seleccionadas, por lo que la última característica añadida (que maximiza el error de predicción) se considera la más importante.

*Análisis visual de los resultados

Ya sea fijando $\text{Alpha}=1$ o $\text{Alpha}=2$ observo que:

La variable **113** es la que sale primera en la mayoría de cálculos de error **113** y **114** salen en ese orden en las primeras posiciones en todas menos cuando el orden del polinomio es 10

En cuanto al análisis de las variables menos relevantes, vemos que las variables **66 60** y **7**, son las que aparecen en las últimas posiciones.

He subrayado de amarillo las más relevantes y de rojo la que menos. Además, he puesto de color verde la letra si se repite la posición en la mayoría de veces y de naranja si pasa en algunas tablas.

Estudio de las gráficas:

Gráfico de error:

Se muestra un gráfico que representa el error en función del número de características seleccionadas. El eje x representa el número de características seleccionadas, mientras que el eje y representa el valor del error. El primer punto en el gráfico representa el error cuando no se selecciona ninguna característica.

Cuando $\text{Alpha}=1$

En cuanto al resultado, la primera "bola" para **Orden poli=2** es **0,23** mientras que los siguientes polinomios de mayor orden, son muy cercanos a **0.21 y 0.2**, mientras que el **orden poli=10** es prácticamente **0.23 aproximadamente**, por lo que el valor de error de la primera variable es prácticamente el mismo.

El error se muestra descendente en todos los casos a medida que aumente el número de variables seleccionadas

Si observamos el recorrido de las variables, es más **constante** cuando el polinomio es de **orden 2 3 4 y 5**, mientras que el polinomio de **orden 10** se muestra más **inestables** los demás polinomios son más **inestables** y se aproxima más a **0**.

En cuanto a las curvaturas, vemos que de forma general tienen forma constante de bajar, a excepción de cuando el orden del **polinomio 10**, que se muestra con mayor curvatura.

Cuando $\text{Alpha}=2$

En cuanto al resultado, la primera "bola" que se observa en la gráfica, es igual para todos porque es cuando no se selecciona ninguna variable, la siguiente para **Orden poli=2** es **0,08** y vemos que los siguientes polinomios de mayor orden, son muy

cercanos , mientras que el **orden poli =10** es prácticamente **0,1**, por lo que hay menor valor de error con la primera característica
El error se muestra descendente en todos los casos a medida que aumente el número de variables seleccionadas
Si observamos el recorrido de las variables, es más **constante** cuando el polinomio es de **orden 2 3 4 y 5**, mientras que el polinomio de **orden 10** se muestra más **inestables** los demás polinomios son más **inestables** y se aproxima más a **0**.

*Análisis Final

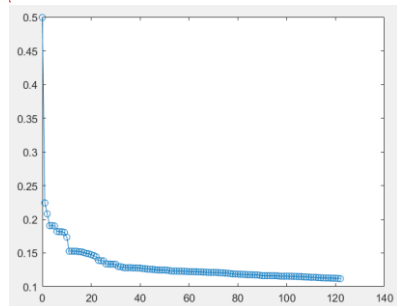
La secuencia de errores y su análisis derivativo pueden proporcionar información adicional sobre la estabilidad del proceso de selección de características y la calidad de las características seleccionadas.

En resumen, este cuarto código proporciona otra estrategia para la selección de características, que puede ser útil en situaciones donde se desea mantener las características que más contribuyen a la precisión del modelo. La elección entre un enfoque de selección hacia adelante, hacia atrás, de eliminación hacia adelante o de adición hacia atrás, son distintas visiones de realizar estos rankings que puede depender de cuál sea el objetivo del proyecto.

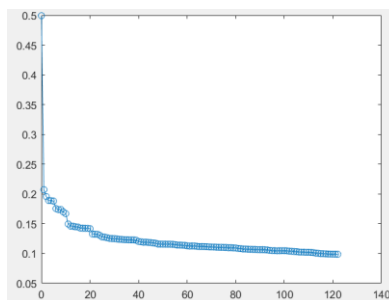
GRÁFICAS

ALPHA1

Orden poli 2



Orden poli 3

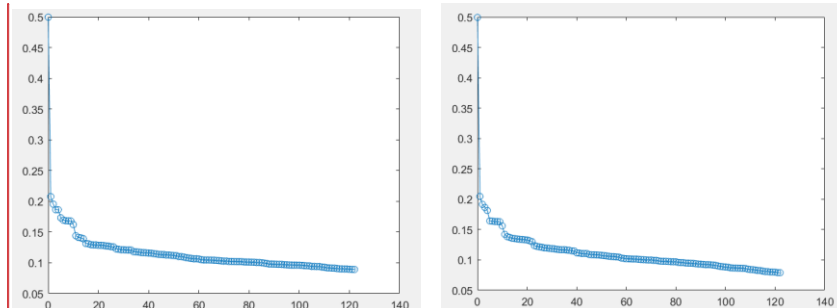


Comentado [CP1]: orden poli 2

Orden poli 4

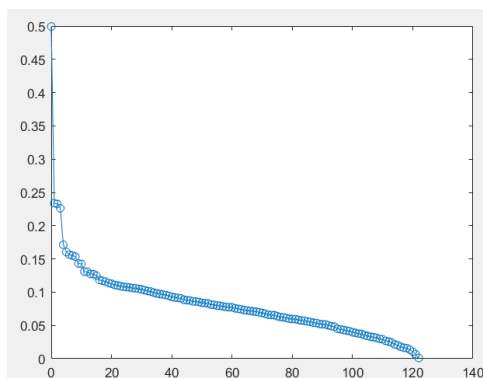
Orden poli 5

Comentado [CP2]: orden poli 4



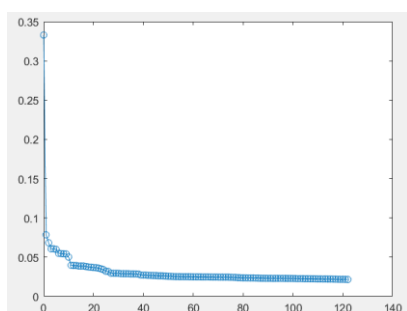
Comentado [CP3]: orden poli 5

Orden poli 10

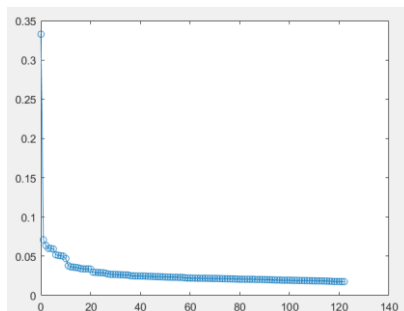


ALPHA2

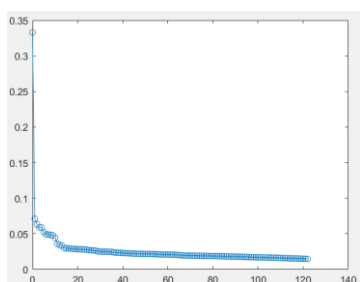
Orden poli 2



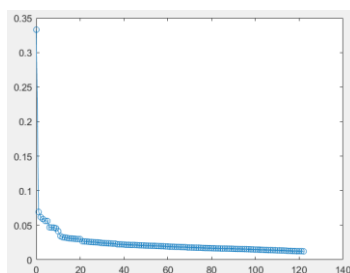
Orden poli 3



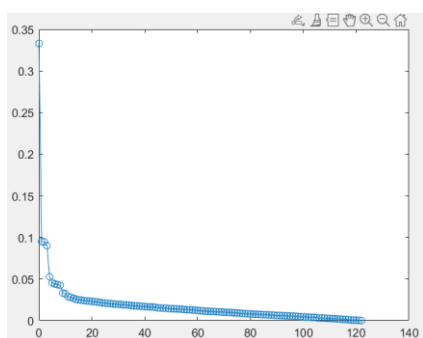
Orden poli 4



Orden poli 5

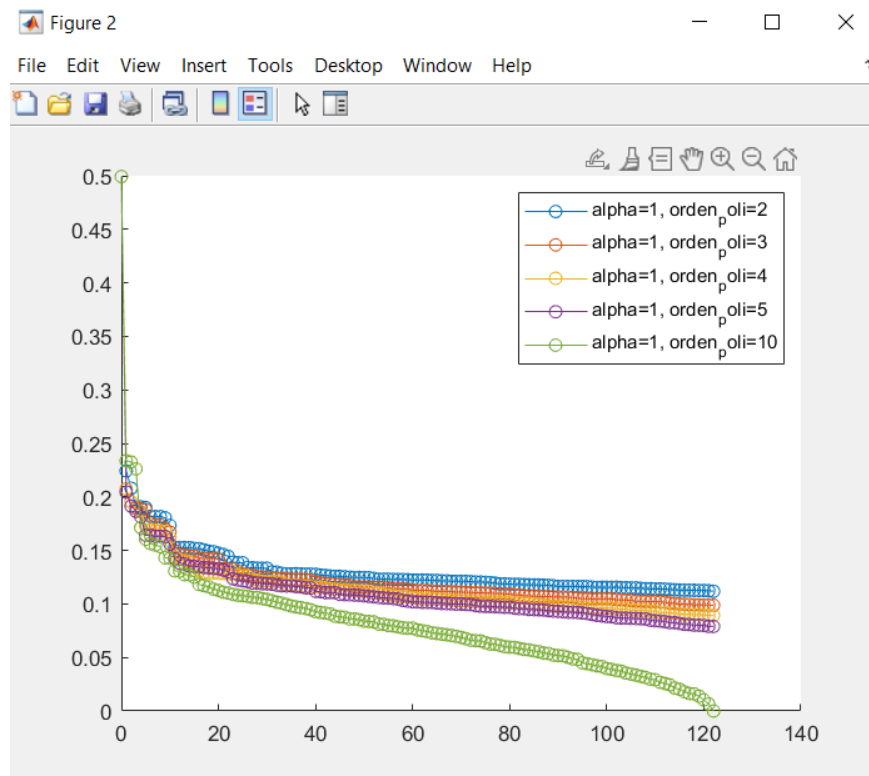


Orden poli 10

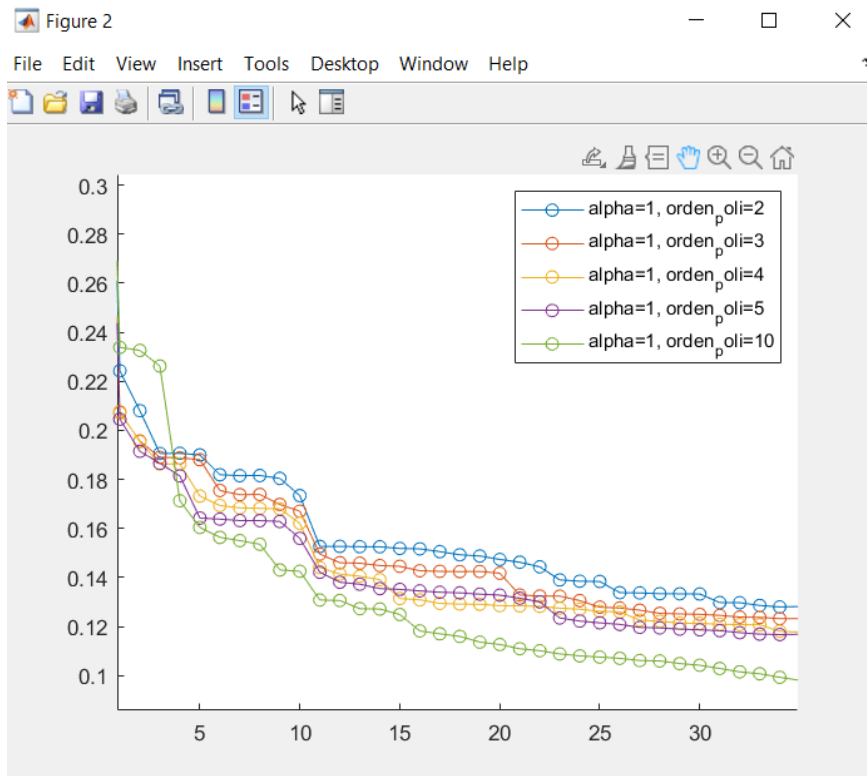


Gráficas comparativas

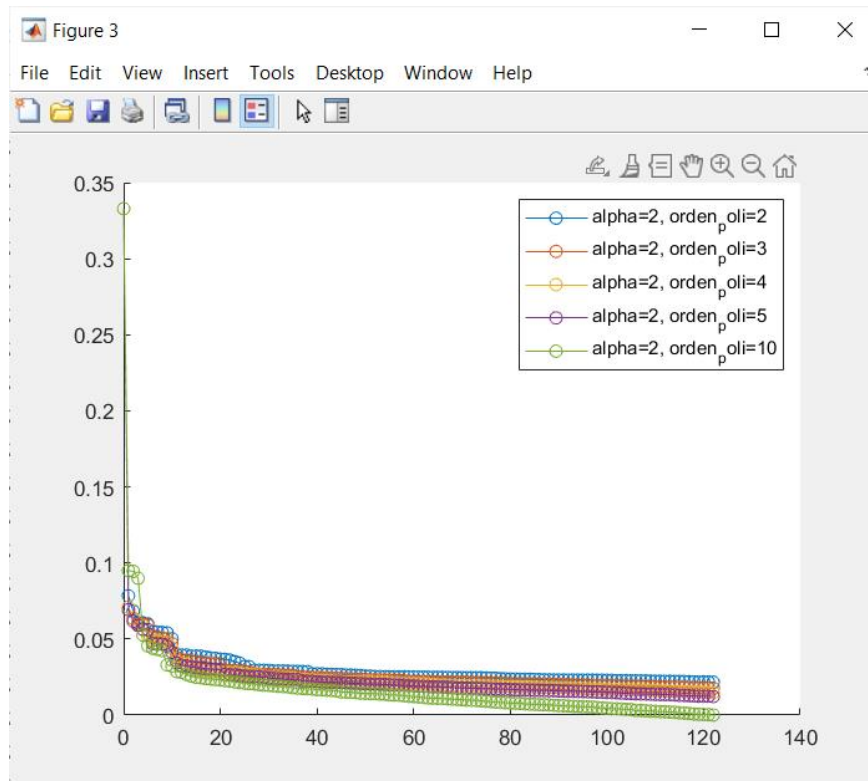
CUANDO ALPHA 1



Zoom:



CUANDO ALPHA 2:



Zoom:

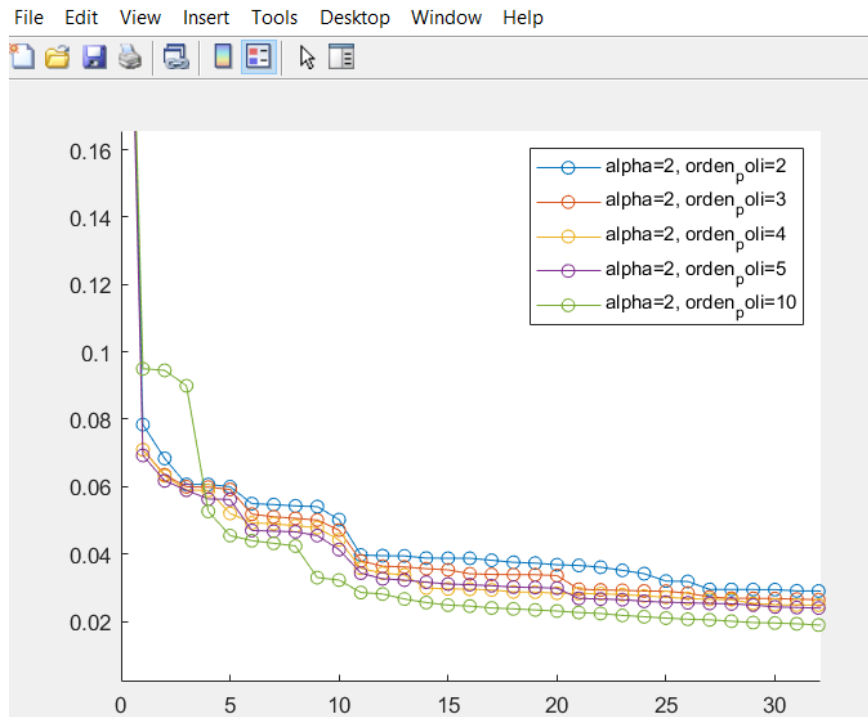


TABLA DE DATOS:

{'Alpha'}	{'Orden Poli'}	{'Error' }	{'Solución' }
{[1]}	{[2]}	{[0.2243 0.2080 ...]}	{[113 114 115 1 ...]}
{[1]}	{[3]}	{[0.2073 0.1956 ...]}	{[113 114 1 115 ...]}
{[1]}	{[4]}	{[0.2076 0.1954 ...]}	{[113 114 1 115 ...]}
{[1]}	{[5]}	{[0.2046 0.1915 ...]}	{[113 114 50 1 ...]}
{[1]}	{[10]}	{[0.2338 0.2326 ...]}	{[115 1 51 113 ...]}
{'Alpha'}	{'Orden Poli'}	{'Error' }	{'Solución' }
{[2]}	{[2]}	{[0.0784 0.0684 ...]}	{[113 114 1 115 ...]}
{[2]}	{[3]}	{[0.0710 0.0635 ...]}	{[113 114 1 115 ...]}
{[2]}	{[4]}	{[0.0709 0.0632 ...]}	{[113 114 1 115 ...]}
{[2]}	{[5]}	{[0.0693 0.0617 ...]}	{[113 114 50 1 ...]}
{[2]}	{[10]}	{[0.0950 0.0945 ...]}	{[1 115 51 113 ...]}

Cuando Alpha =1

[illegible][illegible]

[illegible]

Orden del Polinomio 4

[illegible]

Orden del Polinomio 5

[illegible]

Orden del Polinomio 10

[illegible]

[illegible]

Cuando Alpha =2

Orden del Polinomio 2

[illegible]

4	15
3	21
	118
	47
1	45
9	99
3	105
4	55
01	109

	15
	22
	88
	63
	26
	5

[illegible]

Orden del Polinomio 4

[illegible]

[illegible]

Orden del Polinomio 5

[illegible]

Orden del Polinomio 10

[illegible]